R@Loc 2014

REREARCH@**L**OCATE'14

Proceedings

Canberra, Australia
April 7-9, 2014

Stephan Winter, Chris Rizos (Eds.)

Editors

Stephan Winter
Department of Infrastructure Engineering
The University of Melbourne
Parkville, VIC 3010
Australia
winter@unimelb.edu.au

Chris Rizos
School of Civil & Environmental Engineering
University of New South Wales
Sydney, NSW 2052
Australia
c.rizos@unsw.edu.au

Preface

Locate is the new annual conference on spatial information in Australia and New Zealand, bringing together for the first time all joint forces of SSSI, SIBA, ANZLIC, OSP, LINZ, CRCSI, PSMA, ASIERA and others. Locate'14 has been the inaugural event, and from here Locate aims to become the annual meeting point of industry, government and academia in one of the fastest growing areas of IT: spatial information.

Research@Locate is the academic research stream at Locate. It is organized independently by the Australasian Spatial Information Education and Research Association, ASIERA (www.asiera.org.au). Research@Locate has been designed to provide an Australasian research conference with a transparent full-paper peer review process, with carefully selected presentations and papers, and with its own annual, open-access proceedings. It aims to become the premier academic event in the Australasian region.

Already in its first year Research@Locate received 42 submissions. After a thorough peer review process, 16 of these papers were selected for the conference proceedings. This result corresponds to an acceptance rate of 38\%. In addition to the 16 accepted full papers, four papers received sufficient reviewer support to be invited for presentation only, and for three of these presentations you will find short abstracts in the proceedings.

What you will find in these proceedings is a selected sample of research in the field of spatial information. The presentations this year were organized around four themes (sessions): sourcing and access, accuracy, processing and analysis, and algorithms. A few review papers are also included.

Research@Locate is organised by the Australasian Spatial Information Education and Research Association, ASIERA (http://www.asiera.org.au). ASIERA represents a significant part of the academic segment of the spatial information industry in Australia, with a workforce of several hundred people in fundamental and applied research and innovation, with a responsibility for educating and training future generations of spatial professionals.

Acknowledgements

Research@Locate would not have happened without the support of the institutions behind Locate, SIBA and SSSI. We also wish to thank our colleagues that served on the International Program Committee, and Maria Vasardani for the production of the proceedings.

April 2014

Stephan Winter, Chair

Chris Rizos, Co-Chairs

Program Committee

Program Committee Chair 2014

Stephan Winter, The University of Melbourne, Australia (chair)
Chris Rizos, The University of New South Wales, Australia (co-chair)

Program Committee 2014

Sharolyn Anderson, University of South Australia, Australia
Jagannath Aryal, University of Tasmania, Australia
David Bruce, University of South Australia, Australia
Nick Chrisman, RMIT University, Australia
Paul Corcoran, University of South Australia, Australia
Matt Duckham, The University of Melbourne, Australia
Ahmed El-Mowafy, Curtin University, Australia
Ori Gudes, Curtin University, Australia
Eric Guilbert, The Hong Kong Polytechnic University, Hong Kong
John Hayes, Queensland University of Technology, Australia
Petra Helmholz, Curtin University, Australia
Mohsen Kalantari, The University of Melbourne, Australia
Allison Kealy, The University of Melbourne, Australia
Lars Kulik, The University of Melbourne, Australia
Ki-Joune Li, Pusan National University, South Korea
Samsung Lim, University of New South Wales, Australia
Bharat Lohani, IIT Kanpur, India
Kim Lowell, Cooperative Research Centre for Spatial Information, Australia
Feng Lu, State Key Laboratory of Resources and Environmental Information Systems, China
Nandakumaran Nadarajah, Curtin University, Australia
Gerhard Navratil, Technical University Vienna, Austria
Kevin McDougall, University of South Queensland, Australia
Antoni Moore, University of Otago, New Zealand
Jon Osborne, University of Tasmania, Australia
Abbas Rajabifard, The University of Melbourne, Australia
Femke Reitsma, University of Canterbury, New Zealand
Chris Rizos, The University of New South Wales, Australia

Wenzhong Shi, The Hong Kong Polytechnic University, Hong Kong
Pascal Sirguey, University of Otago, New Zealand
Egemen Tanin, The University of Melbourne, Australia
Martin Tomko, The University of Melbourne, Australia
Xiaohua Tong, Tongji University, China
Bert Veenendaal, Curtin University, Australia
Jinling Wang, The University of New South Wales, Australia
Geoff West, Curtin University, Australia
Cecilia Xia, Curtin University, Australia
Bisheng Yang, Wuhan University, China
Zhangcai Yin, Wuhan University of Technology, China

Steering Committee

Research@Locate is backed by the current members of ASIERA:

David Bruce, University of South Australia, Australia
Nick Chrisman, RMIT University, Australia
John Hayes, Queensland University of Technology, Australia
Kevin McDougall, University of South Queensland, Australia
Xiaoli Deng, University of Newcastle, Australia
Jon Osborne, University of Tasmania, Australia
Femke Reitsma, University of Canterbury, New Zealand
Chris Rizos, The University of New South Wales, Australia
Pascal Sirguey, University of Otago, New Zealand
Stephan Winter, The University of Melbourne, Australia
Bert Veenendaal, Curtin University, Australia

MEMBERS

The University of Melbourne
The University of New South Wales
RMIT University
Curtin University of Technology
The University of Newcastle
University of Tasmania
University of Southern Queensland
University of South Australia
Queensland University of Technology
University of Canterbury
University of Otago

Table of Contents

Sourcing and Access

Accuracy

Processing and Analysis

Algorithms and Review

Identifying salient topics for personalized place similarity

Benjamin Adams
Centre for eResearch
The University of Auckland, New Zealand
b.adams@auckland.ac.nz

Martin Raubal
Institute for Cartography and Geoinformation
ETH Zürich, Switzerland
mraubal@ethz.ch

Abstract

The ability to find similar places is an important component to geographic information retrieval applications as varied as travel recommendation services, marketing analysis tools, and socio-ecological research. Using generative topic modelling on a large collection of place descriptions, we can represent places as distributions over thematic topics, and quantitatively measure similarity for places modelled with these topic signatures. However, existing similarity measures are context independent; in cognitive science research there exists evidence that when people perform similarity judgments they will weigh properties differently depending on personal context. In this paper we present a novel method to re-weight the topics that are broadly associated with a place, based on users' interests inferred from sample place similarity rankings. We evaluate the method by training topics associated with texts about places, and perform a user study that compares user-provided similar places to those provided by automatically personalised place rankings. The results demonstrate improved correspondence between user rankings and automated rankings when personalised weights are applied.

1 Introduction

People commonly use known places as referents for communicating information about other parts of the world. For example, when we describe a place as being *like Canberra, Australia*, that description comes laden with semantics that allow us to infer attributes about that place that are otherwise left unspoken. These implicit attributes of places often differ from explicitly represented properties of places, such as census data, econometric data, and spatial footprints, that are commonly found in geographic information databases. Instead, these attributes are emergent from people's experiences and are revealed through their communication about those places. However, the implicit attributes that a person associates with a given place will also be highly contextual, and thus personalisation is required in order to make useful geographic information retrieval applications that utilise these kinds of attributes.

Operationalising background knowledge about places is potentially useful for several kinds of information applications, including geographic information retrieval and exploration systems (Larson, 1996; Jones and Purves, 2008), mobile applications, tools for social science research and education, large scale proposals such as the Digital Earth system (Grossner et al., 2008), and applications in the nascent field of spatial humanities (Bodenhamer, 2010). Additionally, in the technology sector, place knowledge is used to create location-based advertising services (Ranganathan and Campbell, 2002); with an understanding of the locational and temporal context in which a mobile user is placed, advertising content can be better targeted (Banerjee and Dholakia, 2008). A core functionality that we require in these different applications is the ability to find semantically similar places based on an operationalisation of this implicit knowledge.

In this paper we consider the problem of personalising the search for similar places, which are represented as vectors of probability values associated with a set of features or attributes. For demonstration we focus on characterising *places*

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

as mixtures of topics derived from probabilistic topic modelling on a corpus of place descriptions. These descriptions can take many forms, including travel entries, encyclopaedia articles, newspaper articles, and social media postings. These data sources are plentiful and all provide insight into the experiences of people in various places and, in particular, attributes that are commonly described for those places. Topics are trained using latent Dirichlet allocation (LDA) topic model inferencing on a set of these crowdsourced documents (Blei et al., 2003). Topic model training applies Bayesian inference to arrive at a representation of each place as a probability vector over a finite set of topics. Once we have a topic distribution for each place, we can compare their similarities in terms of the relative entropy of these topic distributions. This measure captures the similarity of the places as derived from an aggregation of descriptions from many people. In other words, the similarity of two places is defined as similarity of the prototypical or "average" representations of the places. There is psychological evidence that for many kinds of categories a prototypical instance of the category can be represented as an average from a set of exemplars in this manner (Rosch, 1978).

However, there is also evidence that when performing similarity (or dissimilarity) judgements on a set of stimuli, people will assign higher salience to some properties than others (Medin et al., 1993). In conceptual space theory, it is proposed that these salience differences can be modelled as weights on the quality dimensions on which these similarity measurements are based (Gärdenfors, 2000; Raubal, 2004). When people judge the similarity of two places, they choose a set of salient properties on which to compare the two places. This set of properties will be smaller than the total set of possible properties on which one could possibly compare the two places (Tversky, 1977). In addition, the set of properties that are salient will vary from person to person and depend on the places being compared. The challenge, therefore, is to identify which weighting to use in a particular context. This feature selection problem becomes one of further reducing the dimensionality of the topic space to a subset of topics that are relevant to a user in a specific context.

In this paper we present a methodology to 1) generate a prototypical mixture of topics associated with places based on crowdsourced natural language descriptions, 2) identify the topics of interest to an individual user, and 3) provide a personalised ranking of similar places to a source place based on the inferred interests of the user. We propose a method to automatically re-weight the topics using a sample similarity ranking for a small subset of the data.

In the following section, we provide background on computing place, personalised search and ranking, and topic models. Section 3 describes our method to identify salient topics. Following that, Section 4 presents the results of a user evaluation of the method. Finally, we conclude and point to future work.

2 Background

This section provides background material on computing place, personalised search and ranking, place recommendation, and topic modelling algorithms used in this paper.

2.1 Place

Several researchers have noted that place is a subjective, experiential phenomenon that is socially constructed (Relph, 1976; Tuan, 1977). Simply put several of the characteristics of places that are salient to people cannot be captured by a spatial footprint and structured data, such as population or median income. The difficulty in developing place representations that take into account common-sense understandings has meant that, to a great degree, place has been represented in a limited manner in geographic information systems (GISs). However, there has been a renewed interest in recent years in modelling place more holistically as it becomes clear that traditional models are insufficient for the needs of many technologies (Jordan et al., 1998; Winter et al., 2009). An increasingly vast amount of data is available online that describe places and types of places in natural language. They come in a variety of forms: e.g., travelogues, encyclopaedia entries, microblogs, literary works, social media sites, etc. These documents provide us with an extraordinary opportunity to understand places not in terms of quantitative, tabular data, but rather in terms of how people write about them.

2.2 Personalised search and ranking

Personalised search and ranking on the World Wide Web has been an ongoing research area for several years (Pitkow et al., 2002; Teevan et al., 2005; Micarelli et al., 2007). Pitkow et al. (Pitkow et al., 2002) have identified two main strategies for personalising search: 1) query augmentation, which uses additional contextual information to refine the query; and 2) result processing, which filters the search results based on a user model. In addition to algorithmic approaches to personalisation, user interface design can play a large role in how well search results can be personalised (Teevan et al., 2005).

Much of the personalisation research relies on using previous search behaviour to develop a model of the user's interests and using this model to personalise search results on the web (Teevan et al., 2005). Developing a user profile based on implicit information has several advantages over asking the user for explicit feedback, because not only does gathering

explicit feedback impose an added burden on the user but also users can provide inaccurate information (Speretta and Gauch, 2005). In mobile search, user-based personalisation can be augmented by location-based personalisation and these factors can interrelate to model how a user's interests change based on location (Bouidghaghen et al., 2011). Formal ontologies can be used to align user interests that are inferred from search history with well-defined concepts described in a semantic hierarchy (Daoud et al., 2007).

Once there is a model that associates user categories with specific interests, it is possible to use different methods of characterising user similarity to provide personalised results for categories of users (McKenzie et al., 2013). Guy et al. (Guy et al., 2010) have explored comparing user similarity in social media along 3 different axes of people, things, and places.

2.3 Semantic similarity in GIScience

The importance of context in semantic similarity measurement is a well-studied research topic in geographic information science (Raubal, 2004; Rodríguez and Egenhofer, 2004; Janowicz et al., 2011). Commonly, context is modelled by applying weights to the attributes that used to describe geospatial features; however, finding the appropriate weights is a challenge (Keßler, 2012). Semantic referencing is an algorithm schema that uses the idea of control similarities provided by a user to semi-automatically calibrate factor weights for similarity measurement of geographic entities (Janowicz et al., 2010). The methods presented here are a form of semantic referencing.

The problem addressed in this paper (personalised similarity of places) is distinct from previous work on personalised web ranking, since we focus on the specific problem of finding similar *places*. Although individuals have different impressions of places, due to the nature of places being manifested in the physical world, there are broad regularities in the types of activities and features that people write about a given place. Because of these regularities, a crowdsourced representation of a place will have broad applicability. Thus, the search for places is not an information retrieval task on an individual document or artefact but rather a similarity search for places, represented as aggregations derived from multiple documents. Personalised place similarity remains a relatively unexplored research area outside of smaller-scale investigations in tourism and sense-of-place research (Moore and Graefe, 1994; Kaltenborn, 1998; Jorgensen and Stedman, 2006) and sociological investigations of place attachment (Stedman, 2002).

As with prior user behaviour for web personalisation, examples of similar places provide a mechanism to infer the interests of the user without requiring him or her to explicitly enter interests. Furthermore, the sample ranking of similar places – and which is provided directly by the user in this evaluation – can also be gathered from social data.

2.4 Topic models

Probabilistic topic models, such as latent Dirichlet allocation (LDA), provide methods to characterise the documents in a natural language corpus as multinomial mixtures of topics (Blei et al., 2003). Each topic is further modelled as a discrete probability distribution over all the words in a corpus. The top most probable words for a topic are usually semantically interpretable by humans. E.g., a topic with the top terms "wine, red, drink, cheese" can be interpreted as being about the activity of wine tasting (Adams and McKenzie, 2013).

LDA models the generation of each document in the corpus as an iterative process; first, a topic is chosen based on the topic distribution for the document; second, a word is randomly selected from that topic. This is repeated for each term in the document. Since each word is drawn randomly and is not based on the previous word, word order makes no difference in the LDA model. Assuming this generative process for how the documents were created, the words of an existing set of documents become the observed variables in the model and Bayesian inference is used to infer the most-likely topics to have generated the data. This inference can be performed programmatically in an approximate manner using a markov chain Monte Carlo Gibbs sampling algorithm (Griffiths and Steyvers, 2004).

Apart from discovering the latent topics in a corpus in an unsupervised manner, this dimensionality reduction allows one to compare the thematic similarity of two documents without requiring that the documents share exact terms. A natural way to measure the similarity of documents is the calculate the Kullback-Leibler (KL) divergence between the two topic distributions, P and Q (see Equation 1) (Steyvers and Griffiths, 2007).

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_i \ln\left(\frac{P(i)}{Q(i)}\right) P(i). \tag{1}$$

Since KL divergence is asymmetric, the Jensen Shannon (JS) divergence can be used if a symmetric measure is desired (see Equation 2).

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2} D_{\mathrm{KL}}(P \parallel M) + \frac{1}{2} D_{\mathrm{KL}}(Q \parallel M), \tag{2}$$

where $M = (P + Q)/2$. These methods of measuring similarity ignore that the importance of individual topics will vary from user to user.

Several extensions to LDA have been developed to characterise the mixture of topics associated with a location or place (Wang et al., 2007; Eisenstein et al., 2010; Hao et al., 2010; Sizov, 2010; Yin et al., 2011; Hong et al., 2012). In these models the generative process for creating a document is extended and some additional form of geographic evidence (either location or place name) is used to condition topics based on that evidence. Thus, a distribution of topics can be associated not only with a document but also a place or location. The technique employed here for evaluation is to represent the topic mixture for a place as a weighted average of the topic probability distributions of all the descriptions for the same place (Adams and Janowicz, 2012). Since the methods described in this paper for weighting topics are post-hoc operations (i.e. assigned after the topic model inferencing), they are broadly applicable to these other flavours of topic models.

3 Identifying salient topics

In this section, we present a method for personalising the weights on the LDA topics, given a source place s and a set of N target places $\{t_1, t_2, \ldots, t_N\}$, and a user specified ordering for those target places based on their similarity to the source place. For each place there is a discrete probability distribution for topics, and for any given individual topic the strengths of that topic for both source and target places can be compared. For example, let the following be a sample user ranking of three cities in terms of similarity to *New York City*:

1. Chicago
2. Los Angeles
3. Houston

Table 1 shows sample topic strengths for each of these cities (source and targets).

City	topic 1	topic 2	topic 3
New York City	0.2	0.6	0.2
Chicago	0.2	0.2	0.6
Los Angeles	0.42	0.38	0.2
Houston	0.8	0.1	0.1

Table 1: Sample topic strengths for source city New York and three target cities.

The set of salient topics Z_S is defined as a subset of all topics, such that $z \in Z_S$ if and only if the Kendall's τ rank correlation coefficient between the user-provided ordering and the topic strengths for topic z is positive (Kendall, 1938). The Kendall's τ measures rank correlation as a function of the number of concordant and discordant pairs in a set of joint observations. A pair of observations (i, j) of two variables (x, y) is concordant if $x_i < y_i \wedge x_j < y_j$ or $x_i > y_i \wedge x_j > y_j$. They are discordant if $x_i < y_i \wedge x_j > y_j$ or $x_i > y_i \wedge x_j < y_j$. Letting C be the number of concordant pairs and D be the number of discordant pairs in a sample and n be the number of observations, Kendall's τ is defined in Equation 3.

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}. \tag{3}$$

Thus, the relative weight (π_i) for each topic is zero for negative correlations. For positive correlations the weight is the correlation normalised such that all positive correlations sum to 1. See Table 2 for an example.

Kendall's τ-B is an extension to Kendall's τ to deal with situations where there are tied rankings. This will occur with topics when one or more documents have a value of 0.0 for the topic strength. When a model is trained using a large

user ordering	topic 1	topic 2	topic 3
Chicago	Chicago (0.0)	Los Angeles (0.22)	Los Angeles (0.0)
Los Angeles	Los Angeles (0.22)	Chicago (0.4)	Houston (0.1)
Houston	Houston (0.6)	Houston (0.5)	Chicago (0.4)
Kendall's τ	1.0	0.33	-0.67
Relative weight	0.75	0.25	0.0

Table 2: Rank ordering of target cities based on topic strength similarity. The difference between topic strengths is shown in parentheses. For each topic the Kendall's τ rank correlation is shown between the user similarity judgement and topic strength difference.

number of topics this will occur often. Kendall's τ-B (Equation 4) handles tied values by changing the denominator of the equation.

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}, \tag{4}$$

where $n_0 = n(n-1)/2$; $n_1 = \sum_i t_i(t_i - 1)/2$; $n_2 = \sum_j u_j(u_j - 1)/2$; n_c = number of concordant pairs; n_d = number of discordant pairs; t_i = number of tied values in the i^{th} group of ties for the first quantity; and u_j = number of tied values in the j^{th} group of ties for the second quantity.

3.1 Using Kendall's τ as product

For each salient topic, z_i, an associated salience weight, w_i, is assigned to the topic equal to the Kendall's τ correlation. Given a set of salience weights $w_1, w_2, \ldots w_n$ we can define relative weights for each topic $\pi_1, \pi_2, \ldots \pi_n$, where $\pi_i = \frac{w_i}{\sum_1^n w}$. Let π be this vector of relative weights, a *weighted Kullback-Liebler divergence A* is now defined in Equation 5.

$$D_{\mathrm{KL}}(\pi; P \parallel Q) = \sum_i \pi_i P(i) \ln \frac{P(i)}{Q(i)}. \tag{5}$$

The divergence is weighted not only on the probabilities of the topic variables but also with the salience weight based on the ordering, and it is thus a weighted average of the logarithm difference between the probabilities. From this new divergence function, the *weighted Jensen Shannon divergence A* is defined in Equation 6.

$$JSD(\pi; P \parallel Q) = \frac{1}{2} D_{\mathrm{KL}}(\pi; P \parallel M) + \frac{1}{2} D_{\mathrm{KL}}(\pi; Q \parallel M), \tag{6}$$

where M is defined as in the normal JS divergence.

This weighted measure can be interpreted as a form of weighted sum model in a multi-criteria decision analysis.The salience weights drawn from a sample ranking can be viewed as a personalised general measure of topic saliency for comparing all places. Table 3 shows a comparison of Jensen Shannon divergence and weighted Jensen Shannon divergence A calculations given the user ranking specified above. Note, in this example the weighted ordering matches the user specified ordering; however, it is not guaranteed to match in all cases, since the salience weights are based on a comparison of rankings, not actual similarity values of the probabilities.

3.2 Using Kendall's τ to alter topic weights

An alternative method re-weights the topic probabilities associated with each place and uses the traditional JS divergence measure to obtain a context-dependent similarity. As with the previous method only positive τ correlations are used. Letting $\mathbf{z}$ be the topic vector for a place, Algorithm 1 describes the steps for re-weighting topics. Each topic probability, z_i, is multiplied by the weight for that topic, π_i, and then the new topic vector is normalized to sum to 1 (see Algorithm 2).

Algorithm 1 Pseudocode for re-weighting topics from user provided ranking.

```
sum ← 0.0
for all z_i do
  if w_i > 0 then
    z'_i ← w_i * z_i
    sum ← sum + z'_i
  else
    z'_i ← 0.0
  end if
end for
for all z'_i do
  z'_i ← z'_i / sum
end for
```

This new topic distribution for a place has a very intuitive interpretation. The re-weighted probability of topic i, z'_i, is a function of the prototypical probability of the topic, z_i, for the place, conditioned on the probability that this specific user is concordant with the average user for this topic. This re-weighted distribution has the advantage that it maintains the desired mathematical properties when comparing similarities using KL and JS divergences (i.e., a minimum divergence of

0 and the square root of the JS divergence is a metric). In Table 3 the Jensen Shannon result using this method is called the *weighted Jensen Shannon B* measure.

city	JS	weighted JS B
Chicago	0.151	0.049
Los Angeles	0.047	0.057
Houston	0.294	0.220

Table 3: Comparison of JS divergence and weighted JS divergence values from New York City given a user ranking of *1. Chicago, 2. Los Angeles, 3. Houston* (lower values indicate more similar). Note, that JS and weighted JS values are not comparable, only ranks.

The method described above works well in cases where the user rankings provide information to the system for all the topics. However, when using a system trained on a large number of topics (e.g., several hundred) there can be individual topics where all of the target places have a weight of 0.0. This means that all the places being ranked are equally similar to the source place in terms of that single topic.

In this case, we only want the topics that are informed by the ranking to affect each other while keeping the probability of all other topics fixed. Let Z be the set of all topics and $Z_{NI} \subseteq Z$ be the set of all topics for which the ranking does not provide any information. The final version of the algorithm is shown in Algorithm 2.

4 Evaluation

In this section, we present the results of a human participants test to evaluate the efficacy of using sample rankings to prime salience weights on individual topic dimensions. LDA topic models were trained on two corpora of georeferenced place descriptions: 1) a set of 200,000 georeferenced Wikipedia articles and 2) a set of 275,000 travel blog entries.[1] Prior to performing topic modelling, standard stemming and cleaning of the documents was performed to remove html tags and other noise in the text. The travel blog entries were matched to places by the authors according to a fixed geographic hierarchy, e.g., *Orlando, Florida, United States*. Each georeferenced Wikipedia article was matched to a named place by intersecting the location associated with the article (based on coordinates template in the article) with the spatial footprint of the named places. Thus, each article is linked to one place. It is possible that the association between text and a place can be refined further using more sophisticated techniques; however, this still remains a open research problem and not the focus of the current work (Vasardani et al., 2013).

The topic signature for a place was calculated by taking all the documents with a relation to the place and averaging over their topic vectors. Thus, topics that have a relatively high probability in a large number of documents associated with a place will be associated with the place (e.g., in travel blog entries a *theme park* topic will have high value for Orlando, FL, but a *skiing* topic will not).

4.1 Design

The user study was designed using the Amazon Mechanical Turk system to gather several place similarity rankings.[2] A qualification test was presented to the Mechanical Turk users to help ensure high-quality (non-spam) results (Kittur et al., 2008). This test was a simple filter that gathers user information and asks a set of simple questions about geographic knowledge. The primary goal of the survey was to determine the degree to which a user-provided ranking on a set of cities can be used to inform the system on the "interests" of the user more generally, and thus tailor search results.

The study focused on comparing 30 U.S. cities rather than other types of places in order to maximise participant familiarity. Although initial variants of the test were done using cities from around the world, it was difficult to find users who had broad familiarity with multiple cities from around the world. Therefore, the tests were limited to residents of the U.S., and a good sample of participants with familiarity of target cities could be gathered. Table 4 shows all the cities that were compared.

Mechanical Turk participants were asked to rank order 7 cities in terms of similarity to another city. Users were also asked to enter how the places are similar as well as describe their familiarity with each of the 8 cities. Table 5 shows two sample responses for cities that are ranked in similarity to Los Angeles.

In total 96 participants provided 5 rankings each, leading to 480 rankings total. Out of the 480 rankings, in 135 cases (28.1%) the participant was 'not' familiar with the main city being compared, in 205 cases (42.7%) the participant was

[1] downloaded from `http://www.travelblog.org`

[2] `http://www.mturk.com`

Algorithm 2 Pseudocode for re-weighting topics from user provided ranking when not all topics are informed by the ranking.

```
sum ← 0.0
sumOthers ← 0.0
for all z_i do
  if z_i ∈ Z_NI then
    z'_i ← z_i
    sumOthers ← sumOthers + z_i
  else if w_i > 0 then
    z'_i ← w_i * z_i
    sum ← sum + z'_i
  else
    z'_i ← 0.0
  end if
end for
for all z'_i do
  if z'_i ∉ Z_NI then
    z'_i ← z'_i * (1− sumOthers )/ sum
  end if
end for
```

Atlanta	Chicago	Houston	Minn.-St. Paul	Phoenix	San Francisco
Austin	Cleveland	Kansas City	Nashville	Pittsburgh	San Jose
Baltimore	Dallas	Las Vegas	New Orleans	Portland, OR	Seattle
Boston	Denver	Los Angeles	New York City	Salt Lake City	St. Louis
Charlotte	Detroit	Miami	Philadelphia	San Diego	Washington D.C.

Table 4: United States cities that were ranked based on similarity. For each comparison 7 of these cities were ranked based on similarity to 1 other city.

'somewhat' familiar, and in 140 cases (29.2%) the participant was 'very' familiar. In 50 cases (10.4%) the participant had lived in (or near) the city and in 290 cases (60.4%) the participant had visited the city.

4.2 Spearman's footrule distance

Spearman's footrule distance is a measure of correspondence (or disarray) between two rankings, similar to Spearman's ρ and Kendall's τ. Let S_n be the set of all permutations of the set of n integers $\{1, \ldots, n\}$. From (Diaconis and Graham, 1977), the Spearman's footrule distance, D_n is defined as in Equation 7.

$$D_n(\pi_n, \sigma_n) = \sum_{i=1}^{n} |\sigma_n(i) - \pi_n(i)|\,, \tag{7}$$

where π_n and σ_n are elements of S_n, i.e., different permutations. Spearman's footrule is used here rather than Spearman's rho, because we want to be able to consider the average footrule distance between several automated rankings and user provided rankings, which is not valid for correlations.

4.3 Average Spearman's footrule results

In Figure 1 a chart is shown of the average Spearman's footrule distance between the human participant rankings and automated similarity calculations. Looking over all responses, the clearest outcome was that for source places with which the participant is not familiar; generally, a participant's assessment is primarily made based on distance rather than other properties. When the participant is somewhat or very familiar with the source place there is an increase in the average Spearman's footrule distance between distance-based rankings.

Looking at Figure 1 it is abundantly clear that similarity ranking based on all travel blog topics or all Wikipedia articles does not correspond in any definitive way with user-provided rankings. In part this is because there is no "average person" – there is very little correspondence between different participants about the appropriate ranking of similar cities. In

Similar to Los Angeles ranking 1	
Miami	active nightlife
Minneapolis-St. Paul	skyscrapers
Dallas	home to corporations
Austin	hot weather
Salt Lake City	big houses
Cleveland	busy
Portland, OR	tourist towns

Similar to Los Angeles ranking 2	
Miami	beautiful people; emphasis on appearance
Dallas	very spread out
Austin	music and performing arts
Portland, OR	music and performing arts
Minneapolis-St. Paul	music and performing arts
Salt Lake City	western US
Cleveland	both are cities with large areas in economic decline

Table 5: Sample rankings of similarity to Los Angeles from two different participants.

addition, individuals have a limited set of properties that they use to compare places as opposed to the large number of different topics found in large text corpora. This is not a problem with the methodology, however, because the task that the system is to perform is not necessarily to match individual human reasoning but rather to allow exposure of patterns and similarities in place description topics that are otherwise opaque to an individual user (or human participant).

4.4 Automatic topic weighting evaluation

Despite this lack of correspondence, we can use the participant studies to evaluate the degree to which it is valid to assume a user has a set of 'interests' (i.e., a salient subset of topics) that are common across different place comparisons and whether we can identify those interests in terms of topics automatically from one or more sample rankings. For example, looking back at Table 5, these two participants consider very different properties when comparing the cities for similarity. Note, however that these participant-provided properties do not necessarily reflect all the properties used by the participant to discriminate between places – e.g., it does not explain the particular ordering of Austin, Portland, and Minneapolis-St. Paul by participant 2, because the same property ("music and performing arts") is given for all of them.

In order to do this evaluation we test whether on average automatic topic re-weighting using the techniques in the previous section result in smaller Spearman's footrule distances for the other rankings by the same participant. In other words, we evaluate the degree to which a single ranking can be extrapolated more generally to other rankings of cities. One sample ranking from the participant is used to calculate weights on the topics and new city rankings are calculated using the weighted JS divergence for each of the other city comparisons done by the same participant. For each ranking the Spearman's footrule distance is calculated between the new weighted ranking and the participant-provided ranking. This process is then repeated for each sample ranking by exchanging which one sets the weights. From this cross-validation an average footrule distance can be calculated for the weighted ranking for a participant, and this can be further averaged over all participants. Figure 1 shows this average footrule distance over all participants when using the weighted JS divergence B technique. The weighted JS divergence A has a similar result, though the overall average footrule distance tends to be slightly higher.

The average Spearman's footrule distance shows a marked decrease over the unweighted results for Wikipedia and travel blogs that were shown in Figure 1, which indicates that this method does help to identify general topics of interest for a person. One very interesting result is that the matching to the participant rankings is distinctly *better* when the participant is *not* very familiar with the source city being ranked against. One explanation for this result is that people who are very familiar with a place will have very specific impressions of the place and thus properties that are salient for that place do not transfer over from other places. Whereas when people are not familiar with a place there is a stock set of "background interests" that they use to help compare places. This indicates that automatic topic weighting will be particularly useful when done as part of a system designed to enable users to learn and explore knowledge about geographic places with which they are not already very familiar.

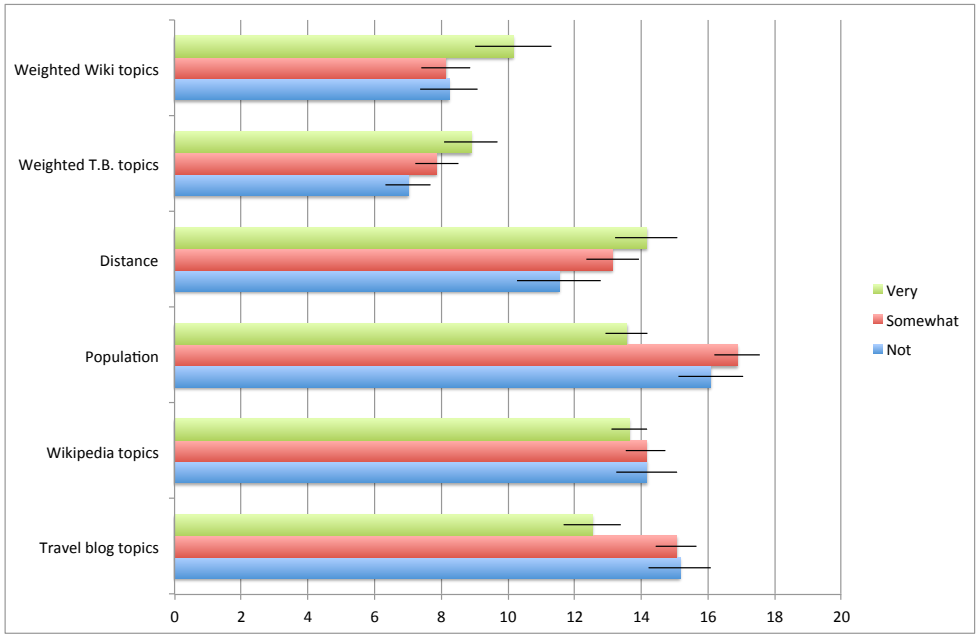


Figure 1: This chart shows the average Spearman's footrule distance between participants' rankings of U.S. city similarities and rankings based on geographic distance, population, Wikipedia topic mixtures, travel blog topic mixtures, and weighted Wikipedia and travel blog topics using weighted JS divergence B. The x-axis is the average Spearman's foot rule distance, indicating total displacement in the orderings of places being compared (Kumar and Vassilvitskii, 2010). Shorter bars indicate overall higher concordance between the participants' similar place rankings and the automated similar place rankings. Each type of automated ranking is split into three categories based on whether the participant is "very" (green), "somewhat" (red), or "not" (blue) familiar with the source place being compared with other places (based on self-report). The statistical significance bar shows $\alpha = 0.05$.

4.5 Discussion

From these results, the automatic weighting of topics based on a sample ranking can legitimately be generalised to other rankings. An interesting alternative would be to extract interests of a user implicitly from social network or other data and apply those to topic weights, but that remains beyond the scope of the current research (Kelly and Teevan, 2003; Ricci et al., 2011). On a case-by-case basis we also examined whether the reasons participants give for similarity correspond to high shared values on specific LDA topics, but apart from anecdotal evidence, it was difficult to do a quantitative evaluation on these data, because the mapping of participant-provided reasons and topics is highly subjective.

Table 6 shows the topics in Wikipedia (from 1200 topics) and travel blogs (from 1500 topics) for *Los Angeles* that increase the most (by ratio) based on re-weightings from the two sample participant rankings shown in Table 5. What is remarkable about the topics shown in Table 6 is that despite the apparent difference between the automatically-determined most-salient topics and the participant-provided reasons, the automatic weighting does significantly increase the concordance between the system rankings and the participant rankings. One explanation is that the topics that are weighted with high salience represent background properties that factor in the participants' conceptualisations of city categories but which are not at the forefront of their conscious comparisons of individual cities.

5 Conclusion

People judge the similarity or difference of places based on different contextual factors, such as their personal interests. Geographic information systems that take advantage of these factors and can provide places that are similar to places known to a user have many potential applications, including travel recommendation services, marketing analysis tools, and socio-ecological research tools. In this paper we presented a new method to automatically identify the topics that are salient to a user when performing similarity judgment. Topics can be any set of features associated with a place, such that a place is represented as a probability vector of topic values. These topic values form a topic signature that is generally associated with a place, and we demonstrated how probabilistic topic modelling can be used to generate such probability vectors for places.

Similar to Los Angeles ranking 1 – Travel blog	
Topic 827	game,basebal,team,play,watch,sport,hockey,fan,stadium,player
Topic 1242	restaur,order,food,tabl,meal,menu,eat,waiter,serv,dinner
Topic 904	money,pay,cost,expens,price,onli,cheap,pound,buy,free
Similar to Los Angeles ranking 1 – Wikipedia	
Topic 761	book,work,publish,life,wrote,histori,year,mani,author,writer
Topic 834	cathol,bishop,dioces,roman_cathol,roman,john,father,priest
Topic 890	golf,cour,club,golf_club,hole,countri_club,countri,locat,golf_cours
Similar to Los Angeles ranking 2 – Travel blog	
Topic 909	mile,road,stop,highway,gas,motel,state,sign,rout,trip
Topic 793	extrem,complet,entir,exact,time,veri,actual,howev,made,ani
Topic 1228	histor,site,build,visit,histori,tour,histor_site,area,museum,mani
Similar to Los Angeles ranking 2 – Wikipedia	
Topic 417	side,east,east_side,west,north,south,locat,north_side,west_side
Topic 164	polic,offic,polic_offic,polic_station,polic_depart,depart,enforc,forc
Topic 847	turkish,ottoman,turkey,greek,byzantin,ottoman_empir,sultan

Table 6: Topics that increase most in salience for Los Angeles based on two different participants' rankings.

Similarity calculations based on probability distributions are commonly context-neutral and only measure the relative entropy or JS divergence of the distributions. We presented a novel approach to use a small, user-provided sample ranking of similar places to automatically re-weight topic weights. This new re-weighting can be used to serve up personalised similarity values between places based on the topics that are of most interest to a user. Since this re-weighting of topic values is done after the initial training used to discover topics associated with places, it is fully compatible with a variety of different methods to create semantic signatures associated with a place, not solely topic modelling as used here.

We evaluated our method with a Mechanical Turk user study and showed that using a small sample ranking of similar places (i.e., a set of control similarities) results in a larger correspondence between automated place similarity rankings and personal users' rankings. Personalised place recommendation and similarity search remains a relatively unexplored research area. A follow-up study to better understand user motivation in performing place similarity will be a valuable next step. Future work will also involve exploring social media data and other sources such as search history to automatically provide sample similar places for a given user.

References

Adams, B. and K. Janowicz (2012). On the geo-indicativeness of non-georeferenced text. In J. G. Breslin, N. B. Ellison, J. G. Shanahan, and Z. Tufekci (Eds.), *ICWSM*, pp. 375–378. The AAAI Press.

Adams, B. and G. McKenzie (2013). Inferring thematic places from spatially referenced natural language descriptions. In D. Sui, S. Elwood, and M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge*, pp. 201–221. Springer Netherlands.

Banerjee, S. S. and R. R. Dholakia (2008). Mobile advertising: does location-based advertising work? *International Journal of Mobile Marketing 3*(2), 68–75.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Bodenhamer, D. J. (2010). The potential of spatial humanities. In D. J. Bodenhamer, J. Corrigan, and T. M. Harris (Eds.), *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, pp. 14–30. Indiana University Press.

Bouidghaghen, O., L. Tamine, and M. Boughanem (2011). Personalizing mobile web search for location sensitive queries. In A. B. Zaslavsky, P. K. Chrysanthis, D. L. Lee, D. Chakraborty, V. Kalogeraki, M. F. Mokbel, and C.-Y. Chow (Eds.), *Mobile Data Management (1)*, pp. 110–118. IEEE.

Daoud, M., L. Tamine, M. Boughanem, and B. Chebaro (2007). Learning implicit user interests using ontology and search history for personalization. In *Proceedings of the 2007 international conference on Web information systems engineering*, WISE'07, Berlin, Heidelberg, pp. 325–336. Springer-Verlag.

Diaconis, P. and R. L. Graham (1977). Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(2), 262–268.

Eisenstein, J., B. O'Connor, N. A. Smith, and E. P. Xing (2010). A latent variable model for geographic lexical variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, Stroudsburg, PA, USA, pp. 1277–1287. Association for Computational Linguistics.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. A Bradford Book. MIT Press.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences 101*(Suppl. 1), 5228–5235.

Grossner, K. E., M. F. Goodchild, and K. C. Clarke (2008). Defining a digital earth system. *Transactions in GIS 12*(1), 145–160.

Guy, I., M. Jacovi, A. Perer, I. Ronen, and E. Uziel (2010). Same places, same things, same people?: mining user similarity on social media. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, New York, NY, USA, pp. 41–50. ACM.

Hao, Q., R. Cai, C. Wang, R. Xiao, J.-M. Yang, Y. Pang, and L. Z. 0001 (2010). Equip tourists with knowledge mined from travelogues. In M. Rappa, P. Jones, J. Freire, and S. Chakrabarti (Eds.), *WWW*, pp. 401–410. ACM.

Hong, L., A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsiouliklis (2012). Discovering geographical topics in the twitter stream. In A. Mille, F. L. Gandon, J. Misselis, M. Rabinovich, and S. Staab (Eds.), *WWW*, pp. 769–778. ACM.

Janowicz, K., B. Adams, and M. Raubal (2010). Semantic referencing - determining context weights for similarity measurement. In S. I. Fabrikant, T. Reichenbacher, M. J. van Kreveld, and C. Schlieder (Eds.), *GIScience*, Volume 6292 of *Lecture Notes in Computer Science*, pp. 70–84. Springer.

Janowicz, K., M. Raubal, and W. Kuhn (2011). The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science* (2), 29–57.

Jones, C. B. and R. S. Purves (2008). Geographical information retrieval. *International Journal of Geographical Information Science 22*(3), 219–228.

Jordan, T., M. Raubal, B. Gartrell, and M. J. Egenhofer (1998). An affordance-based model of place in GIS. In T. Poiker and N. Chrisman (Eds.), *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)*, Vancouver, Canada, pp. 98–109.

Jorgensen, B. S. and R. C. Stedman (2006). A comparative analysis of predictors of sense of place dimensions: Attachment to, dependence on, and identification with lakeshore properties. *Journal of Environmental Management 79*, 316–327.

Kaltenborn, B. P. (1998). Effects of sense of place on responses to environmental impacts: A study among residents in svalbard in the Norwegian high Arctic. *Applied Geography 2*(18), 169–189.

Kelly, D. and J. Teevan (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum 37*, 18–28.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika 30*(1/2), pp. 81–93.

Keßler, C. (2012). What is the difference? a cognitive dissimilarity measure for information retrieval result sets. *Knowl. Inf. Syst. 30*(2), 319–340.

Kittur, A., E. H. Chi, and B. Suh (2008). Crowdsourcing user studies with mechanical turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 453–456. ACM.

Kumar, R. and S. Vassilvitskii (2010). Generalized distances between rankings. In M. Rappa, P. Jones, J. Freire, and S. Chakrabarti (Eds.), *WWW*, pp. 571–580. ACM.

Larson, R. (1996). Geographic information retrieval and spatial browsing. In L. C. Smith and M. Gluck (Eds.), *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information*, pp. 81–124.

McKenzie, G., B. Adams, and K. Janowicz (2013). A thematic approach to user similarity built on geosocial check-ins. In D. Vandenbroucke, B. Bucher, and J. Crompvoets (Eds.), *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography, pp. 39–53. Springer International Publishing.

Medin, D. L., R. L. Goldstone, and D. Gentner (1993). Respects for similarity. *Psychological Review 100*(2), 254–278.

Micarelli, A., F. Gasparetti, F. Sciarrone, and S. Gauch (2007). The adaptive web. Chapter Personalized search on the world wide web, pp. 195–230. Berlin, Heidelberg: Springer-Verlag.

Moore, R. L. and A. R. Graefe (1994). Attachments to recreation settings: The case of rail-trail users. *Leisure Sciences 16*, 17–31.

Pitkow, J., H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel (2002). Personalized search. *Commun. ACM 45*(9), 50–55.

Ranganathan, A. and R. H. Campbell (2002). Advertising in a pervasive computing environment. In M. S. Viveros, H. Lei, and O. Wolfson (Eds.), *Workshop Mobile Commerce*, pp. 10–14. ACM.

Raubal, M. (2004). Formalizing conceptual spaces. In A. C. Varzi and L. Vieu (Eds.), *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004)*, pp. 153–164. IOS Press.

Relph, E. (1976). *Place and Placelessness*. Pion.

Ricci, F., L. Rokach, and B. Shapira (2011). Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor (Eds.), *Recommender Systems Handbook*, pp. 1–35. Springer.

Rodríguez, M. A. and M. J. Egenhofer (2004). Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science 18*(3), 229–256.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization*, pp. 27–48. Hillsdale (NJ), USA: Lawrence Erlbaum Associates.

Sizov, S. (2010). Geofolk: latent spatial semantics in web 2.0 social media. In B. D. Davison, T. Suel, N. Craswell, and B. Liu (Eds.), *WSDM*, pp. 281–290. ACM.

Speretta, M. and S. Gauch (2005). Personalized search based on user search histories. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pp. 622–628.

Stedman, R. C. (2002). Toward a social psychology of place : Predicting behavior from place-based cognitions, attitude, and identity. *Environment and Behavior 34*(5), 561–581.

Steyvers, M. and T. Griffiths (2007). Probabilistic topic models. In T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates.

Teevan, J., S. T. Dumais, and E. Horvitz (2005). Personalizing search via automated analysis of interests and activities. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 449–456. ACM.

Tuan, Y.-F. (1977). *Space and Place: the Perspective of Experience*. The Regents of the University of Minnesota.

Tversky, A. (1977). Features of similarity. *Psychological Review 84*(4), 327–352.

Vasardani, M., S. Winter, and K.-F. Richter (2013). Locating place names from place descriptions. *International Journal of Geographical Information Science 27*(12), 2509–2532.

Wang, C., J. Wang, X. Xie, and W.-Y. Ma (2007). Mining geographic knowledge using location aware topic model. In R. Purves and C. Jones (Eds.), *GIR*, pp. 65–70. ACM.

Winter, S., W. Kuhn, and A. Krüger (2009). Guest editorial: Does place have a place in geographic information science? *Spatial Cognition and Computation 9*, 171–173.

Yin, Z., L. Cao, J. Han, C. Zhai, and T. S. Huang (2011). Geographical topic discovery and comparison. In S. Srinivasan, K. Ramamritham, A. Kumar, M. P. Ravindra, E. Bertino, and R. Kumar (Eds.), *WWW*, pp. 247–256. ACM.

Urban Data Hubs Supporting Smart Cities

Phillip Delaney
The Australian Urban Research Infrastructure Network (AURIN)
University of Melbourne, VIC, 3010 Australia
phillipd@unimelb.edu.au

Chris Pettit
The Australian Urban Research Infrastructure Network (AURIN)
University of Melbourne, VIC, 3010 Australia
cpettit@unimelb.edu.au

Abstract

Discovering and accessing digital data, which is predominantly spatial, is a problem often faced by researchers and policy- and decision-makers in Australia. Significant efforts are underway to provide access to discovery mechanisms to large data repositories, such as the INSPIRE geoportal in the European Union and the Research Data Australia portal developed by the Australian National Data Services (ANDS) in Australia. However, whilst such portals are significant vehicles to data discovery, they typically do not provide direct access to the data assets themselves. In this paper we review the state of play of 'data hubs' where data can be searched, discovered, downloaded and in some cases analysed and visualised. Data hubs are typically web services or data download playgrounds accessible via a portal. In recent times there has been a push for government agencies to open data access to support research and development and innovation in industry. In this paper we will focus on reviewing a number of data hubs initiatives across Australia and internationally in relation to data relevant to enabling smart cities. We will provide specific attention to the Australian Urban Research Infrastructure Network (AURIN), which is developing a portal infrastructure focused on supporting urban researchers, and policy and decision-makers. The AURIN portal aims to facilitate programmatic access to data held in many emerging data hubs across Australia. AURIN is implementing a federated data approach, providing a single access point and common interface for interrogating and visualising datasets. This paper outlines the data hub concept, describing the process and benefits of data hub integration within the AURIN e-infrastructure context, and critically examines this approach and other similar approaches being undertaken by comparable initiatives around the world.

1 Introduction

In this paper we will introduce the concept of data hubs as a mechanism to provide better access to data for urban researchers, policy and decision-makers in Australia. With the advent of the digital city, also referred to as the ubiquitous or smart city, there is a growing need for data to be more accessible and to better support evidenced based decision-making (Batty, 2013). An ACIL Tasman report in 2008 found that there are "few, if any, sectors of the economy that have not begun to use modern spatial information technology", and as such data with spatial information will be integral to this decision making process. This report also identifies that inefficient access to spatial data in Australia has significant productivity impacts in many sectors, including planning and policy development for urban areas. Roche Et. Al. (2012) connects these ideas by identifying key opportunities for the smart city and spatially information communities to collaborate to provide the best outcome for cities.

The Australian Urban Research Infrastructure Network (AURIN, http://aurin.org.au/) is a project funded by the Australian Government's Super Science initiative, tasked with building an e-infrastructure oriented to the needs of Australia's urban researchers. AURIN aims to assist in improving connections between researchers and data custodians across Australia, and to "offer open access to data arising from research infrastructure provided through the AURIN" (AURIN, 2011). As stated Data hubs are considered an integral part of the AURIN e-infrastructure. In this paper we will discuss how AURIN is establishing a number of data hubs in a federated architecture to provide better access to Australia's growing digital urban data asset, and will discuss how these hubs compare to other initiatives around the world, particularly the Infrastructure for Spatial Information in the

In: S. Winter and C. Rizos (eds.): Research@Locate14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

European Community (INSPIRE) project. The Melbourne Data Hub is presented as a case study as the first data hub realised through the AURIN project, and the only hub project fully completed and integrated within the AURIN e-infrastructure. The paper concludes by discussing the next steps in implementing a series of federated urban data hubs across Australia, linked via the AURIN e-infrastructure, and discusses some of the challenges and opportunities in doing so.

2 The Emergence of Data Hubs

Data hubs have been created to address issues arising from the discovery, access, and format diversity of research data. The Open Knowledge Foundation (OKF) defines a data hub as 'a community-run catalogue of useful sets of data on the Internet', and notes that in addition to data searching, the data hub 'may also be able to store a copy of the data or host it in a database, and provide some basic visualisation tools' (OKF 2013B). From a technical perspective, a data hub can be deployed using a 'Hub-and-Spoke' architecture, where data is housed in a central system, or hub, with users accessing the data from many locations, or spokes. While these definitions are useful, they also limit the potential application of data hubs. For example, a data hub can benefit the data user community (community), but the stored and operated by a private company or government institution. It is important to note that the data hub concept does not just apply to free or open data, but to data in general, including data which requires a fee for access. As such, the data hub concept can be seen to focus more on making data accessible rather than making data free.

As such, the AURIN project has broadened the definition of a data hub to represent the scope and variety of applications of the data hub concept. At the core a data hub needs to be a single access point that:

1. Allows users to search for a variety of data;
2. Allows users to access and use this data;
3. Provides data in a discrete set of formats;
4. Creates a discussion/feedback loop between custodians and users;
5. Allows users/data custodians to contribute data to the hub; and
6. Provides information about the data (metadata).

Based on the above definition (Delaney, 2013), traditional data warehouses, data marts and geoportal technology align with many elements of a data hub, and have been designed to allow users to discover, access and share data. However, the collaboration and interaction between the community and data owners is what differentiates the AURIN data hub concept from the data warehouse and geoportal concepts. This collaboration has some direct benefits for data custodians, and clearly addresses the issues of data discovery and access identified in ACIL Tasman (2008), allowing for a much more collaborative research environment. In addition, the data hub reduces duplication of data storage and creation by providing a single online access point for data. Finally, allowing data to be consumed online from the data hub allows users and developers to develop customised products and services using these datasets (Guiliani and Peduzzi, 2011).

One of the main limitations of the data hub is that any downtime of the data infrastructure results in users losing access to data, unless redundancy is factored into the system. This is an inherent risk in the consolidated data hub approach, one shared with data warehouses and geoportals (Engström et al. 2000). However, as Yang (2010) highlights, continuous developments in the configuration of these systems have resulted in improved management methods, improved distribution and minimised impacts to users. In addition, distributing data hubs across research domains and geographies distributes this risk, where downtime for one hub does not mean reduced usage of all available research data.

The primary focus for AURIN has been to access data with a spatial component and provide urban researchers with a system to discover and analyse this data, which aligns closely with the geoportal concept. Tait (2005) describes a geoportal as 'a web site considered to be an entry point to geographic content on the web or, more simply, a web site where geographic content can be discovered'. These geoportals evolved from ongoing developments in Spatial Data Infrastructure (SDI), which were created by various national mapping agencies, like United States Geological Survey (USGS), to manage their spatial information (Williamson et al., 2003). SDI were initially developed to manage and distribute data internally within related organisations, but have matured into a technology infrastructure for managing and distributing spatial information to various users and applications.

The data hub as a concept has many similarities to the concept of geoportals, but with the increasing focus on big data and open data distribution, data hub has been applied to many disparate fields and many types of data. Large technology organisations such as Microsoft have invested heavily in developing enterprise level solutions

to the data hub concept, with specific application development using SQL server to develop a 'hub and spoke architecture' solution to data storage and distribution (Theissen and Kraemer, 2009). For example, this solution has been deployed in CROSSMARK, a company that distributes sales and marketing data to customers across the world through a SQL based portal using the hub and spoke model to distribute on demand access to complex data (Redmond, 2012).

The advent of the data.gov open data movement in 2008, discussed in detail in Section 3.2, resulted in increasing volumes of data available and distributed to the public, and cloud computing allowed a new way to think about this storage and distribution (Yang et al., 2010). The variety of options now available to store and distribute data led to the application of the data hub concept in several large-scale projects across the globe, discussed further in Section 3. A brief timeline of some key contributors to the development of the data hub concept can be found in Figure 1, along with some of the future data hub initiatives currently in planning across Australia.

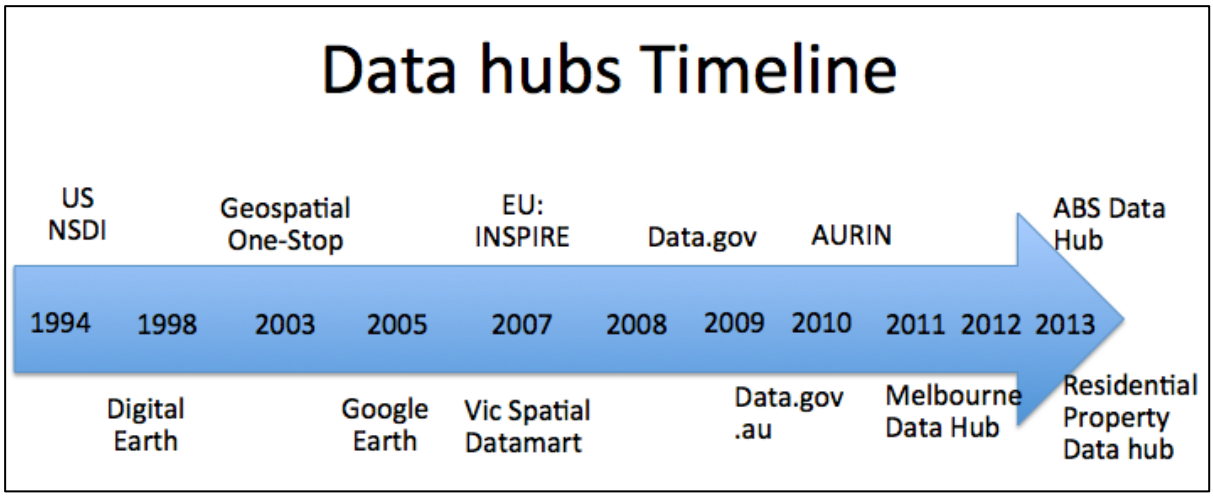


Figure 1 – Evolution of the data hubs (modified from Yang et al., 2010), United States National Spatial Data Infrastructure (US NSDI), the European Union Spatial Information in the European Community (EU INSPIRE) directive, the Victorian Spatial Data Mart, Australian Bureau of Statistics Data Hub (ABS)

3 Data Hubs and International Initiatives

The data hub concept has been realised in many locations and contexts globally. Many scientific fields have collaborated to create research specific data hubs to store, discover and distribute research data to other researchers. Examples of these hubs can be found in the fields of health, environmental and engineering research, for example the "ObamaCare" online heath portal manages data through a Federal Services data hub, drawing information from distributed federal government sources such as the IRS, Social Security Administration and Homeland Security, and is used as a distributed, secure means to verify identity (Vijayan, 2013). The Tropical Data Hub, set up through the Queensland Cyberinfrastructure Foundation (QCIF), has been established to allow researches to collate, distribute and discover research data from tropical research projects being undertaken across the globe (Myers et. al. 2012).

In the urban research community, the European Union (EU) has established the Urban Audit Data Hub as a collection of comparable city statistics across the EU (Manninen, 2008), and the Buildings Performance Institute Europe Data Hub (BPIE, 2012) to specifically collate open data on building stock and building policy. While there is an apparent paucity in specific urban data hubs, many federal and local government agencies have geoportals for the display and interrogation of urban data. The INSPIRE project provides the largest scale application of this open geoportal environment, and will be discussed below. In addition, much data relevant to the urban environment is being distributed through broad application data hubs which support multiple sectors, for example the data distributed through the data.gov platforms discussed in Section 3.2. The Canadian Geospatial Data Infrastructure provides another example of a wide scale geoportal, though has not been discussed in detail in this paper as the limitations are similar to those in INSPIRE.

Broad application data hubs are much more prevalent than research domain specific hubs, and have been set up to create a community of general knowledge sharing. They have the advantage of not being constrained by a particular type of dataset or field of interest, but can be more difficult to search, and contain less detailed information than the field specific hubs.

Initiatives such as www.datahub.io have been launched by the OKF as a free data management and distribution system to allow any user or organization to host and distribute data. In addition OKF has released CKAN, an open source software solution for data publishers to use to create data hubs, making their data accessible for the public. There are also many examples where the private sector has developed technical platforms to support the establishment of new data hubs, including the Oracle Life Sciences Data Hub (Oracle, 2013), the MongoDB Data Hub (MongoDB, 2014)) and the Socrata Open Data Portal (Socrata, 2013).

3.1. INSPIRE

Framework

One of the largest geoportal initiatives underway anywhere is the world is the Infrastructure for Spatial Information in the European Community (INSPIRE) directive. This directive aims to create a standard EU Spatial Data Infrastructure, which will improve the management of spatial data and metadata, data interoperability and data sharing between the 27 member states of the EU, across 34 identified spatial themes (INSPIRE, 2007). The INSPIRE project represents a significant investment from all member states, and has resulted in close to 300,000 spatial datasets being made available to the community through a standardised data discovery site.

The main INSPIRE portal allows users to search for datasets from across the EU from a single interface, and allows advanced search filters to be used to narrow down searches by geography, format or spatial theme. The INSPIRE portal only displays metadata for each dataset, it does not allow users to directly access any of the datasets, either manually or programmatically. However, each metadata resource contains a link to the data source, which may be a file, service or web application.

Limitations

While INSPIRE represents a significant achievement in spatial data standardisation and management, there are still significant steps required to improve usability through effective user interface design and improved data access (Larson Et. Al, 2006). In support of this, preliminary evaluation from the research team in testing INSPIRE has found that it is hard for a user to assess or use datasets without the ability to preview or explore the data itself, or access the datasets for analysis purposes through a single interface.

In the case of data being made available through web applications, the link provided through the INSPIRE metadata is often to a web mapping interface where a user has to go through the process of discovering the dataset again from a new search window as the link often does not go to the individual dataset identified in the metadata. As a significant portion of the datasets in INSPIRE are orthophotography tiles or georeferenced cadastral plans (~40-50%), which are generally accessed through a separate portal, this process can be time consuming and often a duplication of effort.

In addition, the number of datasets listed in this portal can be misleading. For example, an orthophotography series over a town represents one data collection. However, technically the data is too large to hold as a single file, and may be divided in to hundreds or thousands of tiles. In INSPIRE, each of these tiles can be listed individually as a dataset, where in reality they should be listed as one. This also applies to series of scanned cadastral plans which represent a single collection of data, not many individual datasets. It may be more efficient from a user perspective in these cases for the resource to be linked as a collection, not as individual file records, which has been the approach of the Australian National Data Service (ANDS, 2014) and their Research Data Australia (RDA) metadata portal (Treloar, 2008).

3.2. Open Government Initiatives

Framework

Open data, tools and software have become increasingly popular and accessible over the past decade. With this popularity, there has been increasing pressure on government organisations to release their data openly as well.

The principle of open access aims to support public policy developments, and provide a support mechanism for citizens to examine information generated by their governments (Janssen, 2011).

As discussed in Section 3, OKF has released an open source data hub platform called CKAN to encourage organisations to distribute datasets to the public. Many local and national government agencies across the world have taken advantage of this open data hub technology, including Berlin, United States of America, Canada, United Kingdom, Germany, Mexico and many more, including the Australian Commonwealth Government, and the State Government of New South Wales, Queensland and South Australia (OKF, 2013A). These data hubs are examplars of Open Government Data Initiatives around the world, notably www.data.gov for the United States of America and www.data.gov.uk, which were both launched in 2009 and were the first two governments to adopt the data.gov framework. Within this initiative, governments aspire to publishing unrefined or raw public datasets in an open, non-proprietary technical format, licensed for use, re-use and re-distribution at marginal or no cost (Bates, 2012). Many more government organisations around the world use similar technology to achieve the same aim, the end result of which exposes raw data generated by organisations to the public for download, use and analysis.

Limitations

Data hubs established using these platforms conform to each of the data hub characteristics as defined in Section 2, though there are some caveats around the enabling of community or public participation. While many open data platforms contain a dedicated 'Suggest/request a dataset' section, which allows for the community to request specific datasets (see https://explore.data.gov/nominate), there are many examples of open data sites where this has not been implemented. A preliminary search of CKAN instances (OKF, 2013A) shows that the open data sites for Africa, Germany and Italy do not provide a mechanism for the community to request data to be made available, limiting some of the collaborative potential for each of these hubs.

Another limitation of the government open data hubs is that data can only be requested as whole datasets for download, subsets of data cannot be extracted, queried or imported in to other applications. A review of Open Government Data identified that this data often required "substantial human workload to clean them up for machine processing and to make them comprehensible" (Ding Et. Al, 2011). The consequence of this is that while the data has been made 'open', use and analysis of this data by the community is somewhat restricted, especially for those that do not have access to specialist software and analysis tools and skills, such as the ability to use Geographic Information Systems (GIS) software. This is particularly a problem for users who want to access spatial data files, such as shapefiles, but do not have the knowledge or software required to analyse, and visualise this information. This also limits the ability of developers to programmatically access this data to provide it in simple visualisation and analysis applications for the public to better understand and use this data.

4 Data Hubs in Australia

In July 2010 the Australian government released a Declaration of Open Government to promote an open government based on 'better access to and use of government held information, and sustained by the innovative use of technology' (Australia, 2010). One of the primary benefits behind the open government initiative is the broad scale release of government information, including many data sets held by both Commonwealth and State Government bodies. To distribute this information, the Australian Commonwealth Department of Finance and Deregulation has established http://data.gov.au/, a broad scale data hub for discovery and access to government data based on the OKF CKAN platform (OKF, 2013A). This site distributes more than 3000 datasets from over 120 different contributing government organisations (Australia, 2014). State Governments in New South Wales, Queensland, South Australia, Victoria and the Australian Capital Territory have also released similar open data policies, and are all on the way to providing searchable data.gov.au data hub sites (Waugh, 2013). In Victoria, this policy has also led to the release of many GIS datasets through both direct download and through a machine-to-machine data hub hosted by the Department of Environment and Primary Industries.

Other notable data hubs which have been established within Australia include the University of Wollongong Smart Dashboard (Wickramasuriya, 2013) and Western Australia's Landgate SLIP and SLIP Future projects (WALIS, 2012). It should be noted that these two hubs are not part of Open Government Initiatives, but are fee services run.

Limitations

Like the other data hubs from within the data.gov discussion, the open data hubs detailed above only link to whole datasets, where subsets of large datasets cannot be requested, and programmatic integration of these datasets becomes difficult. The Smart Dashboard and SLIP hubs overcome the dataset subset problems, and allows for the data to be programmatically accessed, machine-to-machine(Wickramasuriya, 2013). However, for the moment these hubs are not completely openly accessible, and require a fee to access some components.

An initiative by the Australia and New Zealand Land Information Council (ANZLIC) is currently looking at addressing the broader accessibility of spatial data through its Foundation Spatial Framework. This framework has the goal of "making common foundation spatial data ubiquitous across Australia and New Zealand" (ANZLIC, 2012). This initiative may address the access and usage limitations listed above, though the framework is still in development stages.

5 AURIN Data Hubs

AURIN has been funded to establish a dedicated urban data portal in Australia, and AURIN aims to use the data hub concept to facilitate machine-to-machine access to datasets across the country held by key national, state and local government agencies, and private sector organisations. AURIN leverages both the data hub hosting and distribution features to access data at the source, and consume the data within the AURIN portal (Delaney, 2013). The AURIN portal will also allow users to contribute their research outcomes back into the portal for other researchers to discover and use for their research purposes, aligning with the collaborative aims of the data hub concept (. This ensures that the AURIN portal is accessing the most up to date information within the data hub, allowing users to have increased confidence in the data from the portal. AURIN is co-funding the development of several data hubs across Australia to support urban research. While the definition of a data hub can clearly be quite broad, for AURIN purposes data hubs need to align with the following principals:

1. **A Data Hub facilitates collaboration and interaction between end users and data custodians.**

A data hub represents an opportunity for data custodians and data users to work together to determine which data is important for release to the user community. This represents a paradigm shift from the data warehousing environments where data custodians made all the decisions regarding which data was made available to users. This benefits both groups in this equation: Data custodians can minimise resources spent releasing data by focusing on what is identified as having the highest demand/need, as well as receiving feedback on the quality and usability of datasets. Data users can then receive more of the data they require, and reduce the time taken negotiating access to datasets held within organisations.

Viewed with these goals, a data hub is as much a hub of people as one of data, though a facilitator is often required at the centre to coordinate communication between parties, prioritise data needs and negotiate for access to datasets. The dynamics of relationships are captured in the communication cycles illustrated in Figure 2. AURIN acts as this facilitator within the Australian urban research environment.

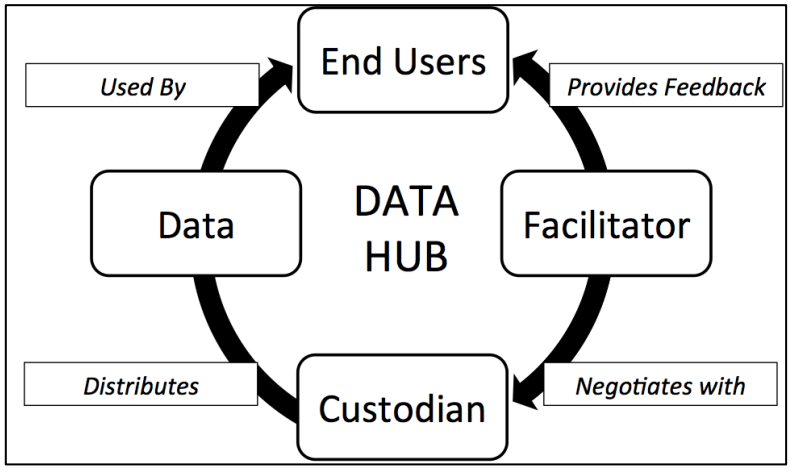


Figure 2 – Data hub conceptual communication feedback cycle

2. Data should be held as close to the source as possible.

End user confidence in the quality of data will increase when data is hosted as close to the creators or custodians of data as technically possible. Where possible, data custodians should be either distributing data through their own hub, or contributing data directly to a third party hub if the custodian does not have sufficient technical resources to set up a programmatic in-house data service. Understandably, not all data custodians have sufficient resources available to enable this programmatic access, and as such AURIN is working to establish data hubs across Australia, focusing on either a specific discipline or focussing on a geography such as a state.

3. Data Hubs should be set up to serve a broad end-user community, not a single project.

The hub and spoke model of data hubs illustrated in Figure 3 shows a hub having multiple end users, and as such a data hub should be established to service long term needs of an end user community. While a facilitator, such as AURIN, may act as a catalyst to establish a new data hub, the resulting networks, technology and data should be set up such a way that it can continue to operate effectively without the original facilitator. This will ensure that users can have confidence in a data hub as a long-term method for accessing relevant data.

4. Sufficient information will be provided for users to understand data.

Research from Dawes et al. (2005) highlights that user understanding of data increases with increased understanding of how a dataset is created. Dawes further notes that in depth metadata allows a user to 'determine if it (the data) would be worth working with for their purposes'. As such, any data served through a data hub must be accompanied by sufficient information to allow a user to assess the suitability of a dataset for a given purpose, including any limitations, restrictions and caveats on any given dataset.

Supporting these principals, the following criteria are used to assess a data hub for suitability for integration in to the AURIN platform (Delaney, 2012). An AURIN data hub is one which:

- provides programmatic access (machine to machine access) to a number of data services.
- aligns to a federated data service model and supports data interoperability so that other portals web mapping tools can build off the hub.
- aligns to data and metadata standards driven approach where possible.
- has a consolidated level of technical operational support at the designed hub.
- encourages a collaboration and consortium approach
- realises economies of scale through multiple data feeds.
- leveraging existing data services infrastructure where possible;
- facilitates licensing arrangements with data custodians through relevant government department or agency.

AURIN aims to connect to many hubs, along the 'spokes', and make the data in these hubs available for discovery, analysis and download. In this way, the AURIN portal can be considered as a hub of hubs – a single access point allowing users to access and combine information from hubs in different geographic locations and from different specialist themes, a concept illustrated in Figure 2.

AURIN is currently exploring incorporating up to 12 data hubs across Australia. The first of these hubs implemented is the Melbourne Data Hub, explored in a case study below. Other data hub projects underway include national projects with the Australian Bureau of Statistics (ABS), and the Australian Property Monitors (APM), as well as state-based hubs looking at housing in NSW, planning and transport data in WA, transportation modelling in Brisbane, and an energy and water data hub in Townsville.

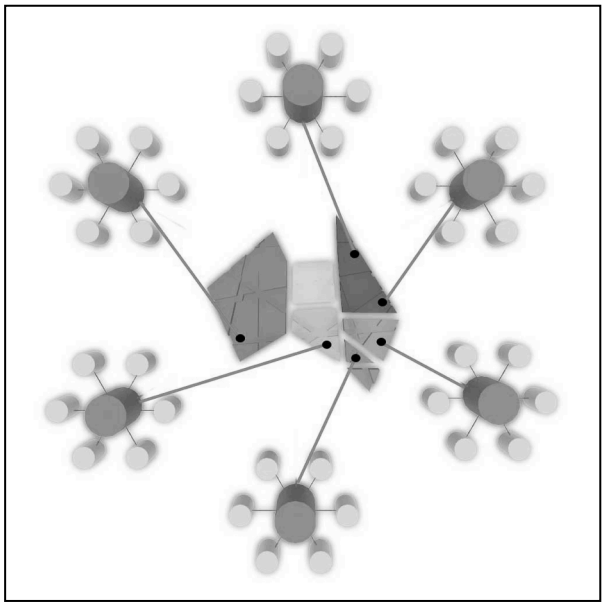


Figure 3 – The AURIN Data Hubs of hubs concept, highlighting geographic locations of select AURIN data hubs.

6 Data Portal Comparison and Discussion

To assess the differences between some of the major data hubs discussed in this paper, a table has been developed identifying some of the key characteristics and metrics of these. This table allows for a non-technical comparison to be made between each geoportal and data hub, including identifying the target user group and technical maturity of each of the stated portals. These examples have been chosen to allow comparisons between international initiatives and Australian examples. This is not an exhaustive list, and other initiatives such as the Canadian Geospatial Data Infrastructure would also provide useful comparisons, but from initial investigations would have yielded similar results. As such, these examples were chosen to provide a comparison of each distinct data hub initiative as discussed in this paper.

This table identifies some clear similarities and differences between these data initiatives. Each portal recognises the significance of metadata by ensuring all data contributed aligns to a minimum standard. However, the detail of this metadata varies within each portal. Both data.gov sites contain varied levels of metadata, from a full ISO compliant implementation to simple information with just a dataset title and abstract. INSPIRE and ANDS both have extensive metadata compliance required before data is published, with INSPIRE using the ISO19115 standard and ANDS recommending different metadata standards for different research data types. AURIN has implemented a simplified ISO 19115 standard, including attribute level metadata, and also a metadata entry for level of measurements (nominal, ordinal, interval, ratio).

Table 1 – Comparison table of major data access initiatives, INSPIRE in Europe, data.gov in the USA, and data.gov.au, ANDS RDA and AURIN data hubs from Australia

Criteria	INSPIRE	Data.gov	Data.gov.au	ANDS RDA	AURIN
Programmatic access to raw data	✖	✓	✓	✖	✓
Programmatic querying of data	✖	✖	✖	✖	✓
Standardised Metadata	✓	✓	✓	✓	✓
Data User/Custodian collaboration	✖	✓	✖	✓	✓
Open data access	✓	✓	✓	✓	Research Only

Facilitates Data Licensing	✓	✓	✓	✖	✓
Total Datasets	0	~55,000	~700	0	~500
Total Metadata Records	~300,000	~55,000	~700	~80,000	~500
Government Data Contributors	✓	✓	✓	✓	✓
Researcher Data Contributors	✖	✖	✖	✓	✓
Private Company Data Contributors	✖	✖	✖	✖	✓
Analytical toolset	✖	✖	✖	✖	✓
Technical Maturity	Production	Production	Production	Production	Beta
Intended User Group	Public	Public	Public	Researchers	Researchers, Policy/Decision Makers

Another clear difference is the number of datasets found in the portal. It has already been discussed in section 3.1 how these numbers may be misleading; in the case of INSPIRE, over 160,000 of the 300,000 metadata records are distributed from just two of the 27 countries. Data.gov clearly provides direct access to the largest set of data, however it should be noted that this is data access without associated analytical toolsets provided to assist in data understanding. The AURIN portal is the only initiative which has enabled data to be queried, analysed and visualised in addition to providing the data access. However, this portal has the fewest number of datasets, which is likely to be a result of the fact that this portal is still in beta development, and not released yet as a full production system as with each of the other initiatives.

All portals distribute data from Government agencies, but AURIN is the only portal to distribute data from the private industry as well. This is due to focused user market of the AURIN platform compared with the other portals. AURIN is designed to support academic researchers and policy/decision makers, not the general public, making the negotiation of private license agreements easier than would be the case for a public portal.

Table 1 highlights that the use of any data portal should be subject to the end-user needs. A user looking to understand data availability would be best served by the INSPIRE and ANDS RDA projects. A user wanting to access full data records should use the data.gov and data.gov.au sites with the understanding that the quality of metadata is inconsistent, and a user will need to undertake any data analysis using a third party application. Researchers in Australia can use the AURIN project to access, download, analyse, query and visualise data from a variety of urban research disciplines, with sufficient metadata to judge the suitability for purpose of each available dataset.

7 Melbourne Data Hub Case Study

Recognising the significant challenges that population growth will have on a city's liveability, the North-West Melbourne Regional Management Forum (NWM – RMF) has identified the need to work collaboratively across government and academia to develop an integrated spatial data platform to support research in the region (Eagleson 2012).

The North-West Melbourne Data Integration project included members from 14 Local Government Authorities as well as AURIN and ANDS, with the aim to use web enabled technology to connect computers, exchange data and undertake analysis. The key component of the NWM project was the ability to access and distribute spatial datasets to various project stakeholders. This was achieved these through the creation of the Melbourne Data Hub, which made an extensive range of health, housing, transport and planning datasets available both to contributing agencies and to urban research across Australia via the AURIN portal (Nasr and Keshtiarast, 2013).

The data hub was established and maintained by the Centre for Spatial Data Infrastructure and Land Administration (CSDILA), and consisted of two main components: a server to distribute the datasets and a tool to harvest and enrich metadata for each dataset. The data was distributed using GeoServer Web Feature Service (WFS), which is an open source spatial data server. CSDILA collaborated with several government agencies to collate urban data, then clean and geocode many of the supplied datasets which were not in a suitable format for distribution. These datasets were then imported to a Postgres/PostGIS database for distribution through the GeoServer. A second GeoServer was also used for the project, housed at the Department of Environment and Primary Industries (DEPI). This GeoServer was used to distribute data which was already held within the

Victorian Spatial Datamart (Nasr and Keshtiarast, 2013). This arrangement allowed data to be kept as close to the custodian source as possible, but also provided an alternate hosting mechanism for departments that lacked the required technical infrastructure to distribute data through their own hub.

To bring these WFS feeds in to the AURIN portal, CSDILA also created a metadata harvesting and enrichment tool using GeoNetwork as the basis, with custom metadata requirements built in to meet the AURIN specifications. This tool was used to capture and enhance title, dataset abstract and attribute abstract information for each of the datasets before being ingested in to AURIN. This allowed the data to be consumed by AURIN with fully compliant metadata, allowing users to understand each component of the datasets.

Common metadata standards are integral to the operation of the AURIN e-infrastructure, as the portal operates using a federated architecture. This allows data to be stored and distributed in a number of different geographic locations and file formats, but understood and integrated in a consistent manner in the AURIN portal.

The data hub was also used to serve various datasets to four demonstrator projects, which were used to demonstrate the value and utility of the Melbourne Data Hub in the urban research areas of health, housing, walkability and employment. The data hub was fundamental in consolidating and distributing datasets for various research application purposes. Each of the four demonstrator projects successfully used the data distributed from the data hub. At the project completion, the Melbourne Data Hub is distributing more than 120 datasets through the AURIN portal, from more than 10 contributing government agencies within Victoria, all with fully compliant AURIN metadata (Nasr and Keshtiarast, 2013). http://blogs.unimelb.edu.au/aurinands/

Setting up a data hub in this manner has also overcome the limitations encountered by the data.gov platforms, and the INSPIRE project, where subsets of data cannot be requested and access to data may be through separate data portals. The GeoServer allows for subsets of each exposed dataset to be extracted either by geography or by a tabular request, which cannot be completed in the data.gov sites. The data held in the hub itself was determined of significance by the urban research community and collaborating government agencies and local councils who had participated in the development of the hub. The project is also supported by the general AURIN "Request a Dataset" option, which allows the research community to identify future datasets for inclusion in the AURIN portal.

Unlike the INSPIRE project, the AURIN portal allows for a programmatic link directly to the data source, so no additional data searches are required. This also ensures that everything can be requested from a single source. Linked to the AURIN federated architecture, this allows data from other data hubs external to the Melbourne Data Hub to also be requested from a single portal. In addition to allowing access to data, the AURIN portal also allows user to visualise and analyse this data using '(i) graphs and charts, (ii) choropleth mapping, (iii) heat mapping, (iv) flow mapping, (v) brushing, (vi) space time cube (STC) representation and (vii) Decision Support Dashboards' (Pettit et. al., 2012).

Figure 4 illustrates a typical AURIN portal session. The data cart highlights many of the contributing agencies to the Melbourne Data Hub, along with data from other data hub projects within AURIN, all of which can be accessed, analysed and visualised in the AURIN portal. The data contains detailed metadata, including links to more information for researchers to assess data quality and lineage. The figure also demonstrates a choropleth map of population, overlayed with a pedestrian catchment map from the AURIN Walkability tool.

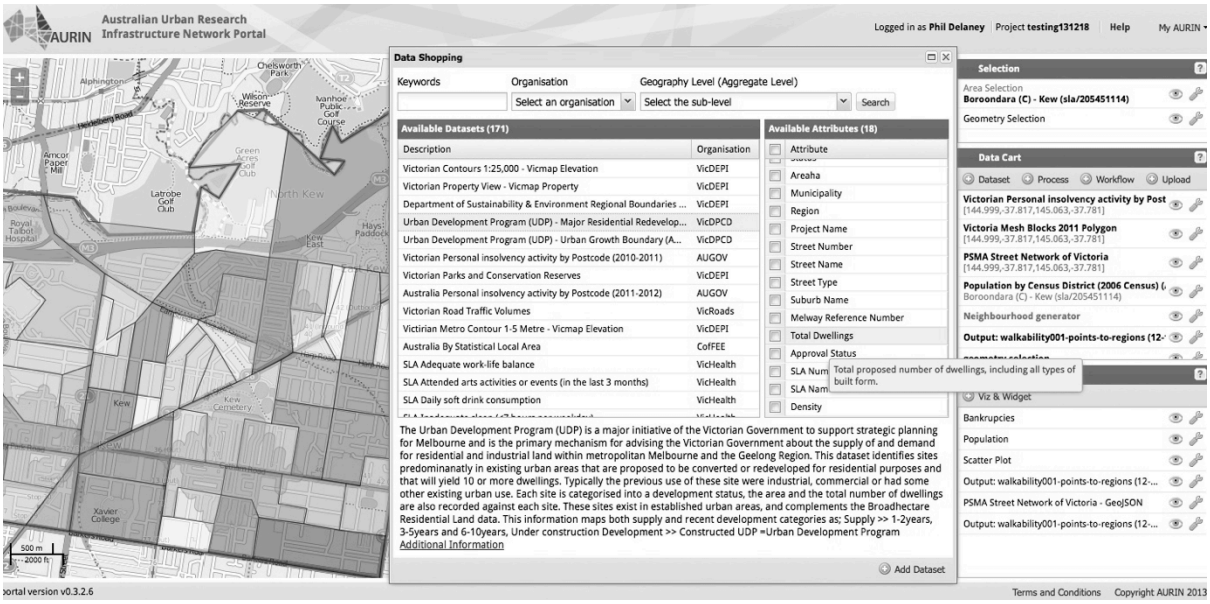


Figure 4 – Example user session of the AURIN portal showing data shopping and map visualisation capability, showing data from the Melbourne Data Hub. (source: https://portal.aurin.org.au, accessed 18 December 2013)

8 Conclusion

The application of the data hub concept in Australia increases the ease of data discovery and access for urban researcher and policy/decision makers from participating Government agencies. By building upon the concept of data hubs, AURIN is also allowing researchers to collaborate on multi-disciplinary research endeavours which cut across domains such as urban health, transport, housing, water and energy supply and consumption and innovative urban design. The data hub model is also connecting researchers with Government agencies in delivering evidenced-based research outcomes that address real-world problems facing cities including liveability, housing affordability, economic prosperity and population growth. The success of the Melbourne Data Hub illustrates the benefits that can be gained from a consolidated data hub across a number of research domains. The hub also overcame the access and usability limitations experienced by some other open data initiatives where data can not be directly accessed, or where subsets of data could not be extracted and analysed. As such AURIN is further pursuing the data hub concept as the means of enabling collaborative, innovative urban research to support better urban planning and design in Australia. This will be achieved through up to 12 data hub projects during the current funding cycle of the AURIN project. These hubs will provide access to public sector, private sector and research data in the support of urban researchers across Australia.

9 Challenges and Future Opportunities

Data hubs provide many opportunities for Australia, and internationally, in enabling a collaborative research environment. However, there are several challenges still to be faced in broad-scale implementation of the data hub concept, including distributing and consuming live data feeds and crowd sourced data, simplifying public and private data licensing, enabling access to a wider variety of data formats, and enabling more efficient incorporation of open government data.

Discussions have already begun between INSPIRE and AURIN, and both initiatives plan to work collaboratively to determine the most appropriate outcomes for delivering data. Further workshops detailing specific technical developments, lessons and outcomes will be required to achieve these outcomes.

The increasing availability of free data will provide more opportunities for the establishment of data hubs, and this data will need a mature framework to work within. As such, a key step for developing the data hub concept will be to develop a common framework for establishing data hub infrastructure. This will increase the interoperability of data hubs and which will support future research endeavours, particularly in multi-disciplinary

collaborative research, and capturing user stories to highlight the benefits of the data hub concepts. Addressing these opportunities will maximise the use, collaboration and volume of data provided through data hubs.

Finally, an array of performance tests will need to be developed to ensure AURIN data hubs meet a consistent set of requirements, supporting stability and speed for users. These metrics will examine performance measures, and allow the assessment of strengths and weaknesses in each system. Presenting the performance framework along with the results of performance testing will be the topic of a subsequent research paper.

Acknowledgements

The authors would like to thank the AURIN groups and committees that are directly shaping these efforts. The AURIN project is funded through the Australian Education Investment Fund Super Science initiative. We gratefully acknowledge their support.

10 References

Australian Urban Research Infrastructure Network. (2011). AURIN EIF Final Project Plan. Retrieved on 09/03/2014, from https://web.aurin.org.au/resources/aurin-documents

Bates, J. (2012). "This is what modern deregulation looks like" : co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *The Journal Of Community Informatics,* 8(2).

Batty M, (2013), Resilient cities, networks, and disruption. *Environment and Planning B: Planning and Design* 40(4) 571 – 573

Buildings Performance Institute Europe, (2014). Data Hub. Retrieved 09/03/2014, from http://www.buildingsdata.eu/

Commonwealth of Australia. (2010). Declaration of Open Government. Retrieved 12/07/2013, 2013, from http://agimo.gov.au/2010/07/16/declaration-of-open-government/.

Commonwealth of Australia. (2014). Data.gov.au Organisations. Retrieved 10/03/2014, 2013, from https://data.gov.au/organization

Ding, L., Lebo T., Erickson JS., DiFranzo D., Williams GT,, Li X., Michaelis J., Graves A., Zheng J., Shangguan Z., Flores J., McGuinness DL., and Hendler JA. (2010). TWC LOGD: A portal for linked open government data ecosystems. *Journal of Web Semantics*, 9(3):325–333, 2011.

Dawes, S., Pardo, T., and Cresswell, A. 2004. Designing electronic government information access programs: A h olistic approach. *Government Information Quarterly* 21(1): 3-23.

Delaney, P., Pettit, C (2013). Realising the Data Hubs Concept in Urban Australia. *International Symposium on Next Generation Infrastructure 2013*. Wollongong, Australia

Engström, H., Chakravarthy, S., and Lings, B. (2000). A User-Centric View of Data Warehouse Maintenance Issues. Paper presented at the *17th British National Conference on Databases*, Exter, England.

Eagleson, S. (2012). North West Melbourne Data Integration Project. In Rajabifard, A, Williamson, I & Kalantari, M (Eds.), *A National Infrastructure for Managing Land Information*, Melbourne: CSDILA, The University of Melbourne.

Giuliani, G. Peduzzi, P. (2011). The PREVIEW Global Risk Data Platform: a geoportal to serve and share global data on risk to natural hazards. *Natural Hazards & Earth System Sciences,* 11(1), 53-66.

INSPIRE (2007) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Official Journal of the European Union*, L 108/1 50 (25 April 2007)

Janssen, K. 2011. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4): 446–456.

Larson, J , Olmos Siliceo, M.A. , Pereira dos Santos Silva, M., Klien E., Schade S. (2006) Are Geospatial Catalogues Reaching their Goals. 9th Agile Conference on Geographical Information Science, 20-22 April 2006, Visegrád, Hungary

Manninen, Asta. (2008). Monitoring urban change and identifying future potentials: the case of the European Urban Audit and the State of European Cities Report. *Urban Research & Practice, 1*(3), 222-229. doi: 10.1080/17535060802476400

MongoDB. (2014). Data Hub. Retrieved 15/07/2013, from http://www.mongodb.com/use-cases/data-hub

Myers, T. Trevathan, J. and Atkinson, I. (2012). The tropical data hub: a virtual research environment for tropical science knowledge and discovery. *International Journal of Sustainability Education*, 8 (1). pp. 11-27

Open Knowledge Foundation. (2013A). CKAN Instances around the world. Retrieved 15/07/2013, from http://ckan.org/instances/.

Open Knowledge Foundation. (2013B). The Data Hub - The easy way to get, use and share data. Retrieved 01/05/2013, 2013, from http://datahub.io/sq/about

Nasr, A., Keshtiarast, A. (2013). Datahub for AURIN and ANDS project. In A. Rajabifard, Eagleson, S. (Ed.), *Spatial Data Access and Integration to Support Liveability: A Case Study in North and West Melbourne*. Melbourne: CSDILA, The University of Melbourne.

Oracle (2013). Oracle Life Sciences Data Hub, Oracle Data Sheet. Retrieved 18/12/2013, from http://www.oracle.com/us/industries/life-sciences/045757.pdf

Pettit, C., Stimson R., Tomko, M., Sinnott, R. (2013). Building an e-infrastructure to support urban and built environment research in Australia: a Lens-centric view. *Proceedings of the Spatial Sciences and Surveying Conference* April 2013. Canberra Australia

Redmond, W. (2012). CROSSMARK Delivers On-Demand Sales Insights in New Self-Service Portal. Retrieved 18/12/2013, from http://www.microsoft.com/en-us/news/press/2012/nov12/11-13crossmarkpr.aspx.

Roche, S., Nabian, N., Kloeckl, K., Ratti, C. (2012). Are 'Smart Cities' Smart Enough? *Global Geospatial Conference 2012*, Quebec, Canada.

Socrata. (2013). Socrata Open Data Portal. Retrieved 13/07/2013, from http://www.socrata.com/open-data-portal/

Stimson, R., Sinnott, R., Tomko, M. (2011). The Australian Urban Research Infrastructure Network (AURIN) Initiative: A Platform Offering Data and Tools for Urban and Built Environment Researchers across Australia. *State of Australian Cities (SOAC) 2011,* Melbourne

Tait, M. (2005). Implementing geoportals: applications of distributed GIS. Computers, Environment and Urban Systems 29(1): 33-47.

Theissen, M., Kraemer, E. (2009). Hub-And-Spoke: Building an EDW with SQL Server and Strategies of Implementation. Retrieved 09/07/2013, 2013, from http://msdn.microsoft.com/en-us/library/dd459147(v=sql.100).aspx.

Treloar, A., & Wilkinson, R. (2008). Access to data for eResearch: Designing the Australian national data service discovery services. *International Journal of Digital Curation*, 3(2), 151-158

Vijayan, Jaikumar (2013, September 11). Obamacare data hub is secure and ready to roll. Retrieved 18/12/2013, from http://www.computerworld.com/s/article/9242342/Obamacare_data_hub_is_secure_and_ready_to_roll

Waugh, P. (2013, July 17). New data.gov.au – now live on CKAN. Retrieved 13/07/2013, from http://agict.gov.au/blog/2013/07/17/new-datagovau-%E2%80%93-now-live-ckan

Western Australian Land Information System. (2012). Location Information Strategy for Western Australia. Retrieved 09/03/2014 from http://www.walis.wa.gov.au/projects/location-information-strategy-for-wa

Wickramasuriya Denagamage, R., Ma, J., Berryman, M. & Perez, P. (2013). Using geospatial business intelligence to support regional infrastructure governance. *Knowledge-Based Systems, 53 80-89.*

Williamson, I., Rajabifard, A., Feeney, ME. (2003). *Developing spatial data infrastructures: from concept to reality*. New York, Taylor & Francis.

Yang, C., Raskin, R., Goodchile, M., Gahegen, M. (2010). Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, 34(4), 264-277.

Developing a Usability Framework to Support Online Rapid Urban Information Discovery and Interrogation

Dr John E. Barton
john.barton@unimelb.edu.au

A/Prof Christopher Pettit
cpettit@unimelb.edu.au

Australian Urban Research Infrastructure Network
University of Melbourne 3010

Abstract

One of the significant challenges facing the development of web portals is optimising the usability of such e-infrastructures for diverse, yet specialised uses. This paper presents a usability framework for supporting the iterative testing and development of the Australian Urban Research Infrastructure Network (AURIN). The AURIN portal serves as a securitised environment for data sharing, and as a repository of open-source analytical routines to support urban decision-making. Harmonisation, security, accessibility and discoverability are critical concerns at the data level as many different kinds of federated data are being referenced by the system. Furthermore, urban researchers are accessing this infrastructure framework across Australia. A broad spectrum of digital data relevant to human settlements needs to be not only discoverable, but also comprehensible and usable by non-experts in the area. Managing the complexity across many areas and supporting a spectrum of naïve to expert users are significant challenges. In this research we introduce a usability framework, which elicits feedback from naïve users through to domain experts. This feedback is captured through hands-on workshops with participants and one-on-one user testing sessions. In this paper we report some of the preliminary findings from early workshops with end-users and some of the one on one testing sessions. We analyse this feedback to understand the current strengths and weaknesses of the system. The end-user feedback provides a dual purpose. Firstly, it is fed-back to the technical development team as part of a cycle of continuous improvement. Secondly, the workshops also provide a vehicle for engaging end-users and raising awareness of the AURIN data search, interrogation and visualisation capabilities. We conclude the paper by discussing the next steps in developing and implementing the usability framework in relation to the AURIN portal.

1 Introduction

User-centric design is of critical importance to design of web-based portals and to ensure success as indicated by the end-users returning (Tatnall 2005). The AURIN portal provides access to data and tools to support

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

evidence-based decision-making in the context of urban settlements in Australia. In concert these data and analytical tools provide the backbone to sophisticated Spatial Decision Support System (SDSS) capability. SDSS are essentially iterative, computer based systems designed to support decision-makers in solving semi-structured *spatial* problems (Sprague Jr and Carlson 1982).
An identifying feature of a Decision Support System (DSS) is the employment of information systems to process data into a form that assists knowledge sharing and communication between experts and non-experts, informing a community of users (decision-makers) focussed on solving a common problem (Stock, Bishop et al. 2008, Geertman and Stillwell 2009). Negotiating solutions requires a degree of flexibility as many different stakeholders negotiate potential approaches to the problem based on consensus between many different disciplines. Thus there exists the challenge in providing a SDSS that is usable for researchers across disciplines and with varying levels of expertise to address semi-structured spatial problems, such as determining the walkability of particular new development or the land suitability for a new industrial estate (Butterworth, Giles-Corti et al. 2013).
Spatial Decision Support Systems add a geospatial component to the information that is presented to decision-makers. This is an essential component when dealing with complex urban coverages, helping identify patterns in spatial composition, clustering, change or the predicted impacts of an intervention on the urban environment (Getis 2010, Barthélemy 2011).
Typically embedded within SDSS are expert tools and models, which aim to be effective in processing data into information that is comprehensible by non-experts in a particular domain. For example, urban planning is a field where researchers, planners, policy-makers and communities need to be involved in the decision-making process. When this information is combined, a better understanding of the structure of the problem and potential solutions is attainable. The effectiveness of these tools can be compromised by discipline-specific approaches that only highlight one aspect of the greater challenge. As such, SDSSs that support multi-disciplinary collaborative approaches to solving urban issues need to be designed with accessibility in mind across a range of experience and expertise to best allow diverse (and often divergent) approaches to be combined as seamlessly as possible.
The usability of such a system's front-end is of pivotal importance in determining the uptake and effective deployment of a SDSS. This paper examines the reiterative testing and development cycle of an online SDSS known as the AURIN portal, which is being developed to support researchers focused on solving problems pertaining to the urban settlements across Australia. The AURIN portal provides a gateway to 100s of geospatial datasets and a suite of analytical and visualisation tools to support this broad user community. Areas of endeavour span across social demographics, housing supply and demand, transport, population health and well-being and innovative urban design applications. These discipline-specific clusters are referred to as 'AURIN lenses' as they conceptually focus on a particular area of national significance and provide an interface between the technical development, the expert community and end-user engagement (Pettit, Tomko et al. 2013). Each particular lens group have identified priority data, analytical tools, models and visualisation routines that are being built into an open source portal e-infrastructure to support the broad end-user community. The fundamental challenge is that each of these resources needs to be easily discoverable and useable in such a way that the respective users goals are met sufficiently to enable rapid goal-seeking behaviour free of heuristic barriers.

2 Background

The challenges presented by urban areas increasing in both size and density, coupled with greater demands on infrastructure, socio-economic policy and sustainable design and planning have given rise to many innovative solutions internationally as to how to best manage these areas. Many groups have formed around the concept of 'smart cities' (Gibson, Kometsky et al. 1992, Schaffers, Komninos et al. 2011) and continued development has seen the emergence of high-fidelity systems primarily developed for urban areas; common instruments being employed include multi-disciplinary information sharing (Fonseca, Egenhofer et al. 2000), web-based portals (Maguire and Longley 2005) and use of three-dimensional visualisation as tools for information delivery (Zlatanova and Gruber 1998, Döllner and Hagedorn 2007).
There is a general direction for governments internationally to move toward an open data sharing model (Shah, Abraham et al. 2012, Tinati, Carr et al. 2012, Fitzgerald, Hooper et al. 2013). One example of note is the Infrastructure for Spatial Information in Europe (INSPIRE); a long-running initiative bringing geospatial information together across Europe (Annoni, Bernard et al. 2004). In the Australian context, the Government has been progressively embracing opening access to data and the employment of open source software (Missingham 2008). Increased levels of access to urban data on a non-commercial basis is progressively a easing a traditional supply-side blockage to the use (and re-use) of geospatial urban data (Vonk, Geertman et al. 2005).

AURIN provides the links to existing datasets federated as closely to their respective custodians as possible (Sinnott, Galang et al. 2011). An additional layer of tools and workflows facilitate urban information to be accessed, analysed, simulated and interrogated (Tomko, Bayliss et al. 2012). This collaborative data-sharing platform, combined with the analytical layers, forms a SDSS that many diverse users will be accessing. This makes usability issues paramount to the uptake, effective use and longevity of the system (Shneiderman and Plaisant 2003).

Typically SDSS are designed by experts for experts; for instance, Planning Support Systems (PSS) aim in engaging an end-user community (in this case planners) in the use of computer assisted decision support tools (Geertman, Toppen et al. 2013, Pettit, Nino-Ruiz et al. 2014). Vonk, Geertman et al. (2005) created a theoretical framework and have identified a number of bottlenecks of widespread adoption of these tools in the planning community (see Figure 1). Key factors include: Intention to use (78%), Awareness (66%), Organisational Facilitators (including training) (61%), Personal Adopter Characteristics, (61%), Ease of Use (58%), and Usefulness (51%). Such factors as relevant to PSS adoptions are considered to be applicable to the use of SDSS tool contained with the AURIN portal which are design to support end-users which includes researchers in the field of urban planning, urban design, geography, and policy and decision makers.

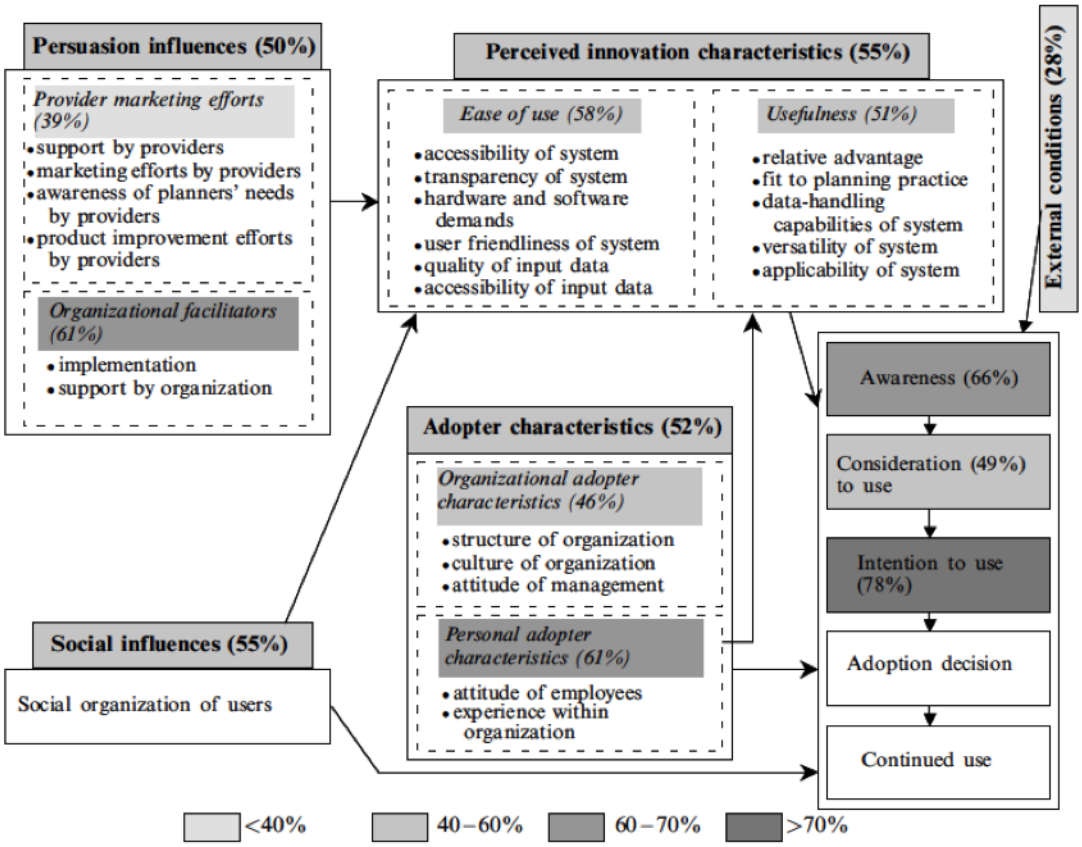


Figure 1: Bottlenecks blocking wide spread usage and adoption of PSS with importance scores (Vonk, Geertman et al. 2005).

After an initial period of technical development, AURIN has begun a series of workshops, user surveys and interviews to introduce the system to end-users from a range of urban disciplines. As part of this engagement process, detailed user testing is being undertaken by a small group of typical end-users. This is to achieve a better understanding of how people are interacting with the system and, importantly, to provide feedback to the technical development team to address usability and functionality issues identified by the group. The AURIN development process follows the agile method (Schwaber 2004). This iterative model successively refines each prototype at incremental beta phases (Huo, Verner et al. 2004). The agile method maintains a close link between developers and the client, where feedback is taken on board the development process in regular

increments. Small teams of developers discuss this feedback face-to-face and maintain a focus on satisfying client needs with working software. The agile method does not focus heavily on verbose documentation, contractually defined scope or rigid long-term planning (Conforto and Amaral 2010, Dingsøyr, Nerur et al. 2012). This development methodology has been adopted by AURIN as it best facilitates the interaction between many clients/stakeholders to prioritise optimally engineered software solutions for complex and evolving projects. The next section will describe formalised workshop sessions designed to expose the software to end-users during the beta phases of development. In turn, the feedback from these workshops is delivered back to the development team and progressively addressed. Pacing the technical development process with user feedback aligns strongly with the concept of SDSS for urban problem-solving and stakeholder community engagement (Andersen and Jger 1999, Batty, Chapman et al. 2000, Balsamo, Di Marco et al. 2004).

3 Usability Framework

3.1 Who were the respondents?

Respondents were predominantly involved in urban research at a university level. Half of the respondents were University Academic (Teaching and/or Research) and a third were University Post-graduate students (PhD or Masters). The remainder were involved in Local Government. A factored gap analysis of the self-assessed level of expertise showed the user groups to rank themselves as more adept in urban issue than technical skills, with one-third of users classifying themselves as 'expert-level' and another third as 'inexperienced'. This is to be expected as the workshops were run in-conjunction with two conferences the State of Australian Cities (SOAC) (2 workshops) and The Australia New Zealand Regional Science Association International (ANZRSAI) (1 workshop) and attracted a participants interested in urban issues and ranging from student-level to senior academics.

Between beta releases, it is important to gather as much information as possible to identify usability issues and feed these back into the development process. AURIN is using several ongoing instruments to gauge the performance of the system as a whole:

1. End-user Workshops
2. User Advisory Group (UAG) Sessions
3. Monitoring and Evaluation

Figure 2 illustrates the temporal framework of the user testing process. Beta 3 was the first public release inviting user feedback, the system having been developed internally only for the first two beta releases. Preceding each subsequent beta release is a User Advisory Group (UAG) session followed by a series of workshops. The process employs screen capture software and an unguided speak-out-loud interview process when conducted in a usability lab. Feedback received by these instruments is fed-back into the development process via the existing bug reporting and feature request framework.

Workshops are conducted between beta releases to give participants hands-on experience in a collaborative environment. These typically last 3 hours. At the conclusion of these workshops, participants are surveyed to gauge usability, confidence, focus areas/scales, self-assessed discipline and technical experience (via Likert scales) and open-ended questions prompting good points, bad points and suggestions for future development.[1]

The UAG has been established to test pre-release functions and provide feedback on the usability and functionality of the system at each beta phase of the portals development. The UAG comprises of professional users from the urban research community, which spans the disciplines encompassed through the AURIN Lenses to form a small group of invited participants for each beta testing phase. The membership of this group may change depending on the functionality associated with a particular Beta release.

Individual sessions have been undertaken with an individual users and at least one AURIN member to take notes and help in session facilitation. Each session lasts 1.5-2 hours. The sessions are tailored to specifically address the new functionality planned for a particular Beta release. Each user session consists of five components:

1. Blue sky: Users would convey their understanding of what the AURIN portal is and what it potentially could provide.

[1] See: https://docs.aurin.org.au/wp-content/uploads/2013/12/131201_survey.pdf

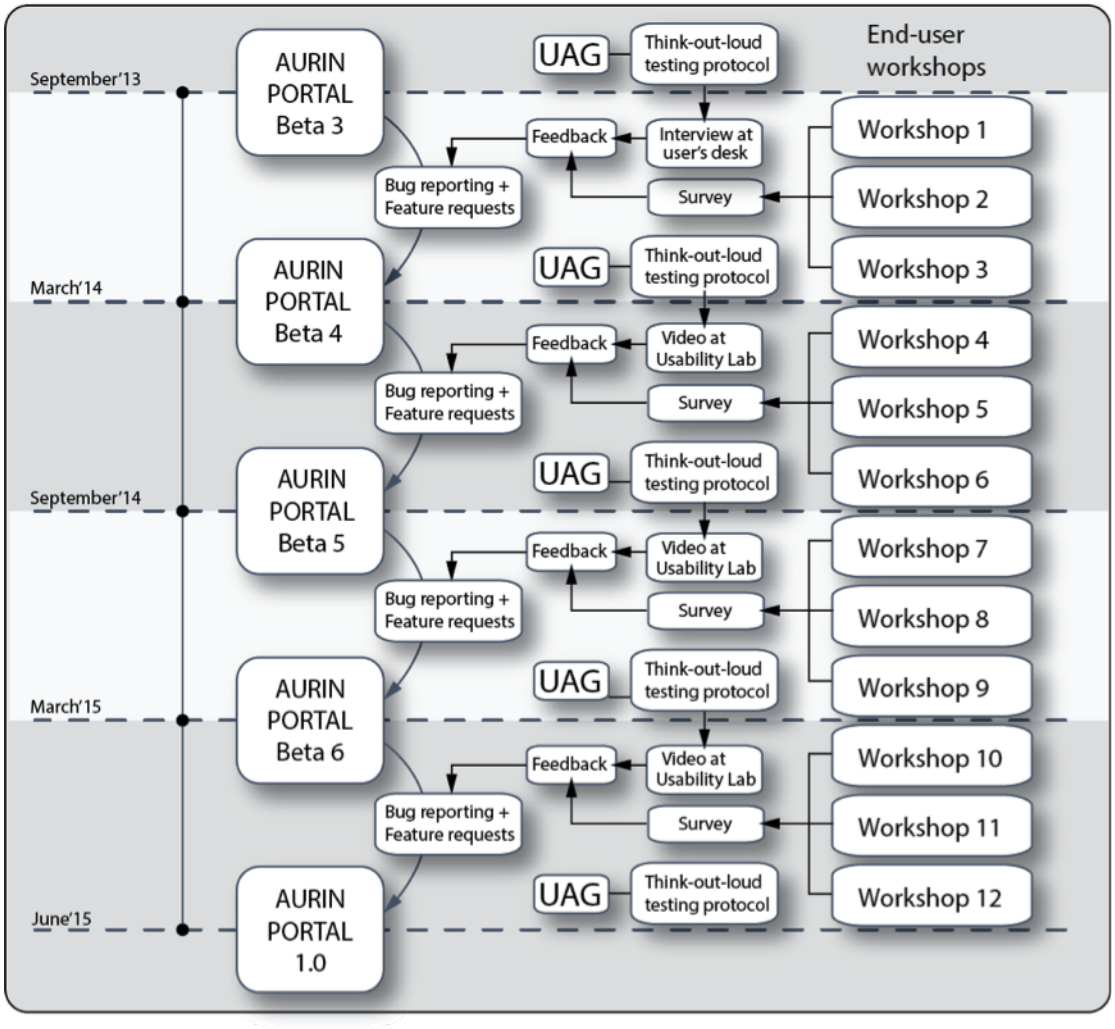


Figure 2: AURIN usability framework.

2. Unguided test: Without any assistance, participants are asked to log into the AURIN portal and a simple exercise such as preparing a chart.

3. Guided test: Any difficulties encountered in pt.2 would be addressed and the user guided through a more difficult task with the support of user manual and tutorial documentation where available.

4. Collaborative brainstorming of features and functionality: discussing user stories describing what/how a user might use data/tool/visualisation resources with basic interaction flows and desired outputs.

5. Debriefing: Finishing this process the user will be given the survey with additional questions focussing on issues of access, interface and task based interactions.

AURIN has undertaken a parallel survey as part of a Monitoring and Evaluation process harmonised with the beta releases to gauge the perceptions of the broader user community, particularly stakeholders that are involved in the broader project who have not had any specific training in the operation of the system.

4 Results

4.1 Workshops

In December 2013, three workshops were facilitated by AURIN introducing a total of 24 participants to the AURIN portal. Participants were invited to fill out a paper survey on completion of each workshop. In total 15

responses were collected.

4.1.1 What were respondents interested in?

Respondents indicated an interest in all scales of urban research ranging from an individual to a National level. Predominantly, 60% of respondents indicated and interest in the precinct/neighbourhood level, with 43% indicating interest in cities and towns, and 30% indicating interest at building scale. 70% of respondents were interested in Urban Housing, and nearly 60% interested in Urban Transport, although no responses were indicated for Urban Logistics.

4.1.2 What were the perceptions of the AURIN portal?

Respondents indicated their main interest was in data download (4.36 on the Likert scale) and Data search and discovery (4.31) while Analytical tools indicated the weakest demand (3.67). This result could confirm that the portal is primarily seen as a data acquisition tool, however still registering a demand for more specialised analytical routines. Only 4 respondents (26%) indicated that they were already familiar with the Australian Urban Research Infrastructure network (AURIN) project prior to using the AURIN portal, however, respondents strongly agreed that the information on the AURIN portal is valuable and that they were likely to visit the AURIN portal in the future. A score of only 3 was registered by respondents indicating that they were neutral as to being able to find what they need quickly on the AURIN portal. This points towards issues associated with usability of the portal that need to be considered.

4.2 User Advisory Group

Over the period of a month concluding mid-September 2013, Six (6) individuals were interviewed, the first two interviews used the Beta2 release and the subsequent four used Beta3. Participants were from academia (3) and Local Government (2) and State Government (1). This initial series of interviews was aimed at gleaning detailed user feedback from a small number of users in order to identify common issues at an early stage. It is important to note that each of the participants had been to an introductory AURIN workshop, or had been recommended by someone who had attended. Consequently, these users had a professional interest in retrieving, analysing and presenting urban information. The interviews were conducted at participants desks in order to identify any access issues via firewalls, software, hardware or bandwidth constraints. Participants were asked to 'Think out loud' to register their perceptions as two AURIN representatives notated responses.
The unguided testing phase replicated the experience of a user logging into the AURIN portal, navigating the interface, loading data and creating a choropleth map without any guidance beyond what was accessible online. Testing revealed that the users all had specific field of interest and were motivated to retrieve data specific from their respective fields- these broadly covered statistical demography, urban planning/conflict issues, urban transport analysis and well-being/social development issues. Once successfully logged in, each of the users were able to explore, find and retrieve relevant data and all were able to create a choropleth map within 30 minutes.

Common issues in unguided tests were revealed:

- at the log-on phase when users needed to correctly enter user names and passwords. If a password had been forgotten, a lengthy process of retrieval and resetting ensued- in one case up to 54 minutes before successful login. This poses a significant barrier to accessing the system with ease;

- at the interface interaction stage of trying to decipher the portal-specific iconography and workflow. Often users saw the system as being analogous to similar systems they had had experience with, however most were reluctant to refer to manuals or help documentation, preferring to intuitively navigate through the system;

- when users followed what appeared to be an intuitive workflow, goals were only revealed when the entire process was carried out (for example, interactive brushing), or if an incorrect path was followed, error messages were not revealed until several steps into the process. Users could not modify input parameters after loading; this was identified as an important part of refining a trial-and-error process.

During the collaborative brainstorming and de-briefing stage, users volunteered techniques to remedy problems or improve effective use of the portal including:

- refining the front end for intuitive interaction for first-time users, or with an interpretive layer to guide the user;
- avoiding custodian-specific acronyms, codes or jargon in preference of verbose, human readable script, with subsequent opportunities for revealing specific detail;
- pre-built use-cases, parameterised queries and industry-standard reports accessible via themes grouped around discipline;
- programmatic access for query and retrieval to bypass the portal.

Users trusted/had no concerns over the curation or reliability of data, and recognized the federated data structure to be a timesaving and trust-worthy feature of the portal. Users liked the fact that federated data was straight from the source, and not derived or static. This also created perceived opportunities for participating in forming and adding to the data network itself.
Diverging opinions were revealed between users- some saw it mostly as a data discovery and retrieval tools, while others saw value in the ability generate a visualisation layer over existing datasets. This is a usability issue that warrants further investigation.

4.3 Monitoring and evaluation

During October 2013, a series of surveys were conducted from a total sample of 106 AURIN stakeholders as part of an ongoing Monitoring and Evaluation process. Stakeholders are represented by the project's community-developers, project leaders, senior researchers and collaborators. Table 1 shows the responses to open-ended questions prompting respondents for three good points ('good'), three bad points ('bad') and three points for improvement ('todo'). These were clustered around common key themes.

Table 1: Clustered results from open-ended survey questions.

	GOOD	BAD	TODO	
Interface	67	42	34	143
Data	52	21	24	97
Feature Requests	7	21	27	55
Bug Reporting	1	30	11	42
Metadata	4	21	16	41
Speed/Performance	23	11	5	39
	154	146	117	417

Respondents primarily focussed on the user interface, both for positive and negative feedback. Respondents also identified the data handling capabilities of the portal as a major strength.

5 Conclusions

The first round of user testing has been an important step in empirically identifying and developing a profile of the system's users. Broadly, users are interested in urban research; this group are mostly members of the academic research community, and the government sector. Their skills range from naïve to expert. This suggests a focus on developing the user experience aspects of the portal to accommodate first-time users, as well as offering guides to more advanced features.
The usability framework and approach will be extended for progressive beta releases to focus in more fidelity on the details of user interaction using a dedicated usability lab where usability testing can be undertaken with a triangulation of approaches including Speak-out-loud, video recording and pre/post questionnaires. Usability inspection through Nielsens (1994) Ten Usability Heuristics are also being considered to be included in the usability framework to identify usability problems in the AURIN user interface design.
Subsequent User Advisory Groups will invite a larger sample of participants, first to confirm and elaborate on established personas, and secondly, to continue identifying issues encountered in the user experience. The cycle of testing at the beta phases enables newly implemented features to be tested, and further related feature requests to be registered. The final UAG meeting and testing will represent an end point to gauge the project's progress

and best ensure the continuity of the project into another phase of development.
The usability of portals and advanced SDSS tools has been identified as a major barrier for the use of such tools (Vonk, Geertman et al. 2005). Our research aims to address this issue by developing and applying a usability framework through-out the development cycle of the AURIN portal and implemented it at critical points in time to align with Beta releases. Future research will report on the refinement of this usability framework and subsequent longitudinal end-user testing results.
Overall, the response from the feedback acquired through the usability framework has been positive, with end-users generally citing AURIN as a valuable initiative filling a gap. Whilst AURIN is currently at the Beta3 phase, more work is required to 'harden' the portal. This will in turn enable a viable comparative analysis to similar portal products that will further assist in positioning and improving this online SDSS. Accessibility, ease of use, functional robustness and an outcome-based focus are crucial factors to the overall usability of the AURIN portal and ultimately its success.

Acknowledgements

AURIN is an Australian Government project conducted as part of the Super Science initiative and financed by the Education Investment Fund. The University of Melbourne has been appointed the lead agent by the Commonwealth of Australia, Department of Innovation, Industry, Science and Research.
AURIN would like to thank members of the User Advisory Group, Workshop attendees and survey respondents.

References

Andersen, I.-E. and B. Jæger (1999). Scenario workshops and consensus conferences: towards more democratic decision-making. *Science and public policy 26*(5), 331–340.

Annoni, A., L. Bernard, K. Fullerton, H. de Groof, I. Kanellopoulos, M. Millot, S. Peedell, D. Rase, P. Smits, and M. Vanderhaegen (2004). Towards a european spatial data infrastructure: the inspire initiative. In *Proceedings of the 7th international global spatial data infrastructure conference, Bangalore, India, February 2*, Volume 4.

Balsamo, S., A. Di Marco, P. Inverardi, and M. Simeoni (2004). Model-based performance prediction in software development: A survey. *Software Engineering, IEEE Transactions on 30*(5), 295–310.

Barthélemy, M. (2011). Spatial networks. *Physics Reports 499*(1), 1–101.

Batty, M., D. Chapman, S. Evans, M. Haklay, S. Kueppers, N. Shiode, A. Smith, and P. M. Torrens (2000). Visualizing the city: communicating urban design to planners and decision-makers.

Butterworth, I., B. Giles-Corti, and C. Whitzman (2013). Perspective 1: Setting the scene for the north and west melbourne data integration and demonstrator projects. *Spatial Data Access And Integration To Support Liveability*, 11.

Conforto, E. C. and D. C. Amaral (2010). Evaluating an agile method for planning and controlling innovative projects. *Project Management Journal 41*(2), 73–80.

Densham, P. J. (1991). Spatial decision support systems. *Geographical information systems: Principles and applications 1*, 403–412.

Dingsøyr, T., S. Nerur, V. Balijepally, and N. B. Moe (2012). A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software 85*(6), 1213–1221.

Döllner, J. and B. Hagedorn (2007). Integrating urban gis, cad, and bim data by servicebased virtual 3d city models. *R. e. al.(Ed.), Urban and Regional Data Management-Annual*, 157–160.

Fitzgerald, A., N. Hooper, and J. S. Cook (2013). Implementing open licensing in government open data initiatives: a review of australian government practice. In *Proceedings of the 9th International Symposium on Open Collaboration*, pp. 39. ACM.

Fonseca, F. T., M. J. Egenhofer, C. A. Davis Jr, and K. A. Borges (2000). Ontologies and knowledge sharing in urban gis. *Computers, Environment and Urban Systems 24*(3), 251–272.

Geertman, S. and J. Stillwell (2009). *Planning support systems: content, issues and trends*, pp. 1–26. Springer.

Geertman, S., F. Toppen, and J. Stillwell (2013). *Planning support systems for sustainable urban development*, Volume 195. Springer.

Getis, A. (2010). Spatial filtering in a regression framework: examples using data on urban crime, regional inequality, and government expenditures. In *Perspectives on spatial data analysis*, pp. 191–202. Springer.

Gibson, D. V., G. Kometsky, and R. W. Smilor (1992). *The technopolis phenomenon: Smart cities, fast systems, global networks*. Rowman & Littlefield.

Huo, M., J. Verner, L. Zhu, and M. A. Babar (2004). Software quality and agile methods. In *Computer Software and Applications Conference, 2004. COMPSAC 2004. Proceedings of the 28th Annual International*, pp. 520–525. IEEE.

Maguire, D. J. and P. A. Longley (2005). The emergence of geoportals and their role in spatial data infrastructures. *Computers, environment and urban systems 29*(1), 3–14.

Missingham, R. (2008). Access to australian government information: A decade of change 19972007. *Government Information Quarterly 25*(1), 25–37.

Nielsen, J. (1994). *Usability engineering*. Access Online via Elsevier.

Pettit, C., R. Klosterman, M. Nino-Ruiz, I. Widjaja, P. Russo, M. Tomko, R. Sinnott, and R. Stimson (2013). *The Online What if? Planning Support System*, Volume 195 of *Lecture Notes in Geoinformation and Cartography*, Chapter 20, pp. 349–362. Springer Berlin Heidelberg.

Pettit, C., M. Nino-Ruiz, R. Klosterman, P. Russo, M. Tomko, R. Sinnott, and R. Stimson (2014). Assisting evidence-based urban planning through the development of an online what if? planning support system.

Schaffers, H., N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira (2011). *Smart cities and the future internet: towards cooperation frameworks for open innovation*, pp. 431–446. Springer.

Schwaber, K. (2004). *Agile project management with Scrum*. O'Reilly Media, Inc.

Shah, N., S. Abraham, and G. Wright (2012). Open government data study: India.

Shneiderman, B. and S. Ben (2003). *Designing The User Interface: Strategies for Effective Human-Computer Interaction, 4/e (New Edition)*. Pearson Education India.

Sinnott, R. O., G. Galang, M. Tomko, and R. Stimson (2011). Towards an e-infrastructure for urban research across australia. In *E-Science (e-Science), 2011 IEEE 7th International Conference on*, pp. 295–302. IEEE.

Sprague Jr, R. H. and E. D. Carlson (1982). *Building effective decision support systems*. Prentice Hall Professional Technical Reference.

Stock, C., I. D. Bishop, A. N. O'Connor, T. Chen, C. J. Pettit, and J.-P. Aurambout (2008). Sieve: collaborative decision-making in an immersive online environment. *Cartography and Geographic Information Science 35*(2), 133–144.

Tatnall, A. (2005). *Web portals: the new gateways to Internet information and services*. Igi Global.

Tinati, R., L. Carr, S. Halford, and C. Pope (2012). Exploring the impact of adopting open data in the uk government.

Tomko, M., C. Bayliss, G. Galang, P. Greenwood, G. Koetsier, D. Mannix, L. Morandini, M. Nino-Ruiz, C. Pettit, and M. Sarwar (2012). The design of a flexible web-based analytical platform for urban research. In *ACM SIGSPATIAL GIS 2012*. ACM.

Vonk, G., S. Geertman, and P. Schot (2005). Bottlenecks blocking widespread usage of planning support systems. *Environment and planning A 37*(5), 909–924.

Zlatanova, S. and M. Gruber (1998). 3d urban gis on the web: data structuring and visualization. *IAPRS 32*(4), 6912699.

Ontology driven VGI filtering to empower next generation SDIs for disaster management

Saman Koswatte[1]
Saman.Koswatte@usq.edu.au

Kevin McDougall[1]
kevin.mcdougall@usq.edu.au

Xiaoye Liu[1]
xiaoye.liu@usq.edu.au

[1]School of Civil Engineering and Surveying, Faculty of Health, Engineering and Science
University of Southern Queensland, Toowoomba Qld 4350, Australia

Abstract

In modern society, the increased popularity of social media has significantly changed the reporting and sharing disaster related information. The Volunteered Geographic Information (VGI) can be considered as a special case of the more general web phenomenon of User Generated Content (UGC). Interestingly, VGI is more current and more diverse than conventional geographic information, although quality and credibility issues exist. The Spatial Data Infrastructures (SDIs) facilitate the spatial data sharing between organizations to discover, access and use available spatial data. As the mobile communication, information technology and other related infrastructures are changing rapidly, the SDI and its architectures should also be considered in order to develop the next generation SDIs. This paper examines the fusion of VGI and SDI as a possible solution to solving SDI's pressing issues like lack of currency and data incompleteness. As ontologies are self-descriptive knowledge of shared conceptualizations, it is argued that the use of domain specific ontologies can successfully improve data currency and quality. The purpose of this paper is to explore the use of domain specific ontologies to filter VGI to be integrated with SDI data by addressing the above specific problems. A conceptual framework is proposed to filter VGI using domain ontologies. The study approach consists of multiple steps including 1) Special domain specific terminology identification of 2011 Queensland-Australia flood tweets, 2) Crowd sourced disaster management ontology development, 3) Tweet processing and improvable spatial data extraction, 4) Validation and Quality control and 5) Integration with SDIs and disaster management applications. The paper discusses a new crowd sourced spatial data life-cycle and provides a comparison of VGI, SDI and next generation SDIs in terms of data quality, standards, currency and breadth of data together with an approach for ontology driven semantic official data generation.

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

Understanding the Gap between the United Nations World Food Programme Crisis Mapping Operations and Crowdsourcing Technology

Sophie E. Richards
Department of Spatial Sciences, Curtin University
GPO Box U1987, Perth WA 6845, Australia
srichards@globalskm.com

Bert Veenendaal
Department of Spatial Sciences, Curtin University
GPO Box U1987, Perth WA 6845, Australia
B.Veenendaal@curtin.edu.au

Abstract

There is increasing pressure from the crisis mapping community for United Nations agencies to adopt crowdsourcing technology as part of existing United Nations crisis mapping, emergency response operations. Whilst United Nations agencies such as the World Food Programme are in support of crowdsourcing initiatives, it is imperative that the technology be assessed before it can be adopted as part of the existing crisis mapping operations. It is frequently argued in theoretical scientific papers that during a crisis situation, the limitations associated with crowdsourcing technology are outweighed by the benefits of its use. However, it can also be argued that in crisis mapping operations, crowdsourcing technology is not of sufficient maturity at present to provide adequate benefits. To understand the capability of crowdsourcing technology for crisis mapping, this was tested by evaluating a number of existing crowdsourced applications. Results of this research indicate that crowdsourcing technology is in its infancy and current applications do not meet the expectations required by the World Food Programmes' crisis mapping operations.

1 Introduction

It has been estimated that by 2050 the number of people affected by crisis events will rise to 2 billion (Aten 2011, pp. 16-20). With this in mind new and emerging technologies such as crowdsourcing must be well understood by the World Food Programme (WFP) for crisis response. Whilst there is increasing pressure from the crisis mapping community for WFP to adopt crowdsourcing technology, the technology can only be adopted if it is of benefit to crisis mapping operations. Assessment based purely from a theoretical scientific perspective, determines that despite numerous limitations associated with implementing this technology, the benefits surpass these limitations when used in support of crisis response (Goodchild & Glennon 2010, pp. 231-241). However, in practice the hypothesis that crowdsourcing technology currently available in the crisis mapping arena is of benefit to existing United Nations (UN) World Food Programme (WFP) crisis mapping operations, needs to be explored. Therefore this hypothesis was tested through a research project. Within this paper the investigation of existing crowdsourcing technology applied to recent crisis events is presented. The degree of benefit associated with the application of crowdsourcing technology to current WFP crisis mapping operations, was determined by evaluating against an identified set of criteria representing the data requirements of these operations.

1.1 Crowdsourcing the geovisualisation production process during emergency response

The term crowdsourcing can be described as 'giving a task for a crowd to execute, instead of executing it oneself' (Vivacqua & Borges 2011, pp. 189-198) suggesting that crowdsourcing is a 'further development of outsourcing' (Hirth, Hoßfeld, & Tran-Gia 2012, n.p.). Within this paper the term crowdsourcing in relation to crisis mapping refers to the outsourcing of geospatial tasks to a large crowd of citizens and volunteers in the production of a geovisualisation. This research is focused on crowdsourcing the Geovisualisation Production Process (GPP), a process commonly used for crisis mapping operations such as those used within WFP. The GPP can be broken

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

down into four stages involving the data: generation, acquisition, processing and publication. During a crisis situation where time is of a premium, all or some of the GPP stages may be crowdsourced by a large crowd rather than be completed by a small team of professionals.

Crowdsourcing for Emergency Response (ER) can be described as the outsourcing of GPP tasks to volunteers and citizens located either in the affected area or at a location anywhere in the world. The volunteers and citizens are recruited by publicising the crowdsourcing needs via an appeal on social networking and internet sites. During the ER to a crisis situation large volumes of crisis data are generated. Geovisualisation and web mapping provide a rich environment to manage and aggregate data and allow user communities to collaborate (Li, Veenendaal & Dragićević 2011). This is especially important at a time when maintaining strong communication channels, between ER officials, volunteers and citizens within the crisis affected area, is 'extremely difficult'(Aten 2011, pp. 16-20).

Through the completion of a literature review as part of this research, a number of technological benefits and limitations were identified. The theoretical benefits associated with crowdsourcing technology for crisis mapping situations include achieving success in a small amount of time, reduced costs to perform geovisualisation tasks, being able to harvest local knowledge and permitting acquisition of collaborative knowledge. There were also several theoretical technological limitations discovered including; positional and attribute accuracy limitations, large volumes of data or big data issues and limitations relating to extracting actionable data. As these are theoretical benefits and limitations, practical research was conducted to understand the full capability of crowdsourcing technology.

1.2 Voluntary and involuntary crowdsourcing

There are two forms of crowdsourcing; voluntary and involuntary crowdsourcing. As the term suggests, voluntary crowdsourcing involves the voluntary participation by the crowd with crowdsourcing technology resulting in volunteered geographic information (Goodchild & Glennon 2010, pp. 231-241). An example of voluntary crowdsourcing is a crowd member who typically is a volunteer actively participating in generating crisis data through a crowdsourcing platform such as OpenStreetMap.

Involuntary crowdsourcing involves acquiring data from social networks such as: Twitter, Flickr, YouTube, Facebook, Blogs and wikis, whilst the original creator of the crisis data is unaware that the data will be used in a crowdsourcing application. An example is Ushahidi which is a popular Web 2.0 platform that has been deployed in more than 30 countries (Greenwald 2010, pp. 43-47). Whilst there is increasing popularity to harvest knowledge from social networking sites in the ER to a crisis situation, this form of involuntary crowdsourcing does present its challenges (Rutsaert et al. 2013, pp. 84-91). Data acquired through involuntary crowdsourcing methods typically result in a large, complex dataset which requires intensive data management to extract the actionable and relevant information. The reliability of the acquired data is also an issue as often the source or metadata is unknown.

2 The World Food Programme

The World Food Programme (WFP) 'is the world's largest humanitarian agency fighting hunger worldwide' (WFP 2013a). Operations conducted by WFP relate to promoting food security through sustainability and the provision of food aid to over 75 countries. These countries are typically developing countries that have the 'most vulnerable populations' (WFP 2012a).

In the ER to a natural disaster or shock event, WFP provides the affected population with access to food aid. WFP responds to a variety of food security crisis situations that 'may last several years' (UNFAO 2006, pp. 24-25). A food security crisis can be described as 'extreme food insecurity, when the main danger is widespread loss of access to food, perhaps leading to famine' (UNFAO 2006, pp. 24-25). A food security crisis like other crisis situations can be defined as either slow or a sudden-onset crisis situation. A sudden-onset food security crisis situation is 'associated with natural disasters triggered by climatic hazards, such as floods or hurricanes' (UNFAO 2006, pp. 24-25) and occurs with little to no warning. A slow-onset food security crisis situation is different in that this type of crisis will 'arise when people who are chronically food-insecure are faced with recurrent or persistent external shocks such as drought' (UNFAO 2006, pp. 24-25). A slow-onset crisis situation will emerge over a longer period of time and may be seasonal or predictable in nature.

2.1 The World Food Programmes' crisis mapping geospatial data needs

WFPs' geospatial data needs required in the ER to a crisis situation relates to humanitarian access, the location or movement of population (including refugees and internally displaced people) and food security. These geospatial data needs were gathered as part of this research, and are presented in Table 1.

Data relating to humanitarian access, the location of population and food security must be rapidly acquired to support WFPs' crisis mapping operations in response to sudden onset crisis situations, as well as sustained data acquisition for slow-onset crisis situations. During crisis mapping operations, WFP need to acquire local knowledge that is relevant to their crisis mapping geospatial data needs to ensure a true and accurate

representation on the crisis situation and affected population is mapped. As WFPs' crisis mapping products affect decision making, the data acquired must be of high accuracy and reliable.

Table 1: The World Food Programmes' Crisis Mapping Geospatial Data Needs

Geospatial Data Need	Specific Geospatial Data Need
Humanitarian access	Roads, rail, navigated waterways, ports, airports, airfields, heliports, air routes, tunnels, bridges, dams, barrages, levees, ferries, humanitarian corridor, entry points/border crossing points, distance matrix between main points of interest, elevation data, slope, access to markets, obstacles and checkpoints.
The location or movement of population	Populated places, classification of populated places (national/provincial/district capital), refugee camps, population figures or population estimates of each populated place.
Food Security	Agricultural areas, grain storage facilities, bakeries, hospitals, schools, colleges, stadiums, built-up areas, location of existing services (water, power, waste), oil and gas wells and communication facilities.

2.2 The World Food Programmes' crowdsourcing application requirements

The degree of benefit associated with the application of crowdsourcing technology to current WFP crisis mapping operations was determined in this research by evaluating applications against a number of criteria related to WFP's requirements. These criteria are identified as: time, local knowledge, collaborative knowledge and crowd participation, positional accuracy, attribute accuracy, big data (spatial data handling potential), actionable data and meets WFPs' geospatial crisis data needs.

Table 2 outlines WFPs' crowdsourcing application requirements in relation to these eight criteria and was identified together with the Emergency Preparedness and Response Branch at the WFP headquarters in Rome, Italy. This research determined crowdsourcing application requirements in relation to data accuracy for decision making, flexibility to support slow and sudden crisis situations and data acquisition to meet WFPs' needs. Research relating to the UN crowdsourcing for crisis mapping examples determined the need for spatial handling capability and accuracy for decision making.

Table 2: WFPs' crowdsourcing requirements

Criteria	WFP Crowdsourcing Application Requirement
Criteria 1: Time	The application would need to be able to support the continual and rapid crisis mapping timeframes for the ER to both slow and sudden-onset crisis situations respectively.
Criteria 2: Collaborative Knowledge (Crowd Participation)	The application would need to have a high level of crowd participation to ensure that crisis mapping tasks are completed within typically limited timeframes. Crowd participation would need to be constant and sustained to support both the slow-onset and sudden-onset crisis situations. The resultant collaborative knowledge needs to be a true representation of the whole affected population.
Criteria 3: Local Knowledge	The application would need to be accessible and useable by the affected population in all countries in which WFP has a presence, despite the countries' level of economic development.
Criteria 4: Attribute Accuracy	Not only would the application need to achieve a high degree of attribute accuracy, but it would also need to have attributes which do not delay crisis mapping operations. E.g. long textual comments associated with each observation or attributes containing uncertainty will delay critical crisis mapping operations and are not practical.
Criteria 5: Positional Accuracy	The application would need to achieve a high degree of positional accuracy and need to generate reliable crisis data. Million dollar food aid decisions need accurate and reliable location information.
Criteria 6: Spatial Data Handling Capability (Big Data)	In the typically time-constrained crisis mapping operations, the application would need to support ease in spatial data handling in order to permit further manipulation of the geospatial data.
Criteria 7: Actionable Data	The application would need to define and extract actionable geospatial data with ease to permit timely assessment of food security and factors affecting food aid decision making.
Criteria 8: Meets WFP Spatial Data Needs	The application would need to be suited to support WFP's geospatial data needs as outlined within this chapter.

3 Study: recent crisis situations and utilisation of crowdsourcing technology

A number of crowdsourcing applications deployed on the Internet in response to two crisis events were selected and investigated for this research. The crisis events were chosen because the WFP undertook crisis mapping operations in the ER to these events during the research period of November to December 2012. The selected crisis events include the November 2012 Gaza crisis and the December 2012 Philippines Typhoon Bopha crisis. For each, the available web-based crowdsourcing applications were discovered and information was obtained on the extent and amount of mapped data, platform used, methods of capture, distribution of data sources and range of useable data. These are identified in the following subsections.

3.1 Gaza Crisis

Israeli–Palestinian violence began in the Palestinian Territory, Gaza Strip on Saturday 10th November 2012. On the 14th November this violence escalated. A cease-fire between Israel and Hamas was agreed upon on 21st November 2012. This crisis is a slow onset crisis situation resulting from conflict and therefore requires ER and crisis mapping operations from WFP for a sustained period of time. With over 2.5 million people affected (UNOCHA 2012a) by this crisis, the magnitude of this crisis situation is great. Two crowdsourcing applications were discovered in relation to this crisis situation; the Palestine Crisis Map crowdsourcing platform (Ushahidi 2012a) and the Tracking Social Media from Israel and Gaza crowdsourcing platform (AlJazeera 2012). The crowdsourcing applications were analysed and evaluated over a period of 10 days. The 10 day period was determined based on the period from when the November violence began (10/11/2012) until when the WFP crisis mapping began (19/11/2012).

3.1.1 Palestine Crisis Map crowdsourcing platform

The Palestine Crisis Map crowdsourcing platform (Ushahidi 2012a) attracted 39 responses over a 10 day period via a web form in order to identify data that may be useful in the ER. The map in Figure 1 illustrates the overall data and categories that were crowdsourced using the Ushahidi platform (Ushahidi 2012a). Of the responses over that 10 day period, only 17.9% contained data about food security, population movement, refugee camps and humanitarian access that was useful to the WFP.

The crowdsourced data obtained was limited and hence the platform was not utilised further as part of WFPs' crisis mapping operations for the November 2012, Gaza crisis situation. Platform analysis results show that this may be due to the following limitations:

a) The low volume of data geovisualised with an average of 3.90 data records geovisualised per day,
b) The low level of spatial data handling capability typically required for time poor crisis mapping operations as there is no data download function available on this platform,
c) The low degree of crowd participation and resultant collaborative knowledge,
d) The low degree of useful data which meets WFPs' geospatial data needs with 82.1% of the 39 data contained crisis information that would be of no benefit to WFPs crisis mapping operations and
e) The low degree of local knowledge and actionable data. This platform acts more as a news article aggregator rather than a harvester of local knowledge for decisions and actions.

3.1.2 Israel and Gaza Tracking Social Media crowdsourcing platform

The Tracking Social Media crowdsourcing platform for Israel and Gaza (AlJazeera 2012) is based on Ushahidi and was able to source data via both web forms and twitter feeds (Figure 2). Of the resulting data captured over the same 10 day period as for the Palestine Crisis Map, 23.3% (as identified through a random sample) were useful to the WFP. This platform (AlJazeera 2012) was not used as part of WFPs' crisis mapping operations for the November 2012, Gaza crisis situation due to the following summary of limitations:

a) The low level of spatial data handling capability. Whilst a large volume of crisis data was geovisualised within a small timeframe (an average of 36.40 data records geovisualised per day), this crowdsourcing platform did not have capability to download or easily acquire the data,
b) 199 of the 364 data could not be verified. The unverified data contains a level of uncertainty,
c) The large portion of data which did not meet WFPs' geospatial data needs. A low level of actionable data crowdsourced through this platform with 76.6% of the 73 sampled data of no relevance to WFPs' existing crisis mapping operations. Low positional accuracy of this data also results in a low level of actionable data, as the geographic locations are not specific enough for accurate delivery of food aid.
d) Low positional accuracy was observed as some of the data was attributed with multiple locations. E.g. One point has the geographic location within the Gaza Strip on the platforms' geovisualisation, however the point was attributed with the textual location of London.

Crowdsourcing Application Summary	
Platform:	Web 2.0 Ushahidi platform
Crowdsourcing Method:	Voluntary and Involuntary crowdsourcing through; iPhone and Android applications, email and the Ushahidi web form.

Crowdsourced Crisis Data Analysis Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Web Form	SMS	Email	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
10	39	3.90	39	0	0	0	2.56	2.56	10.26	2.56	38	1

Figure 1: Summary of the Palestine Crisis Map crowdsourcing platform

Crowdsourcing Application Summary	
Platform:	Web 2.0 Ushahidi Crowdmap platform
Crowdsourcing Method:	Voluntary and Involuntary crowdsourcing through; SMS, iPhone and Android applications, email and the Ushahidi web form.

Crowdsourced Crisis Data Analysis Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Web Form	SMS	Email	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
10	364	36.40	213	0	0	151	2.74	0.00	1.37	19.18	165	199

Results reflect analysis performed on a 20 % random sample of total data records.

Figure 2: Summary of the Israel and Gaza Tracking Social Media crowdsourcing platform

3.2 Philippine Typhoon Bopha

Typhoon Bopha hit the Philippines on the evening of the 3rd December 2012 and caused a crisis event affecting a great proportion of the population. This crisis situation is a sudden onset crisis situation resulting from a natural

disaster and therefore is required to be sustained to a high degree over a short period of time. With 6.2 million people in the Philippines affected (UNOCHA 2012c) by Typhoon Bopha, the magnitude of this crisis situation is great. At the time of conducting this research, four crowdsourcing applications were discovered in relation to this crisis situation including; the Philippine Disaster Watch (Ushahidi 2012b), the Super Typhoon Pablo (Ushahidi 2012c), the Typhoon Pablo Google Crisis Map (Digital Humanitarian Network 2012) and DOST Nationwide Operation Assessment of Hazards (Philippines Government 2012). The analysis period for each application was determined based on the period from which Typhoon Bopha hit the Philippines (03/12/2012) until when the last crowdsourced record was generated at the time of compiling this research.

3.2.1 Philippine Disaster Watch crowdsourcing platform

The Philippine Disaster Watch crowdsourcing platform (Ushahidi 2012b) was based on the Ushahidi system and utilised during the period when the disaster struck (Figure 3). The 10 responses received were obtained via web forms and provided information on population movements; 90% of the data was useful. However, an analysis of the limitations for use in the WFP crisis mapping operations reveals the following:

a) The low positional accuracy suggested by the data is attributed with multiple locations. E.g. One point has the textual location of Bislig City Airport (BPH), Bislig, Philippines but the geographic location of Mangagoy, Philippines. This creates uncertainty in the data.
b) The low volume of data geovisualised with an average daily crowdsourcing rate of 5 data records geovisualised per day,
c) The low degree of relevance of this data in meeting WFPs' geospatial data needs with 0% of crowdsourced data relating to food security, refugee camps or humanitarian access,
d) Despite the geovisualised data having a high degree of actionable data, due to the low level of attribute accuracy (in the form of accurate metadata) this actionable data is not reliable.
e) Most data was able to be captured by other means and hence crowdsourcing was largely redundant.

The majority of the crisis information crowdsourced through this platform were extracted from the National Disaster Risk Reduction and Management Council (NDRRMC) reports. WFP is aware of the NDRRMC reports and already uses these reports during existing WFP crisis mapping operations. As the NDRRMC reports are published every 6 hours, there is a high probability that more up-to-date information is available and this crowdsourced data is potentially redundant.

Crowdsourcing Application Summary	
Platform:	Web 2.0 Ushahidi platform
Crowdsourcing Method:	Voluntary and Involuntary crowdsourcing through; iPhone and Android applications, email, social media and the Ushahidi web form. Twitter (via a specified hashtags).

Benefits and Limitations Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Web Form	SMS	Email	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
2	10	5.00	10	0	0	0	0.00	90.00	0.00	0.00	10	0

Figure 3: Summary of the Philippine Disaster Watch crowdsourcing platform

3.2.2 Super Typhoon Pablo Crowd Mapping crowdsourcing platform

Through use of the Super Typhoon Pablo crowdsourcing platform (Ushahidi 2012c) 27 responses were crowdsourced through Ushahidi web maps (Figure 4). Of the responses capturing a range of ER data, 85.2 % was

considered useful. An analysis of this data in the context of this Philippines disaster identified the following limitations;

a) Whilst the spatial data handling capability associated with platform is of benefit (data downloadable in csv and kml formats), the low volume of data geovisualised through crowdsourcing methods is a limitation. With an average of 9 data records geovisualised on a daily basis, the 9 data records is not a large enough measure to aid in developing accurate situational awareness for such a large-scale sudden onset crisis situations.
b) Of the 8 geovisualised crisis data responses relating to location or movement of population, 5 of these were attributed with data extracted and copied directly from the NDRRMC reports. As previously indicated, the NDRRMC reports are published every 6 hours with a high probably of more up-to-date information being available and hence this crowdsourced data is potentially redundant.
c) Whilst this crowdsourcing platform shows potential in crowdsourcing geospatial data to meet WFPs' geospatial data needs, the low volume of data meant that this platform is of no direct benefit to WFPs' existing crisis mapping operations. Only 14.8% of the 27 data records contained information that was not relevant to WFP crisis mapping operations.
d) Only 7 of the 27 data contained local knowledge. E.g. one point included mentions local weather conditions and the possibility of flooding near the banks of the Agusan River (e.g. a local reference).

3.2.3 Typhoon Pablo (Bopha) Google Crisis Map crowdsourcing platform

The Typhoon Pablo Google Crisis Map (Digital Humanitarian Network 2012) was generated using a Google Map platform (Figure 5) and generated 101 crowdsourced responses over a 4-day period. On 5th December the Digital Humanitarian Network (DHN) was activated by UNOCHA following the impact of Typhoon Bopha which impacted the Philippines on the evening of 3rd of December 2012. The Standby Volunteer Task Force (SBTF) which consists of a crowd of volunteers was activated at this time, as SBTF is part of the DHN. Twitter was determined by UNOCHA as a source of crisis information. Through involuntary crowdsourcing methods, the task outsourced to the SBTF was to acquire and process video and photo crisis data from Twitter and publish this data to a Google Crisis map to assess the damage caused by Typhoon Bopha. The SBTF crowd was required to geo-tag, categorise, generate and attach metadata and time stamp the acquired crisis data.

Crowdsourcing Application Summary	
Platform:	Web 2.0 Ushahidi platform
Crowdsourcing Method:	Voluntary and Involuntary crowdsourcing through; iPhone and Android applications, email, social media and the Ushahidi web form. Twitter (via a specified hashtags).

Crowdsourced Crisis Data Analysis Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Web Form	SMS	Email	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
3	27	9.00	27	0	0	0	3.70	29.63	0.00	51.85	26	1

Figure 4: Super Typhoon Pablo Crowd Mapping crowdsourcing platform

The following limitations were identified:

a) The platform geovisualised a large volume of crisis data over a short period of time (with an average of 25.50 data records geovisualised per day) and has high degree of spatial data handling capability. However the crowdsourced data was not downloaded in this instance due to the low degree of relevance of the data to WFPs' crisis mapping operations and geospatial data needs. 64.71% of the 101 data records were not of any use to WFPs' crisis mapping operations. For example, only 7 data records were relevant to food

security. WFP cannot accurately make decisions based on 7 data records as there is high probability that this is not a true measure of the entire population affected by the crisis event.

b) The 101 data records crowdsourced through this platform were photo and video multimedia data extracted by the SBTF from news websites, Twitter and YouTube. It would not be practical for WFP to view all the photo and video links during crisis mapping operations due to time constraints. This platform has the potential to be of benefit in giving a broad overview of the crisis situation. However to analyse the data at a more detailed scale, (such as viewing the multimedia) would be impractical during the critical time dependent crisis mapping operations.

c) Positional accuracy could be improved since some of the crowdsourced data records were inaccurately geotagged. E.g. a point with obvious low positional accuracy is a point containing the attributes 'Large scale housing damage' which was located within the Bohol Sea instead of on land.

Crowdsourcing Application Summary	
Platform:	Web 2.0 Google Crisis Map platform by Google Crisis Response
Crowdsourcing Method:	• This application is an example of where involuntary crowdsourcing is used to outsource the data generation stage of the Geovisualisation Production Process. • Voluntary crowdsourcing has also been used to complete the data processing stage of the Geovisualisation Production Process. The United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) employed the Digital Humanitarian Network to curate, process and publish the crowdsourced multimedia crisis data from social networking sites. • There are no functionalities which allow the crowd to voluntarily participate with the crowdsourcing application in generating crisis data, as there is no availability of a platform web form.

Crowdsourced Crisis Data Analysis Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Other (including News websites)	YouTube	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
4	101	25.50	50	26	26	6.86	15.69	0.00	16.67	No verification	

Figure 5: Summary of the Typhoon Pablo (Bopha) Google Crisis Map crowdsourcing platform

3.2.4 DOST Nationwide Operation Assessment of Hazards crowdsourcing platform

Figure 6 shows the crowdsourced map of the DOST Nationwide Operation Assessment of Hazards platform (Philippines Government 2012) implemented using Web 2.0 technology. Only a small number of responses were received over a 2-day period with 62.5 % of them providing useful data. However, the information was not considered appropriate for the WFP's crisis mapping operation due to the following limitations:

a) The very low crowd participation and resultant collaborative knowledge with an average of 4 data records geovisualised per day. For the magnitude of this crisis, the low volume of crisis data would not add any extra benefit to existing WFP crisis mapping operations. However this platform has spatial data handling capability with the crowdsourced data can be downloaded to json format. The crowdsourced data records once downloaded are attributed with the coordinates, date, time and with a numerical value which corresponds to specified flood heights e.g. ankle high.

b) Five of the 8 responses contained potentially actionable crisis information, as they contained information relating to flood heights and therefore humanitarian access. However none of the 8 data met WFPs' geospatial data needs in terms of food security, movement or location of population or refugees.

Crowdsourcing Application Summary	
Platform:	Web 2.0 platform
Crowdsourcing Method:	• Originally suspected as a crowdsourcing application due to the Twitter feed present on the Google maps API. However, on closer inspection it appears that the Twitter feed contain Tweets which are tweeted by two official organisations rather than a large crowdsourcing crowd. • Whilst this application is focused on the disaster management aspect of emergency response, there is a function on this platform which allows for crowdsourcing the data generation task of the GPP. • Within this platform there is an application which allows the crowd to generate flood crisis data through voluntary crowdsourcing methods. • This function allows the user to search for and pinpoint a geographic location where the flood event is occurring and submit this flood event location.

Crowdsourced Crisis Data Analysis Summary

Analysis Period (days)	# of Crowdsourced Data Records	Average Daily Crowdsourcing Rate	Web Form	SMS	Email	Twitter	Food Security (%)	Movement/Location Population (%)	Refugee Camp (%)	Humanitarian Access (%)	Verified	Unverified
2	8	4.00	8	0	0	0	0.00	0.00	0.00	62.50	0	0
				No SMS, Email or Twitter functionality							No verification information	

Figure 6: Summary of the DOST Nationwide Operation Assessment of Hazards crowdsourcing platform

4 Study results and discussion

A total of six crowdsourcing platforms involving the November 2012 Gaza and the December 2012 Philippines Typhoon Bopha crisis situations were assessed in this study. None of the six crowdsourcing platforms were used for the crisis mapping operations undertaken by WFP. Each of the crowdsourcing applications was evaluated in terms of the following criteria: time, local knowledge, collaborative knowledge and crowd participation, positional accuracy, attribute accuracy, big data (spatial data handling capability), actionable data and whether or not the data meets WFPs' geospatial data needs. This criteria was determined following a literature review on the technological limitations and benefits of crowdsourcing for crisis mapping.

Figure 7 summarises the benefits and limitations of the six assessed crowdsourcing applications with regards to these criteria. Given that timeliness of appropriate data at the ER phase of a crisis event is crucial, it is not surprising that most applications did not provide sufficient data in a timely manner. Only the Tracking Social Media from Israel and Gaza (AlJazeera 2012) and the Typhoon Pablo (Bopha) Google Crisis Map (Digital Humanitarian Network 2012) crowdsourcing applications produced data that were timely. However, the data that was acquired through these applications was of a low level of relevance to WFP's crisis mapping, geospatial data

needs. The one platform of the six analysed that did meet WFP's geospatial data needs (Ushahidi 2012c) did not geovisualise enough data to make this application of use, with only 27 data records generated over a three day period. Applications such as the Philippine Disaster Watch crowdsourcing platform (Ushahidi 2012b) and the Super Typhoon Pablo Crowd Mapping crowdsourcing platform (Ushahidi 2012c) had a low degree of attribute accuracy with no metadata associated to the crowdsourced data content. This lead to uncertainties and made the data unreliable.

Even if simply considering the overall number of benefits and limitations across the six applications, the total of 34 limitations outweighs the 14 benefits identified. It must be noted here that this is one simplistic view and that the analysis presented within this paper was based purely on one situation; WFPs' crisis mapping operations. The results reflect the analysis performed in association with the requirements of WFPs' emergency response crisis mapping operations. Assessment of the eight criterion used in this study will vary in the number of benefits and limitations based on the crisis mapping situation that the crowdsourcing application is assessed against. A weighting could be applied to each of the eight criteria to reflect the level of requirement of the criteria in relation to the crowdsourcing application, and the crowdsourced outcome to the crisis situation.

Figure 7 shows that five of the six crowdsourcing platforms that geovisualised geospatial data do not meet WFPs' geospatial data needs (Criteria 8). This is a major limitation. The attribute and positional accuracy of the crowdsourcing platforms is also a main limitation with five of the six analysed crowdsourcing platforms containing uncertainties. Analysis of the six crowdsourcing applications determined that Criteria 6 showed the highest degree of benefit with three of the six crowdsourcing platforms having a high degree of spatial data handling capability. This is not consistent for similar platforms (such as the Ushahidi platform) where a data download functionality appears to be added based on the individual or organisations preference during creation and deployment of the platform. For the Google Crisis Map crowdsourcing platform (such as the Digital Humanitarian Network (2012)) a data download functionality appears to be a standard feature of the platform.

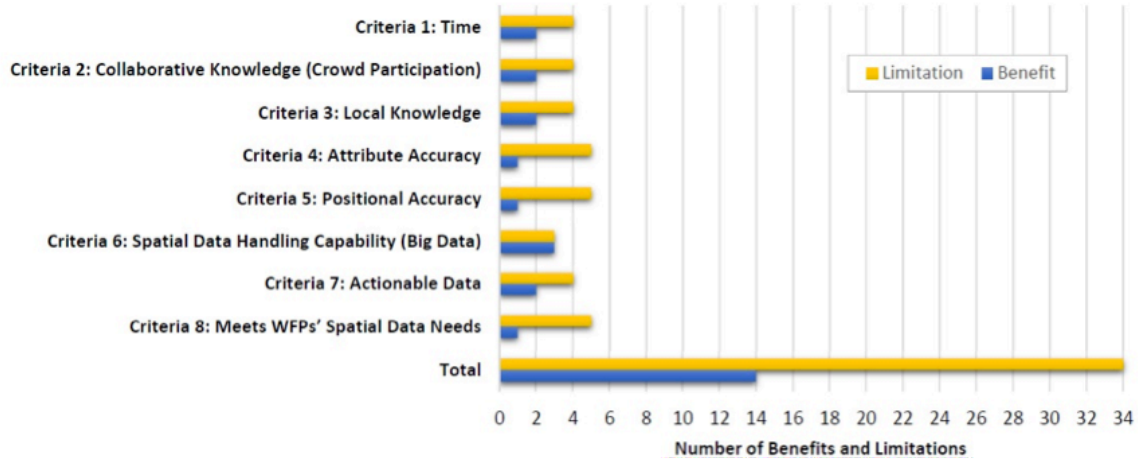


Figure 7: Limitations and benefits resulting from the analysis of different crowdsourcing platforms

5 Conclusions

A number of crowdsourcing applications were investigated in order to determine the benefit of crowdsourcing to the World Food Programmes' (WFP) existing crisis mapping operations. For each application, details regarding the crowdsourced information were extracted and recorded. This information was evaluated in regards to criteria that define the essential requirements for WFP's crisis mapping operations. As WFP responds to a variety of both slow and sudden-onset crisis situations in over 75 countries in which the most vulnerable populations are located, crowdsourcing technology needs to be highly flexible in order to be adopted by WFP. Thesis research identified WFP crowdsourcing capability requirements including; acquisition and sustainability of crowd participation, acquisition of local knowledge (despite a countries level of economic development), achievement of a high degree of attribute and positional accuracy, spatial data handling capability, being able to extract actionable data and to acquire data that meets WFPs' geospatial data needs.

The outcomes of this investigation demonstrate that the limitations outweigh the benefits of applying current crowdsourcing technologies to WFP's crisis mapping operations. This is primarily due to the inability of current crowdsourcing techniques and systems to capture required data in sufficient quantities and of sufficient quality in a timely manner. Crowdsourcing technology appears to be in its infancy at present and will need to develop further before it will be adopted as part of WFPs' existing crisis mapping operations.

Despite not discovering an immediate crowdsourcing solution to be adopted by WFP, this work will lead to further research and development which will be of greater benefit to humanitarian, crisis mapping operations. The

limitations associated with crowdsourcing technology as highlighted in this research will steer future technological developments. For example, highlighting countries where WFP has a presence and the related degrees of information and communication vulnerability will steer technological development in the direction of advancement for developing countries. It is hoped that this research will lead to technological development in terms of a crowdsourcing application that meets; critical time frames for both slow and sudden-onset crisis situations, sustained and consistent crowd participation, affected population participation (which is not dependent on a countries economic development), practical and accurate attributes, accurate positional accuracy, effective spatial data handling capability, an ease in extraction of actionable information and an application which meets the geospatial data needs required for humanitarian crisis mapping operations.

Acknowledgements

Thank you to Kashif Rashid and the GIS unit within the Emergency Preparedness and Response branch of the United Nations World Food Programme (WFP). The support given by Kashif and the GIS unit during the WFP practical placement made researching an ease. I would also like to thank Prof B. Veenendaal for support as research project supervisor. Thank you to my employer, Sinclair Knight Merz for granting me (Sophie Richards) the Bruce Sinclair Scholarship to undertake the postgraduate research within the WFP headquarters in Rome, Italy.

References

AlJazeera 2012, *Tracking Social Media from Israel and Gaza* viewed 10 December 2012, http://www.aljazeera.com/indepth/interactive/2012/11/20121116121728820347.html.

Aten, J 2011, 'Everyday Technologies for Extraordinary Circumstances: Possibilities for Enhancing Disaster Communication', *Psychological Trauma: Theory, Research, Practice and Policy*, vol. 3, no. 1, pp. 16-20.

Digital Humanitarian Network, the United Nations Office for the Coordination of Humanitarian Affairs 2012, *Google Crisis Map - Typhoon Pablo (Bopha),* viewed 27 December 2012, http://google.org/crisismap/2012-pablo.

Goodchild, M & Glennon, J 2010, 'Crowdsourcing Geographic Information for Disaster Response: A Research Frontier', *International Journal of Digital Earth*, vol. 3, no. 3, pp. 231-241.

Greenwald, T 2010, '35 Innovators under 35: Who Will Be the Next Helen Greiner, Mark Zuckerberg, Larry Page, Evan Williams, Jonathan Ive, Marc Andreessen, Daniel Schrag, Sergey Brin, Max Levchin?', *Technology Review*, vol. 113, no. 5, pp. 43-47.

Hirth, M, Hoßfeld, T & Tran-Gia, P 2012, 'Analyzing Costs and Accuracy of Validation Mechanisms for Crowdsourcing Platforms', *Mathematical and Computer Modelling*, vol. no. n.p.

Philippines Government, Department of Science and Technology Agency 2012, *Dost Nationwide Operation Assessment of Hazards,* viewed 27 December 2012, http://nababaha.com/report/input.php.

Rutsaert, P, Regan, Á, Pieniak, Z, McConnon, Á, Moss, A, Wall, P & Verbeke, W 2013, 'The Use of Social Media in Food Risk and Benefit Communication', *Trends in Food Science & Technology*, vol. 30, no. 1, pp. 84-91.

UNFAO, Food Agriculture Organization of the United Nations 2006, *The State of Food and Agriculture 2006,* Rome.

UNOCHA, United Nations Office for the Coordination of Humanitarian Affairs 2012a, *Occupied Palestinian Territory: Escalation in Hostilities. Gaza and Southern Israel, Situation Report (as of 26 November 2012, 1500 Hrs),* viewed 20 January 2013, http://www.ochaopt.org/documents/ochaopt_gaza_sitrep_26_11_2011_english.pdf.

UNOCHA, United Nations Office for the Coordination of Humanitarian Affairs 2012c, *Philippines: Typhoon Bopha, Situation Report No. 12 (as of 31 December 2012),* viewed 20 January 2013, http://reliefweb.int/sites/reliefweb.int/files/resources/OCHA%20Philippines%20Typhoon%20Bopha%20Situation%20Report%20No.%2012%2C%2031%20December%2C%202012.pdf.

Ushahidi 2012a, *Palestine Crisis Map,* viewed 10 December 2012, https://bindup.crowdmap.com/main.

Ushahidi 2012b, *Philippine Disaster Watch - (Typhoon Pablo: Bopha),* viewed 27 December 2012, https://philippinedisasterwatch.crowdmap.com/main.

Ushahidi 2012c, *Super Typhoon Pablo Crowd Mapping,* viewed 27 December 2012, https://supertyphoonpablo.crowdmap.com/main.

Vivacqua, A & Borges, M 2011, 'Taking Advantage of Collective Knowledge in Emergency Response Systems', *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 189-198.

WFP, The World Food Programme 2012a, *World Food Programme. Fighting Hunger Worldwide,* viewed 12 February 2013, http://home.wfp.org/stellent/groups/public/documents/communications/wfp215812.pdf.

WFP, The World Food Programme 2013a, *United Nations World Food Programme - Fighting Hunger Worldwide,* viewed 19th January 2013, http://www.wfp.org/about.

Dynamic Datum Transformations in Australia and New Zealand

Nic Donnelly
National Geodetic Office
Land Information New Zealand
Wellington, New Zealand
CRC for Spatial Information
ndonnelly@linz.govt.nz

Dr. Chris Crook
National Geodetic Office
Land Information New Zealand
Wellington, New Zealand
CRC for Spatial Information
ccrook@linz.govt.nz

Joel Haasdyk
Survey Infrastructure and Geodesy
NSW Land and Property Information
Bathurst, NSW
CRC for Spatial Information
Joel.Haasdyk@lpi.nsw.gov.au

Dr. Craig Harrison
Geoscience Australia
Canberra, ACT
CRC for Spatial Information
Craig.Harrison@ga.gov.au

Dr. Chris Rizos
Civil and Environmental Engineering
University of New South Wales
Sydney, NSW
CRC for Spatial Information
c.rizos@unsw.edu.au

Dr. Craig Roberts
Civil and Environmental Engineering
University of New South Wales
Sydney, NSW
CRC for Spatial Information
c.roberts@unsw.edu.au

Richard Stanaway
Civil and Environmental Engineering
University of New South Wales
Sydney, NSW
CRC for Spatial Information
richard.stanaway@student.unsw.edu.au

Abstract

Dynamic datums, which account for crustal dynamics, are widely used in Australia and New Zealand, although in many cases users may not be aware that they are using such a datum. The most widely-used global dynamic datums in Australia and New Zealand are the International Terrestrial Reference Frame (ITRF) and the World Geodetic System 1984 (WGS84). The use of WGS84 presents particular challenges, as it cannot be directly accessed at the sub-metre level by most spatial professionals. The relationship between WGS84 and ITRF is explored in detail, leading to recommendations on how users should deal with WGS84 datasets. New Zealand and Australia have taken different approaches to modelling the relationship between global dynamic datums and their local geodetic datums. Australia has calculated a set of local transformation parameters, which incorporate tectonic movements. New Zealand has adopted a set of global parameters, and uses a deformation model to account for crustal dynamics. As both countries seek to modernise their geodetic datums, the availability of simple yet accurate transformations will be critical to the success of the modernisation process. This paper outlines the options for dynamic datum transformations within a modernised datum, discussing the strengths and weaknesses of each approach.

1 Introduction

Global positioning technologies have driven a profound change in the way in which geospatial data is collected and referenced. There is an ongoing shift away from relative measurements, made in local coordinate reference systems, to absolute positioning, made in terms of global coordinate reference systems. Global Navigation Satellite Systems (GNSS), of which the U.S. Global Positioning System (GPS) is the most widely used, provide positions in a global geocentric reference frame. Such positions may be regarded as three-dimensional position vectors originating at the centre of mass of the Earth, and terminating at the physical feature or point being measured with the GNSS device.

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

The points being measured are attached to the surface of the Earth, which is continually moving due to crustal dynamics and other localised deformations. Thus if a point is regularly re-measured, its coordinates will change to reflect the reality that the point's relationship to the centre of the Earth has changed due to crustal dynamics. A reference frame or datum which enables this changing position to be tracked is referred to as 'dynamic'.

The two most widely used global dynamic reference frames in Australia and New Zealand (and much of the rest of the world) are the International Terrestrial Reference Frame (ITRF) and the World Geodetic System 1984 (WGS84). There are in fact a number of versions (or realisations) of each of these reference frames, and correct management of the various realisations is critical for high-precision positioning applications.

It is common practice for individual countries or geographic regions to define their own geodetic datums. These datums are often mandated for use in activities such as national mapping and cadastral surveying. Even where their use is not mandated, the national datum is usually the most appropriate reference frame to use for geospatial data, as that data is then easily integrated or compared with other datasets using the same datum. For Australia the national datum is the Geocentric Datum of Australia 1994 (GDA94) (ICSM, 2013a). This is a static datum aligned to ITRF92 at epoch 1994.0. For New Zealand, the national datum is New Zealand Geodetic Datum 2000 (NZGD2000) (Blick, 2003). NZGD2000 is referred to as a 'semi-dynamic' datum, as it incorporates a deformation model to refer measurements and coordinates to a common epoch. It is aligned with ITRF96 at epoch 2000.0.

GNSS measurements are often made in terms of a global reference frame, but presented as a derived final product in terms of the national geodetic datum. An important component of a nation's positioning infrastructure is the determination and maintenance of transformations from commonly used global reference frames to the national geodetic datum. A coordinate transformation is a mathematical operation that enables coordinates expressed in one reference frame (or datum) to be represented in terms of another (ISO, 2008). It is distinguished from coordinate propagation, which enables coordinates referenced to a particular epoch to be represented at another epoch, within the same reference frame. Official transformations in New Zealand and Australia focus on the relationship between the various ITRF realisations and the national datum. Conversely, despite the widespread use of WGS84, transformation parameters from WGS84 coordinates to the national datums are generally not widely known. This is due to factors such as the inaccessibility of WGS84 as a reference frame for sub-metre positioning (since there is a lack of accurate WGS84 ground control for relative positioning), and the consequent relatively high uncertainties associated with WGS84 datasets. Nevertheless, accurate transformations between WGS84 realisations and national datums can be determined via the relationship between ITRF and WGS84. For some datasets, knowing these transformations may be important.

The concept of 'uncertainty' is an important, but often overlooked, aspect of coordinate transformations. When considering positioning in terms of global reference frames, uncertainty of Cartesian coordinates is the relevant measure. This concept is referred to as 'positional uncertainty' in Australia (ICSM, 2013b), and 'network accuracy' in New Zealand (LINZ, 2009). There is usually little value in performing a transformation if the accuracy of a geospatial dataset is low, and consequently the uncertainty of the coordinates is substantially greater than the coordinate change determined from those transformation parameters. On the other hand, the uncertainty resulting from the process of the transformation itself is also rarely accounted for (Haasdyk & Janssen, 2011), and more attention will need to given to this process as measurements and datums improve in accuracy.

This paper describes the dynamic global reference frames used in Australia and New Zealand, with a focus on WGS84, which is widely used but frequently misunderstood. It reviews the transformation parameters used in both countries between ITRF and the national datum, extending these for use with WGS84 datasets. Finally, as both Australia and New Zealand work towards modernisation of their national datums, options for dynamic datum transformations are considered in terms of future trends in positioning and geodetic datums.

2 The International Terrestrial Reference Frame (ITRF)

The International Earth Rotation and Reference Systems Service (IERS) is responsible for the maintenance and development of the International Terrestrial Reference Frame (ITRF), upon which all modern national geodetic datums are based. A reference frame is a realisation, according to a set of agreed conventions (Petit and Luzum, 2010), of the idealised International Terrestrial Reference System (ITRS).

A terrestrial reference system is one which is co-rotating with the Earth about a conventional pole (ibid, 2010). Various models are applied to remove the impact of phenomena such as ocean loading and earth tides, with the effect that the coordinates of a point on the Earth's surface do not change in response to these geophysical effects (ibid, 2010). The reference system needs to be very stable for monitoring long-term global phenomena such as sea-level change.

The latest realisation of the ITRF is ITRF2008 (Altamimi *et al*., 2011). A realisation of ITRF produces estimates of station coordinates and linear velocities for hundreds of stations worldwide. ITRF is now stable at the centimetre-level or better (ibid, 2011). Transformation parameters between the two most recent realisations of the ITRF (ITRF2008 and ITRF2005) are at the sub-centimetre level, a level of agreement which is likely to continue (or improve) for future realisations.

There is also a global tectonic plate model published in conjunction with ITRF2008 (Altamimi *et al*., 2012) which may be used to propagate coordinates between epochs using the defined Euler pole for 14 major tectonic plates. This model is a No-Net-Rotation (NNR) model derived from the station velocities published for ITRF2008, which aligns the orientation of all ITRFs to each other, and to the available geophysical models. Such a plate motion model works well for most of the Earth, where it is assumed that tectonic plates are non-deforming (that is, they are rigid plates that rotate about a point). This assumption is not valid near plate boundaries, which is the case for New Zealand, or where high accuracy over large distances is required, which is the case for all national geodetic datums. Both New Zealand and Australia have their own models for propagating coordinates, and do not directly use the ITRF2008 plate motion model.

The majority of users access the ITRF through GNSS technology, and associated products and services. The most precise products (such as satellite orbits) are produced by the International GNSS Service (IGS). The IGS produces its own ITRF-aligned reference frame, the current realisation being aligned to ITRF2008 and denoted IGb08 (IGS, 2010). The high degree of alignment between these frames means that they can be considered identical for all but the highest precision geodetic applications.

3 The World Geodetic System 1984 (WGS84)

The term 'WGS84' is one of the more ambiguous in global geodesy, which sometimes leads to confusion about the exact nature of this datum. Firstly, depending on context, WGS84 may refer to a reference system, a reference frame or a reference ellipsoid, which are each defined in more detail below. Secondly, even where it is clear from the context that it is the reference frame WGS84 which is the subject of discussion, the fact that there have been several different WGS84 reference frames is often not made clear. In general all are simply referred to as 'WGS84'.

WGS84 is managed by the National Geospatial Intelligence Agency (NGA), formerly the National Imagery and Mapping Agency (NIMA), which was itself the successor to the Defense Mapping Agency (DMA). The DMA was responsible for developing the WGS84 reference system and for the initial reference frame realisation.

3.1 WGS84: the reference system

The WGS84 reference system is designed to coincide as closely as possible with the ITRS (NIMA, 2000). The origin is the centre of mass of the Earth, scale is that of the local Earth frame (in the sense of the theory of general relativity) and initial orientation was consistent with that defined by the Bureau International de l'Heure (BIH) in 1984. The evolution of the orientation in time is such that there is no overall rotation of the system with respect to the Earth's surface (NIMA, 2000).

WGS84 is a right-handed, orthogonal reference system. Its X-axis is consistent with the IERS Reference Meridian (which is approximately equal to the Greenwich Meridian), the Z-axis is coincident with the IERS Reference Pole (which is approximately the geographical North Pole), and the Y-axis is oriented at ninety degrees with respect to the other two axes (NIMA, 2000).

It is the consistency of definition between the WGS84 and ITRS reference systems which results in the high levels of consistency between coordinates realised in the WGS84 and ITRF reference frames.

3.2 WGS84: the ellipsoid

The WGS84 ellipsoid is a reference surface approximating the size and shape of the Earth. Its origin is the centre of mass of the Earth, and it is formed by rotating an ellipse about the Z-axis. It is almost identical to the Geodetic Reference System 1980 (GRS80) ellipsoid associated with the ITRF. The slight difference in flattening is due to a truncation during the computation of the WGS84 flattening value, but is insignificant for all but the highest precision geodetic applications (NIMA, 2000). The difference in ellipsoidal heights calculated using the two ellipsoids is a maximum of just 0.1mm, at the poles. Table 1 lists the key parameters of the two ellipsoids.

Table 1: GRS80 and WGS84 ellipsoid parameters

Ellipsoid	Semi-major axis (m)	Inverse flattening
GRS80	6 378 137 m	298.257 222 101
WGS84	6 378 137 m	298.257 223 563

3.3 WGS84: the reference frames

A WGS84 reference frame is a realisation of the WGS84 reference system through the defined coordinates (and more recently, velocities) for a set of reference stations (Wong *et al*., 2012). These reference stations and their coordinates are then used for applications such as calculating the orbits of GPS satellites. It is through orbital products such as the GPS broadcast ephemeris that the geospatial community is able to access the WGS84 reference frame. As at the start of 2014, the current realisation of the WGS84 system is the fifth such realisation.

The first realisation is the only one officially denoted as 'WGS84'. To avoid confusion, in the remainder of this paper, we denote this first WGS84 reference frame as 'WGS84(Doppler)', since it was based on the TRANSIT Doppler system, the predecessor positioning system to GPS. Each subsequent WGS84 realisation is distinguished by appending the GPS week in which the reference frame was implemented (NIMA, 2000). For example, the current realisation is denoted 'WGS84(G1674)', as it was implemented by the GPS Operational Control Segment (OCS) on 8 February 2012, which is GPS Week 1674 (Wong *et al*., 2012).

Wong *et al* (2012) provide full details of the current WGS84 realisation. This was aligned to ITRF2008 through direct adoption, where possible, of ITRF2008 coordinates published by the IERS for 11 WGS84 monitor stations distributed around the globe. Corrections were made to account for discontinuities due to activities such as equipment maintenance. For two stations, BHR2 in Bahrain and OSN1 in South Korea, the ITRF2008 coordinates could not be constrained without introducing large residuals, so these stations had new coordinates calculated. New coordinates were also calculated for the six United States Air Force sites which comprise the OCS. Velocities for the monitor stations were adopted from the ITRF2008 solution. For the OCS stations, velocities were adopted from a nearby International GNSS Service (IGS) station.

Table 2 summarises each of the five WGS84 reference frame realisations, indicating that since 2002 the positional uncertainty of the reference frame relative to the ITRF is about one centimetre, estimated based on the magnitude of similarity transformation parameters calculated between the ITRF and WGS84 precise orbits.

Table 2: WGS84 reference frames 1987-2013

Reference Frame	**Implementation Date (OCS)**	**Reference Epoch**	**Positional Uncertainty (m) (1-sigma) Relative to ITRF2008**	**Reference**
WGS84(Doppler)	1987, 1 January	NA	1.50	NIMA (2000)
WGS84(G730)	1994, 29 June	1994.0	0.10	NIMA (2000)
WGS84(G873)	1997, 29 January	1997.0	0.05	NIMA (2000)
WGS84(G1150)	2002, January	2001.0	0.01	Merrigan *et al*. (2002)
WGS84(G1674)	2012, 8 February	2005.0	0.01	Wong *et al*. (2012)

3.4 WGS84: An operational dynamic datum

WGS84 is probably the most widely used global reference frame. This popularity stems from its use as the reference frame for GPS orbits – the broadcast ephemeris used for GPS single point positioning. The reference frame of the GPS orbit determines the reference frame of the user position for this type of absolute positioning. On the 1 January each year, the coordinates of the GPS master control stations are propagated to an epoch in the middle of the year, and GPS satellite coordinates are determined relative to these master stations. Thus WGS84 coordinates determined by GPS single point positioning are related to the epoch of the middle of the calendar year in which the observations are made (ICG, 2012), not the reference epoch for the realisation. Thus a full reference for a dataset collected using GPS single point positioning during 2013 would be *WGS84(G1674) Epoch 2013.5*.

The dynamic nature of WGS84 is typically ignored by users who fail to account for the WGS84 reference frame realisation or the epoch of their derived coordinates in their metadata. Indeed it is practically impossible to rigorously reference WGS84 (and other dynamic datums) in most commonly used software packages. In the case of WGS84, ignoring the dynamic details does not usually compromise the data as the positional uncertainties associated with a WGS84 dataset collected via single point positioning are at the metre-level at best.

3.5 The myth of precise WGS84 coordinates

Outside the US military and other authorised users, direct access to sub-metre WGS84 coordinates is almost impossible. The NGA has published coordinates for only 11 tracking stations worldwide, with one in Australia and

one in New Zealand (Wong *et al*., 2012). The GPS precise ephemeris data is published, which in theory enables the use of Precise Point Positioning (PPP) to generate centimetre-accurate WGS84 coordinates. In reality though, very few users have the capability to do the required processing.

Within the geospatial community, therefore, one would expect precise datasets that are referenced to WGS84 to be exceedingly rare. This is not the case, largely due to misunderstandings about WGS84 and its role in GPS positioning. Accurate positioning using GPS almost always involves relative pseudorange or carrier phase positioning to achieve accuracies ranging from centimetres to a metre. In these techniques, some of the stations occupied have known coordinates, and it is the datum of these coordinates which determines the datum for the subsequent geospatial dataset. In Australia the datum would typically be GDA94, in New Zealand, NZGD2000. While many of these techniques do use the broadcast ephemeris during processing, it is not used as a source of coordinates, so does not determine the datum of the coordinates being generated.

The mislabelling of precise geospatial datasets as WGS84, when they are not WGS84, can cause problems if the data epoch is incorrectly assumed. For example, consider the case where a relative positioning dataset is generated in 2013 from a base station with GDA94 coordinates, but is referenced as WGS84. A future user of the dataset, knowing it was generated in 2013, might reasonably assume that the reference frame is WGS84(G1674) with an epoch of 2005.0 (which is the reference epoch for this particular WGS84 frame) or an epoch of 2013.5 (which is the epoch at which WGS84 coordinates are realised for that year). In fact the epoch for the data is 1994.0, which is derived from the GDA94 coordinates of the base-station, so a coordinate error of up to 0.8 m (due to the 7 cm/year tectonic motion of Australia) is immediately introduced.

A smaller systematic error is caused by the incorrect specification of the reference frame. Using the same example, the difference between WGS84(G1674) and GDA94 is nearly 0.1m in height. These problems are often time-consuming and difficult to identify with confidence. Users should be wary if a dataset is purported to be referenced to WGS84, and positional uncertainties are less than a metre.

Even the definition of WGS84 as a dynamic datum, while technically correct, is misleading in terms of the way users access the datum. From the user perspective, it is a series of epoch reference frames, each of which is at the mid-year epoch as discussed in section 3.4. Thus the 'dynamic' WGS84 coordinates derived from point positioning have up to half a year's worth of error due to tectonic movement in them. While this is not noticeable to users accessing WGS84 via single point positioning, it is further reason not to use WGS84 where precise coordinates are required.

4 Transformations Between ITRF, WGS84 and National Datums

The first WGS84 realisation preceded the first ITRF realisation. All subsequent realisations have been aligned to the ITRF through the use of stations with ITRF coordinates in the realisation of the WGS84 reference frame. The alignment between WGS84 and ITRF in each instance is sufficiently close that the two are considered identical, within the uncertainty of the WGS84 reference frame. Table 3 lists the ITRF reference frames to which successive WGS84 reference frames are aligned.

Table 3: Alignment of WGS84 and ITRF realisations

WGS84 realisation	ITRF realisation
WGS84(Doppler)	Not applicable
WGS84(G730)	ITRF91
WGS84(G873)	ITRF94
WGS84(G1150)	ITRF2000
WGS84(G1674)	ITRF2008

WGS84, by definition, tracks a particular realisation of the ITRF, therefore the official transformation parameters between the national datum and ITRF in each country can be used as a proxy for WGS84 parameters with careful attention being paid to the epoch of the WGS84 coordinates as previously discussed.

4.1 ITRF/WGS84 transformation parameters for New Zealand

The parameters in Table 4 should be used to transform from a particular ITRF or WGS84 reference frame to NZGD2000, which is aligned with ITRF96. To transform between reference frames, the transformation parameters must first be determined at the epoch of transformation, which in many cases will be the epoch of the coordinates being transformed. The parameter T_x at time t (in years) is calculated using:

$$\delta t = t - 2000 \quad (1)$$

$$Tx,t = Tx + \delta t . \dot{T}x \quad (2)$$

The transformation is then carried out using:

$$XYZNZGD2000=XYZITRF/WGS84+cTTx,tcTTy,tcTTz,t+cSSt-cRRz,tcRRy,tcRRz,tcSSt-cRRx,t-cRRy,tcRRx,tcSStXYZITRF/WGS84 \quad (3)$$

where

$c_T = 0.001$ (millimetres to metres, applies to T_x, T_y and T_z)
$c_S = 1.0\times10^{-9}$ (part-per-billion to ratio, applies to S)
$c_R = \pi/(180\times60\times60\times1000) = 4.84814\times10^{-9}$ (milli-arcseconds to radians, applies to R_x, R_y and R_z)

With one exception, these parameters are derived from those published by the IERS (IERS, 2013). The transformation parameters between ITRF96 and ITRF97 as calculated by the IGS (Soler and Snay, 2004) were used in preference to those specified by the IERS. The IERS had determined that no significant transformation existed between ITRF96 and ITRF97. The calculations by the IGS determined a non-zero transformation, which New Zealand has adopted since NZGD2000 is based principally on GNSS observations. However, at the time of writing, New Zealand is reviewing its use of the IGS values, which could lead to a change in the transformation parameters in Table 4.

Once the transformation has been carried out, the NZGD2000 deformation model would normally be used to propagate the coordinates to the reference epoch (2000.0).

Table 4: Transformations to NZGD2000

	T_x (mm) $\dot{T}_x$ (mm/yr)	T_y (mm) $\dot{T}_y$ (mm/yr)	T_z (mm) $\dot{T}_z$ (mm/yr)	S (ppb) $\dot{S}$ (ppb/yr)	R_x (mas) $\dot{R}_x$ (mas/yr)	R_y (mas) $\dot{R}_y$ (mas/yr)	R_z (mas) $\dot{R}_z$ (mas/yr)	Reference Epoch
ITRF2008/ WGS84(G1674)	4.8	2.09	-17.67	1.40901	-0.16508	0.26897	0.11984	2000
rates	0.79	-0.6	-1.34	-0.10201	-0.01347	0.01514	0.01973	
ITRF2005	6.8	2.99	-12.97	0.46901	-0.16508	0.26897	0.11984	2000
rates	0.49	-0.6	-1.34	-0.10201	-0.01347	0.01514	0.01973	
ITRF2000/ WGS84(G1150)	6.7	3.79	-7.17	0.06901	-0.16508	0.26897	0.11984	2000
rates	0.69	-0.7	0.46	-0.18201	-0.01347	0.01514	0.01973	
ITRF97	0	-0.51	15.53	-1.51099	-0.16508	0.26897	0.05984	2000
rates	0.69	-0.1	1.86	-0.19201	-0.01347	0.01514	-0.00027	
ITRF96	0	0	0	0	0	0	0	2000
rates	0	0	0	0	0	0	0	
ITRF94/ WGS84(G873)	0	0	0	0	0	0	0	2000
rates	0	0	0	0	0	0	0	
ITRF93	28.8	0.2	5.4	-0.49	1.71	1.48	0.36	2000
rates	2.9	-0.4	-0.8	0	0.11	0.19	-0.05	
ITRF92	-8	-2	8	0.71	0	0	0	2000
rates	0	0	0	0	0	0	0	
ITRF91/ WGS84(G730)	-20	-16	14	-0.69	0	0	0	2000
rates	0	0	0	0	0	0	0	
ITRF90	-18	-12	30	-0.99	0	0	0	2000

	T_x (mm) $\dot{T}_x$ (mm/yr)	T_y (mm) $\dot{T}_y$ (mm/yr)	T_z (mm) $\dot{T}_z$ (mm/yr)	S (ppb) $\dot{S}$ (ppb/yr)	R_x (mas) $\dot{R}_x$ (mas/yr)	R_y (mas) $\dot{R}_y$ (mas/yr)	R_z (mas) $\dot{R}_z$ (mas/yr)	Reference Epoch
rates	0	0	0	0	0	0	0	
ITRF89	-23	-36	68	-4.39	0	0	0	2000
rates	0	0	0	0	0	0	0	
ITRF88	-18	0	92	-7.49	-0.1	0	0	2000
rates	0	0	0	0	0	0	0	
WGS84(Doppler)	-78	505	253	10.01	-18.3	0.3	-7	2000
rates	0	0	0	0	0	0	0	

4.2 ITRF/WGS84 transformation parameters for Australia

The transformation parameters for Australia have been published in Dawson and Woods (2010). Note that the signs of the rotations in Table 5 have been reversed from those published in Dawson and Woods (2010), so that the values are consistent with Equation (3).

Table 5: Transformations to GDA94

	T_x (mm) $\dot{T}_x$ (mm/yr)	T_y (mm) $\dot{T}_y$ (mm/yr)	T_z (mm) $\dot{T}_z$ (mm/yr)	S (ppb) $\dot{S}$ (ppb/yr)	R_x (mas) $\dot{R}_x$ (mas/yr)	R_y (mas) $\dot{R}_y$ (mas/yr)	R_z (mas) $\dot{R}_z$ (mas/yr)	Reference Epoch
ITRF2008/ WGS84(G1674)	-84.68	-19.42	32.01	9.710	0.4254	-2.2578	-2.4015	1994
rates	1.42	1.34	0.90	0.109	-1.5461	-1.182	-1.1551	
ITRF2005	-79.73	-6.86	38.03	6.636	0.0351	-2.1211	-2.1411	1994
rates	2.25	-0.62	-0.56	0.294	-1.4707	-1.1443	-1.1701	
ITRF2000/ WGS84(G1150)	-45.91	-29.85	-20.37	7.070	1.6705	-0.4594	-1.9356	1994
rates	-4.66	3.55	11.24	0.249	-1.7454	-1.4868	-1.2240	
ITRF97	-14.63	-27.62	-25.32	6.695	1.7893	0.6047	-0.9962	1994
rates	-8.60	0.36	11.25	0.007	-1.6394	-1.5198	-1.3801	
ITRF96	24.54	-36.43	-68.12	6.901	2.7359	2.0431	-0.3731	1994
rates	-21.80	4.71	26.27	0.388	-2.0203	-2.1735	-1.6290	

4.3 International Case Studies

By way of comparison, two international case studies are presented below. The focus is on transformations from ITRF to the national datum, although using Table 3, transformations to and from the various WGS84 realisations could be inferred.

4.3.1 United States

The national datum in the United States is the North American Datum 1983 (NAD83). Like WGS84, this datum was originally based on TRANSIT Doppler observations and consequently is not strictly geocentric, being offset by approximately 2 metres (Soler and Marshall, 2003). NAD83 has been periodically re-realised to enable the incorporation of modern observation techniques such as GNSS, with the most recent realisation being NAD83(2011), which has an epoch of 2010.0 (Pearson and Snay, 2013).

NAD83(2011) is fixed to the North American plate, so that the velocities of stable points are minimised across the country. For most users, such as those working in central and eastern United States, station velocities in this plate-fixed frame are negligible. This allows those users to treat NAD83 as a static datum. For regions such as the western United States and parts of Alaska, the proximity to the plate boundary means that a simple plate motion model cannot accurately account for the more complex motions present. For these areas, more complex displacement models, including models which account for earthquakes, are required (Pearson and Snay, 2013).

Transformation parameters from ITRF96 to NAD83 were defined jointly by Canada and the United States using 12 Very Long Baseline Interferometry (VLBI) stations with coordinates in each datum. The NUVEL-1A plate motion model (De Mets et al, 1994) was used to determine the rotations for the North American plate (Soler and Snay, 2004). The IGS transformation is used from ITRF96 to ITRF97, but for all subsequent realisations of ITRF, the transformation parameter values published by the IERS are used. The total transformation from ITRF2008 to NAD83 is obtained by addition of these individual sets of transformation parameters (Pearson and Snay, 2013). This is identical to the approach currently taken in New Zealand, except that NZGD2000 is by definition aligned to ITRF96.

4.3.2 Great Britain

The national mapping datum for Great Britain is Ordnance Survey Great Britain 1936 (OSGB36). This is a local datum, with its ellipsoid positioned to best fit the Earth's surface over the landmass of Great Britain. Consequently, the fit to the Earth as a whole is relatively poor with the origin of OSGB36 being offset from the geocentre by over 700 metres. For high precision positioning applications, the official datum is the European Terrestrial Reference Frame 1989 (ETRF89), which is a realisation of the European Terrestrial Reference System 1989 (ETRS89). ETRF89, along with all subsequent ETRF realisations is fixed to the stable part of the Eurasian plate, in much the same way as NAD83 is fixed to the North American plate. This enables users to ignore the effects of tectonic plate motion (Ordnance Survey, 2013).

To transform from ITRF to OSGB36 requires that the coordinates first be transformed to ETRF89. This is achieved using the procedure recommended by the IAG Subcommission for the European Reference Frame (EUREF), as outlined in Boucher and Altamimi (2011). It involves a two-step procedure. Firstly, IERS-published published parameters are used to transform from the ITRF realisation of the coordinates to ITRF89. Secondly, three rotation rates (which account for plate motion) are applied to transform from ITRF89 to ETRF89.

Because OSGB36 is based on triangulation, distortions in the network mean it is not possible to transform coordinates from ETRF89 using a 7-parameter transformation to any better than 5 metres. Therefore, a gridded displacement model is used for this transformation (Ordnance Survey, 2013).

4.4 Differing approaches to dynamic datum transformations

Inspection of Tables 4 and 5 reveals substantial disparities in the numerical values used in Australia and New Zealand, which cannot be solely attributed to the different versions of ITRF and epochs to which each national datum aligns. These disparities are mainly due to the different approaches taken by New Zealand and Australia to define their official transformation parameters.

The New Zealand parameters keep the process of reference frame transformation quite separate from the process of coordinate propagation. Propagation of coordinates to or from the reference epoch of 2000.0 is achieved using the NZGD2000 deformation model.

In contrast, Australia has calculated a localised set of transformation parameters between ITRF and GDA94 which are only applicable on the Australian tectonic plate (Dawson and Woods, 2010). Transformation and propagation are carried out in a single process enabling ITRF/WGS84 data at any epoch to be transformed to GDA94 using a 14-parameter transformation. Note that a standard 14-parameter transformation assumes the same data epoch for both the input and output coordinates, but the Australian transformation is quite unique in that the GDA94 coordinates always refer to 1994.0. A user may validly choose to propagate their ITRF coordinates to epoch 1994.0 (e.g. using the ITRF2008 plate motion model) before or after the ITRF to GDA94 transformation, but will notice coordinate differences of up to 20 mm compared to a direct transformation at the current epoch (Haasdyk & Janssen, 2011).

The Australian continent is extremely stable, with very little relative deformation across its landmass. Thus the fourteen-parameter transformation is able to accurately incorporate the national-scale plate motion. From the user perspective, there is no need to utilise a deformation model and some geospatial software packages can handle fourteen-parameter transformations. Those that cannot perform a 14 parameter transformation usually at least have the capacity to carry out seven-parameter transformations, giving the user the opportunity to enter the appropriate seven-parameter values for the epoch at which the transformation is required.

This approach would not work well in New Zealand. Even at a national scale there is significant relative deformation that would not be modeled adequately using a simple fourteen-parameter transformation for coordinate transformation and propagation. Thus it is necessary to separate the processes of coordinate transformation and propagation. Currently, the lack of support for the deformation model in commercial software makes its use impractical for many users, who continue to rely on a dense network of passive marks which they can use to calculate local transformations to account for deformation.

The case studies of the United States and Great Britain further highlight the variation in transformation approaches, to reflect local circumstances. While crustal dynamics can be simply handled in Great Britain (as in Australia), the continued use of a local geodetic datum based on terrestrial measurements means that accurate

transformations to geocentric datums requires the use of a gridded model. The United States has taken the approach of satisfying the current desire of most users for a static datum by fixing NAD83 to the North American Plate, while still providing the necessary means for users in areas of significant deformation to accurately transform coordinates.

5 Transformation Options for Modernised National Dynamic Datums

Both New Zealand and Australia are investigating how best to modernise their datums so that they continue to support high accuracy positioning in each country. In Australia's case a new dynamic datum is proposed, to be released in approximately 2020. For New Zealand, modernisation may occur within the framework of the existing datum which already accounts for dynamics, or a new datum may be developed. For both countries, accurately representing the dynamics of the Earth's surface will be a key challenge, as will maintaining compatibility with international systems.

The development of modernised datums will follow established international conventions and utilise accepted global models, such as those specified in the IERS conventions (Petit and Luzum, 2010). This will ensure maximum consistency with preferred positioning methodologies such as GNSS, which also follow these conventions. As well as providing consistency, the use of established conventions and models provides a level of traceability to a datum, which flows to the positions derived.

5.1 Option 1: Local 14 parameter transformation, including national-scale tectonics, plus residual deformation model

This is an enhancement of the approach already implemented for GDA94. Currently GDA94 assumes that the Australian continent is stable and that coordinates do not change over time.

This option assumes that any future dynamic datum retains a single reference epoch. In this option, a 14-parameter transformation continues to be provided that includes both the reference frame transformation and propagation between ITRF and GDA, and between the epoch of the dataset and the reference epoch. A gridded deformation model can then be used to account for the residual deformation, which is not included in the 14-parameter transformation. Most users will only need to use the simpler and more widely supported 14-parameter transformation.

For New Zealand, such an approach is unlikely to provide a practical solution. Even national-scale deformation is so complex that the size of the residual deformation would be significant for most geospatial applications. Thus both the 14-parameters and the deformation model would need to be applied in the majority of cases.

In Australia, by contrast, the stable tectonic setting means that the residual deformation field will be insignificant for many applications, although it is noted that in the future, the accuracy demands of many applications is likely to increase as centimetre-level absolute positioning becomes mainstream. Applications that only require decimetre-level accuracy could continue to use the 14-parameter transformation, as they do currently. This option may present the simplest approach for a large proportion of Australian spatial users.

Disadvantages of this approach for Australia is that it may not be favoured internationally, as a number of countries are at least partially straddling plate boundaries (although 94% of the Earth's surface lies within the stable portion of a tectonic plate). In practical terms this may not matter, given that software needs to handle 14-parameter transformations for global reference frame transformations.

5.2 Option 2: Global 14 parameter transformation, excluding national-scale tectonics, plus deformation model

This is the approach currently used for NZGD2000.

The use of global transformation parameters provides the maximum level of consistency with, and traceability to, global standards and conventions, an important consideration for a national datum, particularly one which aims to support the use of global positioning technologies.

A future dynamic datum may use all available global ITRF stations in its processing, following the approach taken by regional reference frames such as the Asia-Pacific Reference Frame (APREF) (Haasdyk *et al.*, 2014). At the very least it will include a substantial subset of global stations distributed over a significant proportion of the globe. If this is the case, then the future national datum is explicitly aligned to a global network of stations. Logically therefore, a globally determined set of transformation parameters is appropriate.

With this approach, coordinate propagation is handled in totality by a deformation model, maintaining complete separation between transformation and propagation. The current disadvantage of this approach is that most software does not have the capacity to incorporate deformation models. This situation is likely to change in the coming years as the necessity for incorporating deformation models to fully utilise accurate geospatial data is better understood.

5.3 Option 3: Local 14 parameter transformation, excluding national-scale tectonics, plus deformation model

This is similar to Option 2, the difference being that rather than using a published set of parameters from the IERS, a local set is calculated using the ITRF stations in each country. The difference from Option 1 is that the 14 parameters are only accounting for the reference frame transformation – tectonic motion is accounted for by the deformation model. The advantage of this approach is that it enables any regional biases in ITRF to be accounted for in the transformation, increasing the alignment of the coordinates generated by the transformation with the ITRF stations in the region. However, the improving precision of successive realisations of the ITRF should make the calculation of local parameters unnecessary. Any local discrepancies are likely to be due to deformation not fully accounted for in the ITRF, which is more properly included within a deformation model. An example of such a local discrepancy is the 4 mm per year average residual rotation rate identified in Australia when comparing the ITRF plate motion model with the recently released NNR-MORVEL56 model (Altamimi *et al*, 2012).

It would still be worth calculating a local set of transformation parameters before making a decision against this option, to prove that they are not significantly different from the global parameters.

5.4 Option 4: Deformation model only

This option combines propagation and transformation in a variable resolution grid of site velocities and offsets. Stanaway and Roberts (2013) discuss this option in detail. A key advantage of this option is that coordinate transformation and propagation are combined, yet the full range of non-secular deformation can be included. Consequently, loss of precision when propagating uncertainty is minimised. If Australia were to define a refined GDA94, such a model would enable the propagation of these refined GDA94 coordinates with an uncertainty of 6 mm at the 95% confidence level (Stanaway and Roberts, 2013).

For New Zealand, this option has similar advantages, albeit that the more complex nature of the country's deformation compared with Australia means that sub-centimetre propagation uncertainties are not realistic. Given that New Zealand already utilises a deformation model, it would not be a major change to amend this model to incorporate components for reference frame transformation.

As for the other options involving deformation grids, the biggest disadvantage to this approach at the current time is the inability of software to support these models.

5.5 Option 5: Reduced parameter sets

Other options could involve variations in the number of parameters used in a transformation/propagation model. For example, a three-parameter transformation of three rotations and the reference epoch could be used to propagate coordinates to the geodetic datum. In this scenario, scale and translations are assumed to be null. With options such as this, increased simplicity is being traded off against decreased precision. For applications that only require limited precision, the accuracy provided by simpler options is likely to be perfectly adequate. Once again, this is particularly relevant for Australia, given its stable tectonic setting.

One issue with this approach is that it will not provide the precision required for more demanding applications, such as engineering or geodetic surveying. Therefore, this option would need to be combined with one of the options discussed above. Having multiple transformation options requires more care from users to record appropriate metadata, so that any transformation applied can be confidently reversed.

6 Concluding Remarks

As it becomes easier to acquire centimetre-accurate geospatial datasets in terms of global reference frames, it is important that reference frame transformations are handled correctly. For dynamic datums this means that the propagation of coordinates between epochs must also be carried out in addition to the transformation itself. While the concept of a dynamic datum may appear new, they are in fact widely used in Australia and New Zealand already, through the global reference frames WGS84 and ITRF. For WGS84 in particular, reference frame transformations are not carried out in a rigorous manner. In many cases this does not matter, due to the relatively high levels of uncertainty present in most WGS84 positions. However, transformation parameters have been recommended in this paper to enable rigorous transformations where they are required. These transformation parameters assume the close alignment of WGS84 with the ITRF. With the increasing prominence of multi-GNSS positioning, users are encouraged to use ITRF as their reference frame in preference to WGS84. ITRF is more accessible, with all global positioning technologies being aligned to it. In addition, the application of epoch is much more transparent and easier to understand for ITRF. The differences in the ITRF/WGS84 transformation parameters between Australia and New Zealand reflect the different approaches taken by each country to transformations between ITRF and their national datum. These differences are primarily due to the highly stable tectonic setting in Australia, contrasted to the relatively unstable plate tectonics in New Zealand.

Both countries are actively investigating datum modernisation through implementation of a dynamic datum (Australia) or improvements to the current semi-dynamic datum (New Zealand). Options for dynamic datum transformations have been considered. There may be advantages to separating transformation from propagation, and utilising globally-determined transformation parameters. There may also be advantages to using a gridded model to handle both transformation and propagation of coordinates. However, a solution that might be preferable from a geodetic perspective may not be the simplest approach for users of the datum. For this reason, Australia will probably continue to advocate the use of 14-parameter transformations, at least as a transitional measure until commonly used software can utilise deformation models. The final determination of a preferred method for dynamic datum transformations will need to be made considering the current state of ITRF and available user tools at the time a new datum is promulgated.

Acknowledgements

This work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centres Programme.

Dr. Chris Pearson, of the University of Otago, contributed to discussions about the most appropriate transformation parameters to use in New Zealand.

References

Altamimi, Z., Collilieux X. and Metivier L. (2011). ITRF2008: an improved solution of the international terrestrial reference frame. *Journal of Geodesy*. Vol 85, pp457-473.

Altamimi, Z., Collilieux X. and Metivier L. (2012). ITRF2008 plate motion model. *Journal of Geophysical Research*. Vol 117, B07402.

AS/NZS ISO 19111:2008 (2008). *Geographic Information – Spatial Referencing by Coordinates*. Standards Australia (Sydney) / Standards New Zealand (Wellington).

Boucher, C. and Altamimi, Z. (2011). Memo: Specifications for reference frame fixing in the analysis of a EUREF GPS campaign, http://etrs89.ensg.ign.fr/memo-V8.pdf (accessed March 2014).

Blick, G. (2003). Implementation and development of NZGD2000. *New Zealand Surveyor*. 293, pp15-19.

Dawson, J. and Woods A (2010). ITRF to GDA94 coordinate transformations. *Journal of Applied Geodesy*. Vol 4, pp189-199.

DeMets, C., Gordon, R., Argus, D. and Stein, S. (1994). Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.* Vol 21(20), 2191-2194.

Haasdyk, J., Donnelly, N., Harrison, C., Rizos, C., Roberts, C. and Stanaway, R. (2014). Options for Modernising the Geocentric Datum of Australia. In: S. Winter and C. Rizos (Eds.): *Research@Locate'14*, Canberra, Australia, 7-9 April 2014, published at http://ceur-ws.org

Haasdyk, J. and Janssen, V. (2011). The many paths to a common ground: A comparison of transformations between GDA94 and ITRF. *Proceedings of IGNSS Symposium 2011 (IGNSS2011)*, 15-17 November, Sydney, Australia.

ICG – International Committee on GNSS (2012). World Geodetic System 1984. http://www.unoosa.org/pdf/icg/2012/template/WGS_84.pdf

ICSM (2013a). Geocentric Datum of Australia 1994 (GDA94), http://www.icsm.gov.au/gda/index.html (accessed December 2013).

ICSM (2013b). Standards for the Australian Survey Control Network, (SP1), version 2.0 http://www.icsm.gov.au/geodesy/sp1.html

IERS (2013). Transformation parameters from ITRF2008 to past ITRFs, http://itrf.ensg.ign.fr/doc_ITRF/Transfo-ITRF2008_ITRFs.txt (accessed December 2013)

IGS (2010). The IGS Tracking Network: IGb08 Reference Frame sites, http://igscb.jpl.nasa.gov/network.refframe.html (accessed March 2014)

LINZ (2009). Standard for the geospatial accuracy framework – LINZS25005, 21 September 2009 http://www.linz.govt.nz/sites/default/files/document/25005-Standard%20for%20the%20geospatial%20accuracy%20framework%20-%20LINZS25005_4.pdf

Merrigan, M., Swift, E., Wong, R. and Saffel, J. (2002). A refinement to the World Geodetic System 1984 reference frame. *Proceedings of the 15th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GPS 2002)*, Portland, OR, September 2002, pp. 1519-1529.

NIMA (2000). Department of Defense World Geodetic System 1984. : Its definition and relationships with local geodetic systems. NIMA Technical Report, TR8350.2, Third Edition, Amendment 1, 3 January 2000.

Ordnance Survey (2013). A guide to coordinate systems in Great Britain, http://www.ordnancesurvey.co.uk/docs/support/guide-coordinate-systems-great-britain.pdf (accessed March 2014).

Pearson, C, and Snay, R. (2013). Introducing HTDP 3.1 to transform coordinates across time and spatial reference frames. *GPS Solutions*. Vol 17(1), 1-15.

Petit, G, and Luzum, B. (eds.) (2010). IERS Conventions (2010), IERS Technical Note 36.Verlagdes BundesamtsfürKartographie und Geodäsie, Frankfurt am Main, Germany.

Soler, T, and Marshall, J. (2003). A note on frame transformations with applications to geodetic datums. *GPS Solutions*. Vol 7(1), 23-32.

Soler, T, and Snay, R. (2004). Transforming positions and velocities between the International Terrestrial Reference Frame of 2000 and North American Datum of 1983. *Journal of Surveying Engineering*. Vol 130(2), 49-55.

Stanaway, R. and Roberts, C. (2013). A High-Precision Deformation Model to support Geodetic Datum Modernisation in Australia, *Proceedings of IAG Scientific Assembly, Potsdam, Germany, 2013* (to be published).

Wong, R., Rollins, C. and Minter, C. (2012). Recent updates to the WGS 84 reference frame. *Proceedings of the 25th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2012)*, Nashville, TN, September 2012, pp. 1164-1172.

Managing the dynamics of the New Zealand spatial cadastre

Don Grant
School of Mathematical & Geospatial Sciences,
RMIT University,
GPO Box 2476,
Melbourne 3001,
AUSTRALIA
donald.grant@rmit.edu.au

Chris Crook
National Geodetic Office
Land Information NZ,
Radio NZ House
155 the Terrace,
Wellington 6145,
NEW ZEALAND
ccrook@linz.govt.nz

Nic Donnelly
National Geodetic Office
Land Information NZ,
Radio NZ House
155 the Terrace,
Wellington 6145,
NEW ZEALAND
ndonnelly@linz.govt.nz

Abstract

In 1995, the concept of a dynamic cadastre, based on a dynamic geodetic datum, was proposed for New Zealand to recognize that all cadastral boundaries in New Zealand are in some form of motion – relative to each other and relative to the geodetic datum which is also in motion. Subsequently New Zealand implemented a semi-dynamic geodetic datum which is accompanied by a deformation model. Later, a survey conversion project resulted in the boundaries of 70% of the land parcels in New Zealand being coordinated to survey accuracy in terms of the semi-dynamic datum. These boundaries continue to be adjusted by least squares as new cadastral survey observations and geodetic control stations are integrated into the network. However the deformation model has not, in practice, been routinely applied to cadastral boundaries. In 2010 and 2011, the Canterbury region in the South Island of New Zealand was subjected to a sequence of earthquakes that caused widespread damage and resulted in some boundaries being ruptured by up to 4 metres. A set of localized deformation models was developed to model the seismic movements. Propagating these movements through to all affected cadastral boundaries has proved to be a major undertaking which is described in this paper.

1 Introduction

Cadastral boundaries, to be useful, need to be able to be realized in the real physical world where public and private rights, restrictions and responsibilities in land apply. As an important part of the land-based property rights system, it is critical that they be well documented, managed and updated in public databases – whether digital or paper-based. We use the term "physical cadastre" here to describe the physical manifestation of boundaries in the real world. The term "spatial cadastre" is used to describe the digital or paper records that describe the shape and location of those boundaries in cadastral record systems.

Boundaries in the physical cadastre may move over time. So may the representation of these boundaries in the spatial cadastre as new survey information is accepted which indicates that the boundary coordinates are incorrect and as the spatial cadastre is reviewed and readjusted as a consequence. The mechanisms that cause, manage and record these movements are naturally related to each other (a new boundary survey may result in a new accepted position for the boundary and therefore changed coordinates) but may also operate quite independently and at different times.

New Zealand sits astride the boundary between the Australian and Pacific tectonic plates. During major earthquakes, the movements of cadastral boundaries in the physical cadastre near the fault are apparent to everyone. At other times or further from the fault, the very slow, broad-scale and inexorable deformation caused by tectonic plate movement is generally invisible to landowners but does become apparent over time to surveyors and managers of the spatial cadastral system. The dynamics in the cadastre induced by earth deformation are more obvious in New

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

Zealand than in other countries such as Australia. Physical cadastral boundaries are in motion relative to other boundaries, relative to the national geodetic datum, and relative to international terrestrial reference frames.

Grant (1995) proposed the development in New Zealand of a dynamic geodetic datum which would support a system of dynamic cadastral boundaries. The most obvious driver for this model was earth deformation. It had become apparent by 1995 that the accuracy of the geodetic system could not be maintained if it was assumed that geodetic control marks were fixed in space in relation to each other or in relation to the axes of the coordinate system.

This thinking led to the implementation of a new datum for New Zealand, NZGD2000 (Grant & Pearse, 1995; Grant & Blick, 1998; Blick et al, 2003). It was implemented as a semi-dynamic datum, meaning that coordinates are defined at a reference epoch 1 January 2000, and that positions at other times are determined by applying a deformation model. The datum is aligned with ITRF96 at epoch 2000.0. Initially the deformation model was a constant horizontal velocity field defined by interpolating on a gridded representation. It was understood that over time the deformation model would be updated, both as better information about the tectonic deformation was acquired, and as events such as earthquakes introduced additional deformation components.

The deformation model handles earthquake related deformation as "patches", localized deformation models of limited spatial and temporal extent, that are added to the main secular velocity model. It was recognized (ibid) that most users of spatial data did not have the knowledge or tools to apply a deformation model, but nonetheless desired spatial data that reflected the current relative positions of data sufficiently accurately. To support these users the concept of "reverse patches" was developed, whereby the effect of earthquakes is added to the "2000.0" reference coordinates, and the patch deformation is subtracted from them to calculate coordinates before the earthquake. While this concept was developed soon after 2000, it was not until the sequence of earthquake commencing in 2010 struck the city of Christchurch and the surrounding area that there was a sufficient business driver to update the deformation model. Although there had been other major earthquakes since 2000 they only significantly affected remote, sparsely inhabited areas. In 2013 a new version of the deformation model was published including patches for eight events, four affecting the south of South Island, and four main events in the Christchurch sequence (Donnelly et al, 2014).

The cadastral system is well connected to the geodetic system and is similarly in motion resulting from earth deformation. Management of the spatial cadastre in response to deformation of boundaries in the physical cadastre, is the least well managed source of cadastral dynamics, and is the subject of this discussion.

One of the difficulties of managing a dynamic cadastre is the increasing number of customers who use the spatial cadastre in their business processes, products and services. On the one hand is the GIS community who may use it as contextual spatial data or may expressly align other datasets to it. For these customers, the dynamics of the cadastre are a nuisance – stability is often more valued by them than spatial accuracy. However on the other hand are cadastral surveyors using the cadastre in the way it was primarily intended to be used – to locate property boundaries. For these users, accuracy is more important and changes to the cadastre are expected – especially as it is surveyors who initiate those changes through the lodgement of new cadastral survey transactions.

2 New Zealand Cadastral System

2.1 Boundaries and plate tectonics

2.1.1 Principles of boundary definition

Survey marks play a very significant role in the definition of boundaries in New Zealand. Through common law set by precedent in court cases, an original and "undisturbed" boundary mark occupies a very high position on the hierarchy of evidence of boundary location. In the absence of such a boundary mark, survey measurements from a nearby mark (assuming it is also undisturbed) can be used to reinstate the original position of the boundary mark. The standards for cadastral survey in New Zealand require "witness marks" and permanent reference marks" to be placed in secure positions near the boundary – specifically so that they can serve this function of witnessing the boundary and allowing its reliable and accurate reinstatement in the event that boundary marks are disturbed or removed.

2.1.2 Undisturbed marks and tectonic motion

An original and undisturbed boundary mark has been taken to mean that a mark is in the same position as when it was driven into the ground by the surveyor that first created the new boundary. However the common law principle relying on "undisturbed" survey marks predates the present day knowledge of plate tectonics. We must now clarify what we mean by "in the same position".

In practice, the process of reinstating boundaries in New Zealand has always depends for its success on treating the

slow and imperceptible movements of plate tectonics as if they were not occurring. Or to put it another way, where a boundary mark has been moved only by tectonic processes, the Courts or surveyors have (perhaps unwittingly) considered it to be "in the same position" as it was originally placed – that is, undisturbed.

As noted in Grant (1995) this allows undisturbed boundary marks to move almost perfectly in concert with the fixed assets of landowners that are firmly resting on or attached to the earth's surface. The ownership of those assets, and the land they rest on, is thereby not affected by tectonic motion.

Of course, these marks are not in the same position in relation to the coordinate axes of the geodetic datum. Therefore coordinates of a mark, if they are to remain accurate, will necessarily change over time even though the mark is considered to be undisturbed. This led to the concept of both a dynamic datum and a dynamic cadastre (Grant 1995). It also means that accurate measurements to distant survey marks – now possible using Global Navigation Satellite Systems (GNSS) – will change over time.

2.1.3 Witness marks and earth deformation

The system of witness marks required by cadastral rules and regulations support the security of ownership because such marks are required to be within a specified distance of the boundary. This greatly reduces the risk that earth deformation will result in differential movement of witness and boundary marks. Originally, this distance restriction was intended to ensure the accuracy of measurement and boundary reinstatement given the limitations of traditional survey techniques.

More recently cadastral rules and regulations have allowed the use of Global Navigation Satellite Systems (GNSS) for cadastral survey. However a distance limitation has been retained for witnessing – not now for measurement accuracy but deliberately to minimize the risk and extent of differential movement between witness marks and the boundaries they serve to witness.

2.2 Management of the spatial cadastre

From 1996 to 2008, Land Information New Zealand developed the automated survey and title system known as Landonline. This resulted in an integrated geodetic, cadastral and land registration system with digital lodgement of structured survey and title transactions which are validated against a database populated with historical cadastral survey and land title information, and which continues to be populated with the new transactions.

2.2.1 Survey conversion project

A spatial definition of all parcels in New Zealand in the primary (ownership) layer was loaded into Landonline from the predecessor Digital Cadastral Database (DCDB). For 70% of the cadastral parcels in New Zealand, the coordinates of boundary points were then upgraded to survey accuracy. "Survey accuracy" means that the coordinates comply with the accuracy standards set in the Rules for Cadastral Survey set by the Surveyor-General).

Population of the Landonline database with accurate cadastral survey information was known as the Survey Conversion Project (Rowe, 2003). To achieve survey accuracy, the following components were required:

- An accurate geodetic datum - NZGD2000. An important attribute of NZGD2000, compared with the predecessor datum NZGD49, was that it was largely free of distortion.
- A network of geodetic control points, coordinated in terms of NZGD2000 and with all geodetic observations and coordinates brought in terms of the datum reference epoch of 2000.0 (1 January 2000).
- Extension of the geodetic network to higher density in the areas identified for survey-accurate coordinate upgrade. These "survey conversion areas" were chosen to cover the most intensive and valuable land uses – urban, peri-urban and intensive rural areas.
- Connections between geodetic control and the cadastral survey network of boundary points and marks.
- Capture of all boundary dimensions for current parcels and such other survey measurements as were necessary to provide a well-connected network (for example, connections across roads or streams).
- Least-squares adjustments of the cadastral network which allowed testing of the coordinate relative accuracies against the accuracy standards specified in the Regulations and Rules.

Those coordinates that met the accuracy standards with 95% confidence were designated as having SDC status – "Survey-accurate Digital Cadastre". The least squares adjustment of observations also provides an estimate of the accuracy of the resultant coordinates, at least in terms of the local control used by the adjustment. The accuracy is represented by an order assigned to the coordinate.

2.2.2 Automated validation and integration of new survey transactions

For the 70% of survey-accurate parcels, this accuracy status of boundary coordinates supports the semi-automated validation of new cadastral survey transactions by least squares adjustment of the new survey information in relation to the survey accurate coordinates. For this to function effectively, the survey accurate coordinates must be maintained and improved as new survey evidence is accepted into the database.

Least squares adjustment is applied to all new cadastral surveys to test self-consistency of the set of new observations and also consistency with the survey-accurate coordinates in the database. This often results in changed and improved survey-accurate coordinates for existing boundary points in the neighborhood of the survey.

2.3 Changes to boundary coordinates

As well as boundary movement resulting from tectonic processes there are a number of processes that cause boundaries to move in either the physical cadastre, or the spatial cadastre, or both. For example:

- Resurvey of water boundaries may result in them moving according to the common law doctrine of accretion and erosion. This movement occurs in the physical cadastre and the new survey definition results in a change in position in the spatial cadastre.
- New survey measurements to undisturbed old boundary marks may result in the coordinates for that boundary in the spatial cadastre being corrected to the new surveyed position.
- A new survey which identifies and resolves an error or conflict in a previous survey and, as a result, reinstates that boundary position with a new mark or new recalculated boundary dimensions. The newly defined position of the reinstated boundary may result in a change in the coordinates for that boundary.
- An upgrade of geodetic control, or a new connection between geodetic control marks and the local cadastral boundaries, may result in a significant shift in coordinates for cadastral boundaries in the area.

The first three of these cases are managed as a standard process within Landonline for every new cadastral survey dataset that is approved. Following approval, the new survey measurements and vectors are integrated into the surrounding cadastral network with a specific local network adjustment.

Periodically, a need is identified in Landonline for a Wide Area Cadastral Adjustment (WACA). This may be due to upgraded geodetic control (the 4th case above) (Donnelly and Palmer, 2006).

3 Categories of cadastral deformation

Section 2 above describes the standard processes applied in Land Information New Zealand for dynamic management of the physical and spatial cadastres. The spatial cadastre is adjusted many times a day as new cadastral survey datasets are lodged with the department by cadastral surveyors, approved by the department, and adjusted into the existing cadastral network by the department. In this sense, New Zealand already has a dynamic spatial cadastre even without accounting for earth deformation.

However managing the dynamics of the cadastre from earth deformation is not routine and spatial data management processes are still being developed. Earth deformation takes different forms and the appropriate spatial model for managing change varies according to the nature of the deformation. The relevant factors are:

- **Spatial variation**. The extent to which the deformation is spatially continuous or discontinuous.
- **Parcel distortion**. Another way of assessing the spatial variation is to consider whether parcel shapes are significantly distorted by the earth deformation.
- **Temporal variation**. The extent to which the deformation is on-going, continuous and linear; on-going, continuous and non-linear; or episodic and near instantaneous (discontinuous).
- **Boundaries follow ground movement**. It is not necessarily the case that deformation of the earth's surface will result in boundaries following that movement.

There are 2 principles of common law which affect the response of boundaries to ground movement. The first principle is that localized movement of the soil, such as occurs in landslips, does not result in boundary movement. The second principle is that moveable water boundaries only move if the accretion or erosion is slow and imperceptible. A sudden shift (avulsion) does not result in movement of the water boundary.

Different factors come into play with different types of ground movement. **Tectonic deformation** is the deformation resulting from the slow and steady movement of the 2 tectonic plates that New Zealand sits astride. This movement is taken up across a broad deformation zone that covers most of the country (Beavan & Haines, 2001). All boundaries in New Zealand are affected by tectonic deformation because the boundaries are in motion relative to each other. These movements are, to a large extent, modeled by the deformation model that accompanies NZGD2000.

Earthquake deformation is caused by the sudden stress release of earthquake, aftershocks and any post-seismic relaxation that follows the rupture. The impact on boundaries depends on whether the fault rupture reached the surface of the earth as well as distance from the fault rupture. Cadastral parcels that are very remote from the earthquake fault do not move significantly. Parcels that are remote may be subjected to block movement without distortion. Parcels nearer to the fault may be subjected to linear (affine) distortion. Parcel boundaries that are very close to the fault or lie across it may be bent (non-linear distortion) or even ruptured if the fault trace reaches the surface of the earth.

Indirect surface deformation may also occur where the surface layers of the earth are indirectly impacted by an earthquake. For example on steep slopes, the shaking may cause rockfalls and landslides. On relatively flat sites with soil or subsoil susceptible to liquefaction, the shaking may cause the surface to flow during the period of strong motion – buildings and other assets on the surface of the ground, as well as survey marks, may move with the flowing soil. Uplift or subsidence caused by the earthquake may also cause rivers to break their banks and follow a new flow-line to the sea. Table 1 summarizes each type of movement and the spatial model used for that movement. All of these types of movement of boundaries have been experienced in New Zealand in the last few years.

Table 1: Categories of boundary movement and spatial modeling – (Grant & Crook, 2012)

Movement Category	Spatial variation	Temporal variation	Parcel shape distorted	Boundaries follow ground movement	Spatial model applied
Tectonic deformation	Continuous - broad scale	Continuous - near linear	No	Yes	Datum deformation model
Earthquake - remote	Continuous - broad scale	Instantaneous + post-seismic	No	Yes	Deformation patch
Earthquake - near field	Continuous	Instantaneous + post-seismic	Near linear (affine)	Yes	Deformation patch
Earthquake - rupture zone	Discontinuous	Instantaneous + post-seismic	Non linear	Complex[1]	Interpolate across rupture, resurvey
Landslip / Rockfall	Discontinuous	Instantaneous	No	No	Not modelled
Liquefaction	Generally discontinuous	Instantaneous	Variable	Complex[2]	Not modelled
Natural boundary avulsion	Continuous but localised	Instantaneous	No	No	Not modelled

4 Canterbury Earthquakes

The Darfield earthquake of 4 September 2010 was the first in a sequence of four substantial earthquakes to impact Canterbury and Christchurch. The four major earthquakes in the sequence are outlined in Table 2 below. These four are the only earthquakes to have caused surface movements of more than 1cm (excluding highly localised movement such as that caused by liquefaction). They have therefore been the key focus of recovery and restoration activities.

Table 2: Significant earthquakes in the Canterbury 2010-2011 sequence

Date and Time	Magnitude (Richter Scale)	Approximate Depth (km)	Distance from Christchurch City Centre (km)
4 September 2010 – 4:35	7.1	11	40
22 February 2011 – 12:51	6.2	5	7
13 June 2011 – 14:20	6.0	6	10
23 December 2011 – 15:18	6.0	6	10

[1] See section 5.5 Deep-seated movement
[2] See section 5.4 Shallow surface movement

Figure 1 – Effects of fault rupture on previously straight fence and water race. (Photo - Survus Consultants)

The Darfield earthquake was the only one to result in surface rupture. The rupture was 24km long and resulted in shearing (Figure 1) across the fault. Numerous cadastral parcels are intersected by the fault rupture as shown in Figure 2. All four earthquakes also resulted in liquefaction and lateral spreading, although this was particularly serious for the 22 February 2011 earthquake, which resulted in extensive property and land damage as well as many deaths.

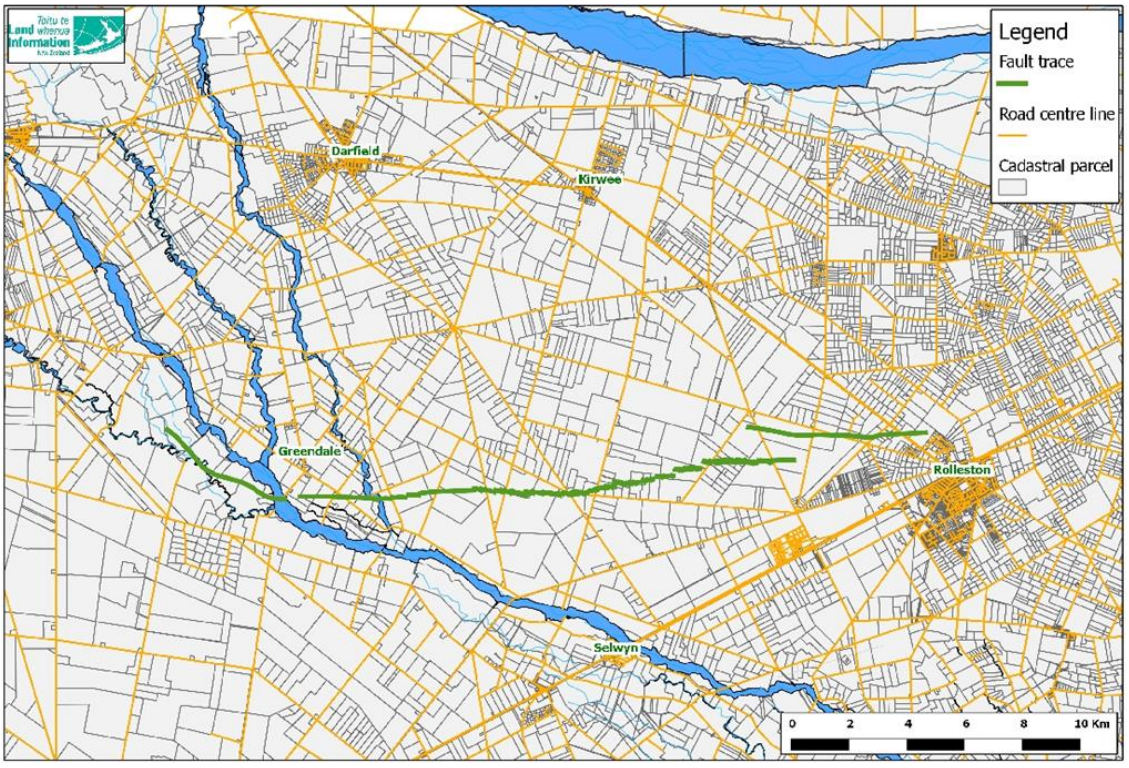


Figure 2: Darfield fault rupture overlaid with cadastral parcel fabric

5 Regulatory response to cadastral boundary movements

The legislative and regulatory responses to the sequence of earthquakes in Canterbury, New Zealand, are described in Smith et al (2011) and Grant et al (2012). Ballantyne (2004) had previously identified that there is little evidence of consistent international best practice for the re-establishment of property boundaries following earthquakes. Nickles (2009) also referred to the unsatisfactory legal position of having no legislation to deal with these situations.

5.1 Initial response under emergency legislation

Shortly after the 4 September 2010 Darfield earthquake, legislation was passed to ensure that the necessary

response and recovery efforts were not impeded by legislation that had been enacted to cover less extreme circumstances. The Canterbury Earthquake Response and Recovery Act 2010[3] provided for Orders in Council to set aside any legislative provisions that were impeding the response and recovery efforts. One of the Acts specified for such flexibility was the Cadastral Survey Act 2002 which regulates the cadastral survey system in New Zealand.

An Order in Council[4] provided for the Surveyor-General to forego the usual requirements of consultation to make Rules (having the power of government regulations) *"specifying how the spatial extent (particularly boundaries) of Canterbury earthquake land must be defined and described"*

These interim Rules and associated guidelines (Land Information New Zealand, 2010) were made to clarify how boundaries would be deemed to have moved in different circumstances and what evidence was required of cadastral surveyors reinstating them. Following the subsequent devastating Christchurch aftershock on 22 February 2011, the emergency powers were further updated by the Canterbury Earthquake Recovery Act 2011[5].

5.2 Enduring response to boundary movements

With the interim Rules in place, an amendment to the Rules for Cadastral Survey 2010 was developed applying the normal process of full consultation. Along with some other changes, these amended Rules for Cadastral Survey 2010 generalized the regulatory response to moving boundaries. This means that the applicable standards and regulations will be in place across New Zealand for comparable future scenarios – including large slow moving landslips. The Rules (Land Information New Zealand, 2012a) and associated guidelines (Land Information New Zealand, 2012b) came into force on 1 January 2013.

5.3 Principles of reinstatement of earthquake affected boundaries

Shortly after the 4 September 2010 Darfield earthquake, concern was expressed by the public on the impact on property boundaries right across the region. The Surveyor-General established the principle that boundaries in New Zealand should continue to move in concert with movements of the bedrock. This matched the status quo for boundaries throughout the country that are affected by slow tectonic deformation.

5.4 Shallow surface movement

A complicating factor, especially within urban areas, is that a great many boundaries had been moved almost at random by the effects of soil liquefaction. In this case the common law is quite clear – where the surface layers of the land move, taking with them boundary marks, fences and other assets, the boundaries do not move.

Figure 3 – The boundary point next to the fence post has moved 2.8 metres due to lateral spreading as a result of liquefaction (Photo - Eliot Sinclair and Partners)

[3] http://www.legislation.govt.nz/act/public/2010/0114/latest/whole.html
[4] http://www.legislation.govt.nz/regulation/public/2010/0467/latest/DLM3424212.html
[5] http://www.legislation.govt.nz/act/public/2011/0012/latest/DLM3653522.html

The task of the cadastral surveyor is greatly complicated in this case by the fact that in the broad areas affected by liquefaction, all existing survey marks are subjected to highly variable movements due to liquefaction and it becomes virtually impossible for surveyors to accurately identify where that original position actually was.

5.5 Deep-seated movement

The task of reinstating boundaries affected by deep-seated movement of the bedrock is made easier by the fact that there is no applicable common law for this situation. Therefore the general principle outlined by the Surveyor-General can be followed. This has the major benefit that it leaves the assets of landowners still in their possession and on their land. This principle could be stated as: "if you owned the land before the earthquake – you still own it afterwards".

The most complex example of this principle occurs in cases where boundary lines have been ruptured by the fault trace (see Figure 1). In this case, new angles will have been introduced to a formerly straight boundary line. This is illustrated in example D in Figure 4 (Land Information New Zealand 2012).

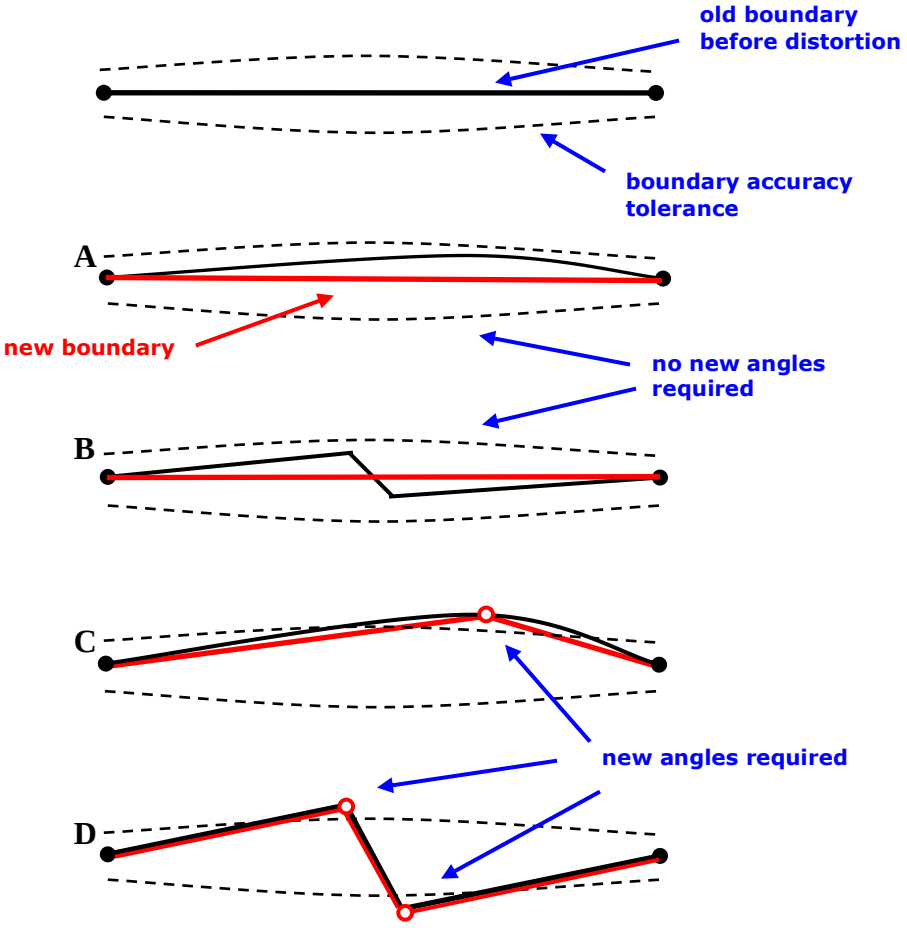


Figure 4: Boundaries affected by deep-seated distortion (Land Information New Zealand, 2012)

6 Deformation Models

A challenge in replicating the deformation from the Canterbury and Fiordland earthquakes to the cadastral fabric is that relatively few marks have been accurately surveyed since the earthquakes. The normal process for generating coordinates of parcel boundaries is by calculating them from surveys when the parcels are defined, but for most parcels there are no post-earthquake surveys. Instead the coordinates are recalculated by using a model that predicts the coordinate change due to the earthquakes, and applying this modeled coordinate change to the pre-earthquake coordinates of all affected parcels. This model is in the same as the NZGD2000 deformation model patch for the earthquake.

The deformation patch is calculated from surveyed coordinate changes together with other geophysical data, such as seismic data and DInSAR (Differential Interferometric Synthetic Aperture Radar) observations to construct a

geophysical model of the fault mechanism causing the deformation. The geophysical model can then be used to calculate the expected deformation at any other point on the surface. In order to provide a simple and efficient means of publishing and calculating the surface deformation, a grid based representation of the surface deformation is calculated from the geophysical model.

The calculation of the patch deformation grid is detailed in Winefield et al. (2010) which uses the 2009 Dusky Sound magnitude 7.8 earthquake as an example. A much more complex model was required for the magnitude 7.1 Darfield Earthquake in 2010 (Beavan et al, 2012). However the approach in each case was very similar, and ultimately is based on equations defining the deformation due to a uniform slip on a rectangular fault plane embedded in an infinite homogenous elastic half space as formulated by Okada (1985). In order to emulate the complexity of the actual deformation, the model combines the deformation on a large number of rectangular sub-faults on each of which a different slip vector is permitted.

The initial fault model is guided by seismic evidence and surface observations which indicate the likely location of the fault plane(s). This model is then refined by numerical inversion to match the observed deformation from survey measurements and DInSAR data.

Even though the models are in some cases very complex (up to 940 separate fault planes were used to model the 4 September 2010 Darfield Earthquake) they still cannot completely represent the actual deformation.

Where the fault breaks the surface and in areas of local deformation the assumptions of elastic behavior become invalid, and the accuracy of the model deteriorates. These are also the areas the ground disturbance may require the physical cadastre to be re-established in any case, so here the model, while inaccurate, provides a useful realignment of the spatial cadastre pending its update by the resurvey and re-establishment of the physical cadastre.

Elsewhere the simplicity of the model is more acceptable, particularly since we are using surface deformation observations to calculate the model, and then using the model to calculate surface deformation where it has not been observed. To some extent the physical unreality of the model relationship between fault movement and surface deformation cancels out. In effect the geophysical model provides a mechanism for smoothing and interpolating between the observations.

7 Application of deformation models to cadastral coordinates

In practice the New Zealand spatial cadastre is embodied in Landonline, the survey and title database system maintained by Land Information New Zealand. Although the primary role of this database is to manage survey and title transactions, the spatial definition of the cadastral fabric is published from this database and is widely copied and used by the New Zealand GIS community for mapping purposes and for associating other spatially defined data with the corresponding property rights. More recently this data is also directly available to customers through web services.

The cadastral fabric is in a continuous state of change. It is updated many times a day through processing of survey transactions, for example when a parcel is subdivided. It is also spatially updated as new survey measurements, are included in the database. These are used to recompute boundary and other coordinates.

All these changes are supplied to client databases by a variety of update processes, both directly and indirectly (via intermediary spatial data service providers). Client databases in turn may use a number of bespoke processes to ensure that their own spatial data remains aligned with the cadastral fabric where this is important.

7.1 Practical considerations

The application of a reverse patch to the datum means that the published reference epoch coordinates are changed by adding the effects of earthquake deformation. The NZGD2000 reverse patch for the Canterbury and Fiordland earthquakes defines coordinate changes over an extensive area. Over much of this area though the coordinate changes are small, less than 1 cm. To reduce the impact on the spatial cadastre only changes greater than 5cm were considered. To avoid a discontinuity in the dislocation field a buffer was added around the model over which the dislocation transitioned from 5cm to zero. Even with this restriction the update still affects about 15% of the spatial data in Landonline, involving over about 500,000 parcels and about 2 million corner nodes. From an operational point of view applying a change of this magnitude proved challenging and required taking the database offline for a weekend.

For clients maintaining copies of the spatial cadastre this update is in principle no different from the day to day changes that they routinely incorporate into their databases. From a practical point of view however there are two significant differences:

- the number of features being updated is much greater than a typical incremental update, and
- the coordinate change at any location is defined a simple grid based model, unlike day to day updates which are piecemeal and not defined by any model.

These two features provide both a challenge and a potential solution for clients. The large number of coordinate changes may challenge maintenance processes that do not scale up to the size of this update (for example manual maintenance procedures). However clients can use the published grid based model to directly update their copy of the cadastral data. Moreover the model can be used to update their own spatial information that is aligned with the cadastre. The coordinate update model has been published in a number of formats to support clients in applying the changes.

One difficult aspect of managing a dynamic cadastre has proved to be the technical management of large-scale coordinate and parcel topology changes within a working publicly accessible database. Landonline is not a read-only public database – it has thousands of transacting professional customers. Surveyors and solicitors are lodging new datasets all the time during working hours. LINZ staff are validating and processing those transactions. The efficient operation of the land-based property market is critical to the health of New Zealand's economy so the database can only be taken offline outside normal working hours and a failed database update cannot be tolerated. The upgrade of 2 million boundary nodes on 500,000 parcels, and in a manner that did not disrupt the efficient operation of the land property market – this proved to be challenging. However a great deal has been learned and it will be easier next time.

7.2 Limitations of the deformation model

The application of a deformation model across a large part of New Zealand, cannot account for all of the evidence that a surveyor would take into account when reinstating a boundary. For example Figure 4 above illustrates situations where new angles may be introduced into boundaries that have been bent or ruptured by the fault trace. The position of these angles can only be determined by close investigation of where the bending or rupture actually occurred along the boundary line. Figure 5 also shows that even where the fault trace has reached the surface of the earth, it is not a zero-width line that can be easily modelled – it often has a complex structure and the impact on boundaries requires judgment to be applied to each boundary line if the usual accuracy standards are to be met.

Similarly, in cases of liquefaction, the deformation model provides a reasonable estimate of the movement of the bedrock. That is of some assistance to a surveyor reinstating a boundary but they will still have to assess the complex evidence provided by survey marks, boundary marks, fences and buildings, all of which may have moved almost independently as the surface or subsurface layers of the soil turned to liquid for some tens of seconds (and which have been further moved several times by successive aftershocks).

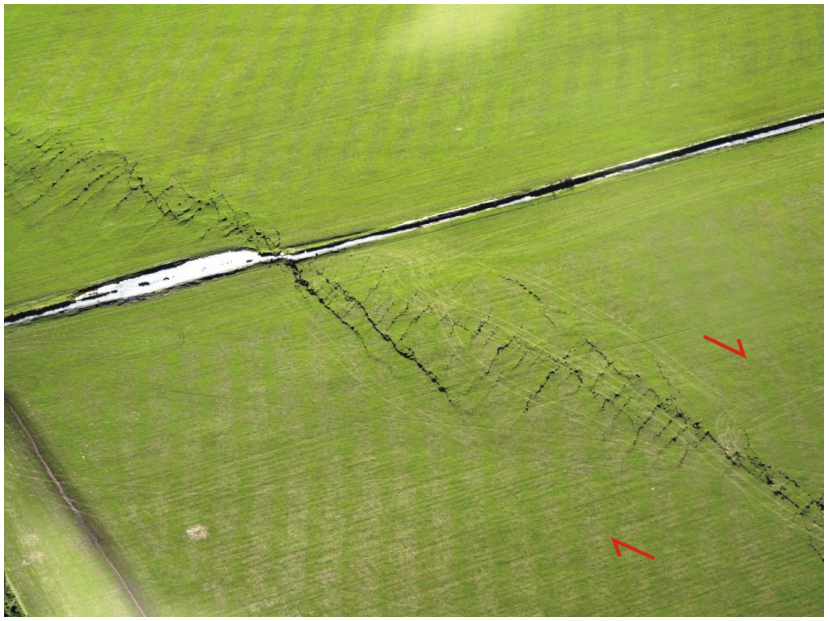


Figure 5: Greendale Fault surface rupture. Arrows indicate direction and width of displacement. Here, ~3.5 m of displacement is distributed across a zone up to 40 m wide. Photo by Richard Jongens (Quigley et al, 2010).

The deformation model therefore provides an approximation of how boundaries have moved – a much better approximation than a null model (the default model for a static cadastre) which assumes no such movement. But it cannot fully substitute for reinstatement by a licensed cadastral surveyor who is able to collect and assess local evidence of movement and apply the correct legal principles.

The ability of the model to represent the real-world changes due to the earthquakes varies across the affected area. The model has a level of uncertainty associated with it, which propagates into the coordinates calculated using it. The agreement between observed and modeled positions increases with increasing distance from the fault, and is lower where there is localised deformation caused by phenomena such as liquefaction. Based on analysis described in Donnelly et al (2014), the uncertainty of the model was used to assess whether the accuracy classification for a particular coordinate required updating, as well as the coordinates themselves.

8 Conclusions

It is apparent that cadastral boundaries physically move as a result of earth deformation and that the spatial cadastre needs to be able to respond to and model those movements. However this problem covers a complex spectrum of specialist knowledge: geophysics; geodesy; management of the spatial cadastre; and land law. The dynamics of the earth are reasonably well known, measured through geodetic techniques and modelled in solid-earth geophysics. New Zealand's geodetic datum has a deformation model associated with it to recognize the motion of "fixed" survey marks attached to the surface of the earth.

At the other end of the knowledge spectrum, land law is based on centuries of common law and precedents formed in a small number of historic court cases. Those precedent setting cases have not, to date, recognized the existence of geodynamics on the surface of the earth. The tectonic motions are slow but continuous, occur across the whole country and mostly cannot be detected by the general public. Motions resulting from earthquakes are localized, frighteningly fast, but of short duration.

The cadastre must bridge the interface between the measured dynamics of the earth's surface and the relatively inflexible, slow moving and slow changing application of land law which nevertheless serves the vital function of protecting property rights in land.

New Zealand has started down the path of bridging this gap. Ballantyne (2004) recommended that principles be established and, if necessary, legislation to address the uncertain impact on property boundaries. The necessity of responding to the Canterbury earthquake sequence has taken us some way forward with the amended Rules for Cadastral Survey 2010 (Land Information New Zealand, 2012) but there is much still to learn. Grant (1995) in proposing a dynamic datum for a dynamic cadastre, anticipated that these issues may be resolved by the year 2010. That proved to be too optimistic but steps towards this goal have been made and the problem cannot be ignored for long. More research in a number of areas of geodesy, sensing of deformation, spatial management of the cadastre and land law will be required.

9 References

Ballantyne, B. (2004). *Managing the New Zealand cadastre after deformation events: applying grit to a slippery slope*. Research report prepared for LINZ.

Beavan, J., Haines, J. (2001). Contemporary horizontal velocity and strain rate fields of the Pacific-Australian plate boundary zone through New Zealand. *Journal of Geophysical Research: Solid Earth (1978–2012), 106(B1)*, 741-770.

Beavan, J., Motagh, M., Fielding, E., Donnelly, N., Collett, D. (2012). Fault slip models of the 2010-2011 Canterbury, New Zealand, earthquakes from geodetic data, and observations of post-seismic ground deformation, *N. Z. J. Geol. Geophys., Vol 55(3).*

Blick, G., Haanen, H. (2005). The Earthquake's Impact on Property Boundaries. *Proceedings, The 1855 Wairarapa Earthquake Symposium, Wellington, New Zealand.*

Blick, G., Crook, C., Grant, D., Beavan, J. (2003). Implementation of a Semi-Dynamic Datum for New Zealand, *International Association of Geodesy Symposia, 30 June – 11 July, (128), Sapporo, Japan*.

Donnelly, N., Crook, C., Amos, M., Grant, D., Ritchie, J., Roberts, C. (2014). Canterbury Earthquakes Deformation Model Accuracy. *Proceedings of Surveying and Spatial Sciences Conference (LOCATE14), 7-9 April, Canberra, Australia* (in press).

Donnelly, N., Palmer, J., (2006). Issues with maintaining spatial accuracy in a nationwide digital cadastral network. *New Zealand Surveyor 296*, 34-39.

Grant, D.B. (1995). A dynamic datum for a dynamic cadastre. *Australian Surveyor, 40(4)*, 22-28.

Grant, D.B., Pearse M. (1995). Proposal for a Dynamic National Geodetic Datum for New Zealand. *Proceedings International Union of Geodesy & Geophysics General Assembly, Boulder, Colorado, USA*.

Grant, D.B., Blick G. (1998). A new geocentric datum for New Zealand. *New Zealand Surveyor No. 288*, 40-42.

Grant, D.B., Crook, C. (2012). Spatial maintenance of the New Zealand cadastre in response to earthquakes. *Proceedings of FIG Working Week, Rome, Italy*.

Grant, D.B., Smith, M., Thompson, M. (2012). What happens to the cadastre when the earth moves: legislative and regulatory responses to the earthquakes in Canterbury, New Zealand. *Proceedings, International FIG Symposium & Commission 7 Annual Meeting, Innsbruck, Austria*.

Land Information New Zealand, Surveyor-General. (2010). *LINZG65702: Guideline for Rules for Cadastral Survey (Canterbury Earthquake) 2010.* LINZ, Wellington.

Land Information New Zealand, Surveyor-General. (2012a). *LINZS65003: Rules for Cadastral Survey 2010 (amended 1 November 2012).* LINZ, Wellington.

Land Information New Zealand, Surveyor-General. (2012b). *LINZG65704: Interim guide to the amended Rules for Cadastral Survey 2010.* LINZ, Wellington.

Nickles, W.L. (2009). Ground movement and its effect on cadastral boundaries. Chapter 8 of *Land title surveys in New Zealand.* Ed. D.F. McKay, New Zealand Institute of Surveyors.

Okada Y. (1985). Surface deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am., 75:4*, 1135-1154.

Quigley, M., Van Dissen, R., Litchfield, N., Villamor, P., Duffy, B., Barrell, D., Noble, D. (2012). Surface rupture during the 2010 Mw 7.1 Darfield (Canterbury) earthquake: Implications for fault rupture dynamics and seismic-hazard analysis. *Geology, 40(1),* 55-58.

Rowe, G. (2003). The survey conversion project – making a survey-accurate digital cadastre a reality for New Zealand. *New Zealand Surveyor 293*, 31-38.

Smith, M., Thompson, M., Grant, D.B. (2011). Re-establishment of cadastral boundaries following the 2010-2011 Canterbury earthquakes. *Proceedings of the Spatial Sciences and Surveying Conference, Wellington, New Zealand.*

Winefield, R., Crook, C., Beavan, J. (2010). The Application of a Localised Deformation Model after an Earthquake, in *Proceedings of XXIV FIG Congress, April 11-16, Sydney, Australia*. Available at: http://www.fig.net/srl/

Options for Modernising the Geocentric Datum of Australia

Joel Haasdyk
Survey Infrastructure and Geodesy
NSW Land and Property Information
Bathurst, NSW
CRC for Spatial Information
Joel.Haasdyk@lpi.nsw.gov.au

Nic Donnelly
National Geodetic Office
Land Information New Zealand
Wellington, New Zealand
CRC for Spatial Information
ndonnelly@linz.govt.nz

Dr. Craig Harrison
Geoscience Australia
Canberra, ACT
CRC for Spatial Information
Craig.Harrison@ga.gov.au

Dr. Chris Rizos
Civil and Environmental Engineering
University of New South Wales
Sydney, NSW
CRC for Spatial Information
c.rizos@unsw.edu.au

Dr. Craig Roberts
Civil and Environmental Engineering
University of New South Wales
Sydney NSW
CRC for Spatial Information
c.roberts@unsw.edu.au

Richard Stanaway
Civil and Environmental Engineering
University of New South Wales
Sydney, NSW
CRC for Spatial Information
richard.stanaway@student.unsw.edu.au

Abstract

Instantaneous, reliable and fit-for-purpose positioning and time services across Australia are the aims of the National Positioning Infrastructure policy. Indeed, it is already possible to achieve centimetre-level positioning almost anywhere (outdoors) and anytime, following advances in positioning technologies over the last few decades. Not surprisingly, this capability has highlighted limitations and distortions in the current Geocentric Datum of Australia 1994 (GDA94) which was not designed to support positioning at this level of accuracy. A 'next-generation' datum or 'Australian Terrestrial Reference Frame' is currently being prepared which will remove the known distortions of the current datum realisation, create a homogenous 3-D datum across Australia based on permanent GNSS datum stations, be adaptable and flexible by incorporating new measurements and technologies as they become available, and assign realistic uncertainties to coordinates. Importantly, this datum will include a comprehensive deformation model to account for real-world dynamics which GDA94 currently ignores, such as the metre-level differences between our maps and mobile positioning devices due to tectonics, earthquakes, ground subsidence and localised deformation. This paper presents the technological progress towards this new datum, reports on recent user-needs assessments across Australia, discusses several options for the final realisation of the modernised datum, and discusses future research required to achieve a fully dynamic datum. While focused on the Australian datum, much of what is presented is also applicable to the modernisation of national and international datums across the globe.

1 Introduction

The purpose of this paper is to investigate the advantages and disadvantages of various options for the development of a modernised Australian datum or 'Australian Terrestrial Reference Frame' (ATRF). Such a modernised datum is required as a framework for the capture and comparison of modern survey, geodetic and other spatial data. These spatial data are collected at discrete epochs but represent an earth which is changing at rates which exceed current measurement precision. With significant improvements in positioning technology now enabling centimetre-level positioning capability via techniques such as Network Real Time Kinematic (NRTK) and Precise Point Positioning (PPP) (Janssen *et al*., 2011; Rizos *et al*., 2012), and decimetre-level accuracy or better soon available to the mass-market (Gakstatter, 2013a), the need to 'modernise' the geodetic datum is once again inevitable.

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

The recent history of Australian datums demonstrates that, even on a relatively stable tectonic plate, improvements in positioning technology and changes in user accuracy requirements have driven datum updates on an approximately decadal time scale. The recent widespread establishment of Continuously Operating Reference Stations (CORS) which permanently gather range data to the satellites of Global Navigation Satellite Systems (GNSS) has highlighted distortions in the existing realisation of the Geocentric Datum of Australia 1994 (GDA94) and metre-level offsets from international datums such as the International Terrestrial Reference Frame (ITRF) and navigational systems such as the World Geodetic System (WGS84) (Haasdyk & Watson, 2013; Dawson & Woods, 2010, Altamimi *et al*., 2011, Donnelly *et al*., 2013).

Improvements in computing power and geodetic adjustment software mean that it is now possible to compute, on a daily basis, a rigorous national adjustment of all available measurements, incorporating new measurements and new positioning technologies immediately when they become available. As a result, survey control station coordinates (including those of CORS) which are used to realise the datum could conceivably change with a reasonably high frequency to reflect real-world dynamics including tectonic motion and episodic deformation (e.g. localised ground subsidence), as well as improvements in coordinate and uncertainty estimation.

The assumption that 'static' unchanging coordinates are sufficient for most spatial data users is being challenged, but users are understandably reticent to describe a locally 'stable' physical feature using 'dynamic' coordinates which change over time. However, it has been demonstrated that localised as well as large-scale deformation exists across Australia (e.g. Ng *et al*., 2010; Featherstone *et al*., 2012; Tregoning *et al*., 2013). Many users also create new spatial datasets in WGS84 (if using GPS broadcast satellite orbits), or in the latest ITRF, currently ITRF2008, (if using precise satellite orbits), in which coordinates of 'stable' physical features *do* change over time. Unfortunately, many users do not understand the implications of these datum differences and fail to apply the necessary transformations to their data, resulting in significant data inaccuracies (Gakstatter, 2013b; Donnelly *et al*., 2013).

The goal of datum modernisation is to supply all users with the most complete yet most straightforward datum products which can define a locally consistent set of coordinates, such that their positioning device agrees with the physical world and associated spatial data to an acceptable level of accuracy. In this context, this paper explores options for datum modernisation which are nominally for application in Australia, but which are relevant for datum modernisation across the globe. Particular attention is given to minimising the risks and costs associated with frequent coordinate conversions, while exploring the benefits of maintaining a 'dynamic datum' in close alignment with the reference frame(s) used to define the GNSS satellite orbits, especially for the growing group of mass-market users who tend to be unaware of datum issues.

1.1 Defining a Dynamic Datum

Geodetic datums can be classified as either 'static', in which the coordinates of locally stable physical features do not change over time, as 'dynamic' (also 'fully-dynamic'), in which coordinates of physical features change continuously (i.e. in real-time) usually to reflect movement in a global context, or as 'semi-dynamic', in which coordinates are officially defined at a particular date or 'reference epoch' but dynamics are catered for via application of a deformation model. (e.g. Blick *et al*., 2009; Stanaway *et al*., 2011). The word 'kinematic' is also frequently used interchangeably with 'dynamic' in many texts.

In the case of both the 'dynamic' and 'semi-dynamic' datum, a deformation model is explicitly defined as part of the datum, providing for official and traceable propagation of coordinates (and measurements) between epochs. Deformation models generally consist of the velocity of rigid tectonic plates, and/or of discrete points representing site-specific motion in the region, and in some cases also define a series of episodic deformation events such as earthquakes or subsidence. The motion of any local station or feature can therefore be interpolated from this model.

In contrast, a 'static' datum is completely insensitive to any secular or deformation information, and defines only the coordinates of physical features as they were at a certain point in time. For example, GDA94 is a 'static' datum, with coordinates defined relative to the ITRF92 at epoch 1994.0. The fact that the Australian continent undergoes tectonic motion is simply disregarded in the definition of GDA94, and therefore GDA94 coordinates do not change over time.

To be explicit in this definition, it is noted that dynamic datums are actually constant in their definition: ITRF for example, is known as a dynamic datum, but the origin of the datum is defined as the centre of mass of the Earth, with the X, Y, and Z axis orientation strictly defined with respect to conventional international agreement (Altamimi *et al*., 2011; Petit & Luzum, 2010). These axes can be thought of as rotating with the Earth and do not change in response to surface deformations. On the surface of the Earth, however, tectonic plates move at rates which can be observed by modern positioning technologies; it is, in fact, the tectonic plates that are moving *within* the constant ITRF and therefore the coordinates expressed in ITRF change with time. The whole purpose of a dynamic datum is to define a stable reference frame against which one can monitor and describe long-term global phenomena such as sea-level change or global mass redistribution.

1.2 The Current Geocentric Datum of Australia

The current Geocentric Datum of Australia, GDA94, was defined by the Australia and New Zealand Intergovernmental Committee on Surveying and Mapping (ICSM) as Australia's first geocentric datum (ICSM, 2006; ICSM, 2013a). The purpose of defining GDA94 was to align Australia's datum and mapping with new satellite geodetic technologies and to correct distortions of up to 6 metres which had been detected by these technological advances (Featherstone, 2013). For all practical purposes, this static GDA94 datum was compatible with GPS and ITRF at the time of implementation.

To form the link to the global reference frames in which the satellite orbits are determined, the coordinates of eight permanent stations of the Australian Fiducial Network (AFN) were defined in three dimensions using GPS, in the most rigorous datum then available – the International Terrestrial Reference Frame 1992 (ITRF92) – at epoch 1994.0 (ICSM, 2006). Uncertainties at the AFN were determined to be approximately 30 mm (horizontal) and 50 mm (vertical) (95% confidence interval (CI) – Dawson & Woods, 2010). Additional ground control (horizontal only) were then adjusted in a hierarchy, in regional sections due to computational limitations, holding higher order control stations fixed, making it necessary to estimate rather than compute absolute positional uncertainty within the adjustment (ICSM, 2013b). The resulting coordinates were given uncertainty estimates (e.g. approximately 250 mm (95% CI) within NSW – Haasdyk & Watson, 2013), and further hierarchical densification is still undertaken today.

1.3 Drivers for Datum Update

Many of the same drivers for datum update which resulted in the development of GDA94 are again providing the need for datum modernisation, and are listed below (Dawson & Woods, 2010; Stanaway *et al*., 2011; Featherstone *et al*., 2012; Haasdyk & Watson, 2013; Stanaway & Roberts, 2013; Tregoning *et al*., 2013):

- Technological improvement and significantly more precise geodetic measurements gathered since 1994 can be used to compute improved coordinates and uncertainties.
- Systematic distortions in GDA94 of up to 300 mm (horizontal) have been detected by modern measurements such as from CORS, and 'site-transformations' are currently required to agree with local ground control coordinates.
- The Australian tectonic plate moves (and rotates) at ~7 cm/yr, but GDA94 is defined by coordinates locked to epoch 1994.0. GDA94 will therefore be offset with respect to ITRF and WGS84 coordinates by ~1.5 m in 2015. The rotation is significant for surveying and geodesy applications, affecting bearings of baselines by 7 mm per 30 km over a 20 year period.
- The 1.5 m offset is large enough to affect expected positioning accuracy of mass-market devices such as smartphones and tablets. These will likely determine coordinates in the latest ITRF − without direct reference to GDA94 − by directly accessing International GNSS Service (IGS) products in real time.
- Ground deformation is readily apparent in subsidence due to water, coal or gas extraction. Deformation due to seismic activity is more pronounced in other nearby countries, such as New Zealand, but is still observable within the Australian tectonic plate.
- A significant 9 cm vertical bias exists between ITRF92, upon which GDA94 is based, and ITRF2008.
- Improvements in computing hardware and software capabilities now make it possible to perform rigorous geodetic adjustments of a virtually unlimited number of station parameters and measurements.

2 Progress towards a new National Adjustment

In recent years, the Geodesy Technical Sub Committee (GTSC) – now the Permanent Committee on Geodesy (PCG) – of the ICSM has been preparing for datum modernisation. The following sections describe recent progress.

2.1 ATRF and APREF as a densification of the ITRF

While the ITRF provides the internationally accepted global reference frame for high-precision applications, the density of the network is low, with only 580 sites world-wide and 117 in the southern hemisphere (Altamimi *et al*., 2011). The Asia Pacific Reference Frame (APREF) has already been established to create and maintain a more densely realised geodetic framework of CORS in the Asia-Pacific region based on the continuous observation and analysis of GNSS data (Geoscience Australia, 2013). Recent APREF solutions report more than 300 stations across Australia and New Zealand, compared to fewer than 30 stations in the same region, in the ITRF solution.

An Australian Terrestrial Reference Frame (ATRF) would provide a further densification of ITRF and serve as the modernised dynamic datum in Australia by combining data from the Australian CORS of the APREF, a national homogenous adjustment of all available geodetic measurements, and a national deformation model (refer to section 2.4). ATRF would remain aligned to ITRF but provide the density of stations and/or deformation models required for the accurate maintenance of spatial data in Australia.

2.2 Collation of Jurisdictional Datasets

In Australia, each state or territorial jurisdiction is the custodian for its own geodetic measurements and legal source for ground control coordinates. Consequently, since GDA94 was defined, each jurisdiction has gathered and adjusted its survey control measurements using different methods, with some jurisdictions holding fixed the geodetic control coordinates originally adopted, and others re-coordinating all control stations as new measurements become available. This has caused discontinuities in coordinates at jurisdictional boundaries.

These measurements are currently being collated for an upcoming national adjustment (e.g. Haasdyk & Watson, 2013). Figure 1 shows some examples of jurisdictional adjustments, with Victoria and New South Wales, for example, providing tens of thousands of stations and up to 100,000 measurements each. The obvious differences between the state diagrams also reflect the different tools and processes being used to store, adjust and analyse the data at this time.

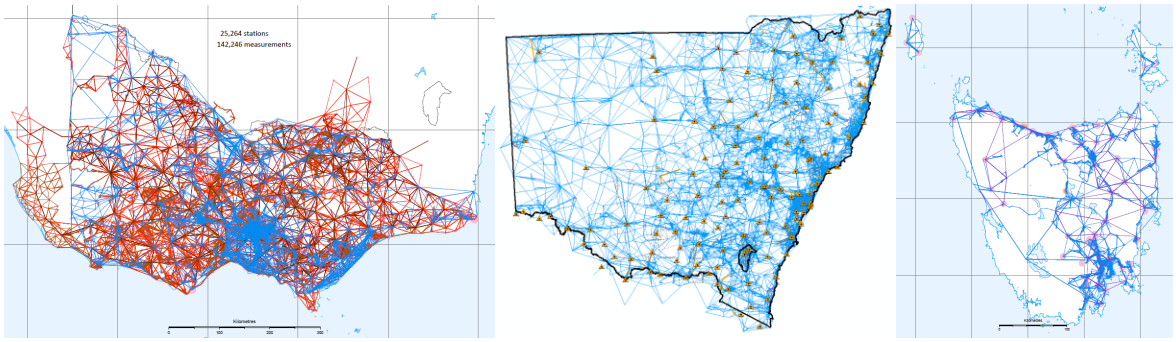


Figure 1: (Left to Right) Victoria, NSW and Tasmanian datasets. Scales vary. Colour schemes vary. (e.g. Victoria Data: Blue = GNSS, Red = Terrestrial)

2.3 New Geodetic Adjustment Software and Hardware

Historically, geodetic adjustments were limited in size due to computing memory and hardware available. More recently, Leahy and Collier (1998) described a 'dynamic phased-adjustment' which can perform a rigorous adjustment on a network of any size. This allows the computation of absolute Positional Uncertainty (PU) of all observed stations across the network, as well as relative uncertainty between any nominated stations, as described in the new Standards and Practices document (ICSM, 2013c).

DynaNet, a program that can perform such a dynamic network adjustment, has been developed at the former Department of Geomatics at the University of Melbourne (Fraser *et al*., 2014). This program has been modified to allow the adjustment to be computed in any datum and reference epoch, for example by rotating and translating all available GNSS baselines and position measurements to account for tectonic motion via published conformal transformations (e.g. Dawson & Woods, 2010). Further development is underway to incorporate the more complex deformation models described in section 2.4.

While DynaNet can be run on a desktop computer, the proposed national adjustment will be carried out on a new Australian supercomputer known as 'Raijin', currently the 27th fastest computer in the world (NCI, 2013). Primarily intended for climate modelling, Raijin is available for other computationally-intensive activities of national interest.

Testing of hypothetical large national adjustments has already been undertaken, with a 400,000 station network, comprising over 1.2 million GNSS measurements, adjusted within approximately 50 hours. Further improvements to the efficiency of the software are planned, which have the potential to substantially reduce the duration and memory requirements of the adjustment.

2.4 Deformation Models

Deformation of the land surface relative to the defined datum includes a complex combination of secular tectonic motion, and more localised, usually episodic deformation events such as those due to ground subsidence or earthquake events. Stanaway *et al*. (2012) provides a useful summary of deformation types and expected magnitudes. A deformation model is therefore required to enable coordinates (and measurements) to be propagated between epochs. For example, such a model would allow coordinates which are determined from contemporary measurements, as well as existing spatial datasets, to be propagated to a common epoch for direct comparison.

2.4.1 Secular Tectonic Motion

The internal deformation of the Australian continent (with the exception of isolated areas of intraplate earthquakes and subsidence) is less than 1.0 mm/yr over 4000 km or 0.25 ppb (Tregoning, 2003; Tregoning *et al*., 2013; Stanaway & Roberts, 2012). As a result, a rigid plate motion model is sufficient to account for most of the

dynamics experienced by the Australian tectonic plate. Stanaway and Roberts (2013) indicate that a site velocity model based on the ITRF2008 plate model can account for tectonic motion in the stable parts of the Australian plate with an uncertainty of only 6 mm at 95% CI between 2013.0 and 1994.0. This model could be further improved by a higher density of observed site velocities at the APREF CORS.

The application of a deformation model based only on rigid tectonic plate motion would be in-line with current practice in North America and Europe, where the North American Datum 1983 (NAD83) and the European Terrestrial Reference System 1989 (ETRS89) are respectively fixed to the North American and Eurasian plates. In this way, NAD83 and ETRS89 are treated as plate-fixed static datums in the same way as GDA94, with the velocities of stable points minimized (Pearson and Snay, 2013; Ordnance Survey, 2013).

For example, ETRF89 (the European Terrestrial Reference Frame realisation of ETRS89 used in Great Britain) was equivalent to ITRF89 at epoch 1989.0. To propagate from current ETRF89 coordinates to the ITRF89 on which it is based requires only the application of the rigid tectonic plate model. Conveniently, the motion of a rigid plate can be described simply by three parameters, namely rotation about the X, Y, and Z axis of the defined datum (Stanaway & Roberts, 2009). These parameters can be easily applied in the form of a 7-parameter conformal transformation, which is commonly accepted in many software packages, with translation and scale parameters set to zero. Propagation from GDA94 to ITRF92 could be accomplished in the same manner. Of course for regions in deforming zones such as the western United States, the proximity to the plate boundary means that a rigid plate motion model cannot accurately account for the more complex motions observed.

Any remaining transformation between datums, e.g. between current ITRF2008 coordinates and ITRF89 on which ETRF89 is based, requires a second step in which a 14-parameter transformation is applied. It is possible to accomplish both transformation between datums and propagation between epochs in one step, by combining the parameters from the two steps above, as in ITRF2008 to ETRF89, or ITRF2008 to GDA94 transformations (Boucher & Altamimi, 2011; Dawson & Woods, 2010).

2.4.2 Gridded Deformation Model

While most of Australia experiences little relative deformation, in some areas such the Perth Basin, Newcastle and the Latrobe Valley, significant deformation has been observed relative to the surrounding stable plate. For example, Featherstone *et al*. (2012) demonstrate subsidence of greater than 5 mm/yr in Perth due to ground-water extraction.

Stanaway and Roberts (2013) have proposed a gridded deformation model for the Australian continent which can be used to propagate ITRF coordinates at any epoch, to GDA94. This Australian continental deformation model consists of a 1 degree grid of site velocities to account for tectonic plate motion and other site-specific velocities, plus a 1 degree grid of coordinate 'patch' corrections at defined epochs to account for distortion between the reference frames as well as episodic deformation events. The gridded deformation model has advantages over the current 14-parameter model (Dawson & Woods, 2010) in that localised deformation can be defined within the model and denser nested grids can be used in areas where greater precision is required, or where more significant deformation is occurring.

The proposed site velocity model is derived from the ITRF2008 plate model (Altamimi *et al*., 2011) with corrections applied based on observed versus modelled site velocities at a number of APREF CORS stations with time series longer than eight years. The national patch model is derived from differences between the gazetted coordinates of these APREF stations and the site velocity model regressed to epoch 1994. Evaluation of the model indicates agreement within several millimetres compared to the current 14-parameter transformation at these sites.

2.4.3 Deformation Model of the New Zealand Geodetic Datum 2000 (NZGD2000)

An example of a deformation model in practice is the NZGD2000 Deformation Model, published by Land Information New Zealand (Crook and Donnelly, 2013; Winefield *et al*., 2010). This model, recently updated, includes a simple linear velocity model at a regular grid of roughly 10 km spacing, and as of February 2014 also includes sub-models or 'patches' for 10 significant earthquakes to have affected New Zealand since 1 January 2000, including the 2010 and 2011 Canterbury earthquakes.

These 'patches' represent any distortion which is measured (e.g. from a network of CORS or geodetic control stations) or inferred (e.g. from large scale geophysical models) and also provide associated uncertainty values. Patches provide variable resolution information of complex motions (e.g. linear, ramp, step or exponential decay) which describe any known deformation events to the maximum ability of the existing measurements and models.

2.4.4 Deformation determined by DInSAR and Other Non-traditional Geodetic Techniques

Ground surface deformation, for example due to natural events such as earthquakes or anthropogenic activities such as water extraction, can be highly variable over a small area, and is often non-linear. These complex deformations cannot be adequately monitored using technologies such as GNSS because of the sparse distribution of ground control stations and the high cost of field operations to occupy and observe these control stations before, during and after the deformation event(s).

Research is currently being undertaken via the Cooperative Research Centre for Spatial Information (CRCSI) to determine how best to take advantage of large-scale quick-repeat remote sensing techniques such as Lidar and Differential Interferometric Synthetic Aperture Radar (DInSAR) to complement traditional geodetic point-based techniques, and how best to model and quickly disseminate deformation information to allow the spatial community to easily create and maintain products which are precise and accurate at the centimetre-level.

For example, DInSAR can measure land movements to a precision of a few millimetres (Simons and Rosen, 2007). This technique uses airborne or spaceborne radar to create deformation maps or grids, from measurements made with a density down to 10 metres. Thus areas of anomalous deformation can be readily identified for additional monitoring with GNSS, Lidar or further DInSAR measurements.

Additional work will need to be undertaken to resolve potential discontinuities in deformation that could result from making measurements at unstable monuments. For example, geodetic monuments, historically placed on reservoirs or silos to assist with long-distance sighting, may be subject to large deformations which are station-specific and do not agree with land deformation in the near vicinity (Haasdyk & Roberts, 2013).

2.5 User Discussion Forums

A number of education and discussion forums have been convened by state, territory or national geodetic authorities across the country in the last two years. These forums aim to introduce the plans for a next generation datum to a diverse group of spatial data users, and to understand the perceived and real costs, benefits and opportunities associated with datum modernisation from the users who will be most affected by these changes.

The first forum, internally run at NSW LPI in late 2012, highlighted the diverse groups (councils, engineering, mining, construction, research, etc.) currently utilising legacy datasets in GDA94. The first open forum was held in March 2013 with predominantly surveyors (around 250) interested in centimetre-level positioning. The second, in April at the SSSI conference, was targeted at geospatial professionals and the audience contained a diverse range of users. Since then, additional forums have been held in NSW, VIC, ACT, NT, TAS and webinars describing datum modernisation issues have been made available (Donnelly & Haasdyk, 2013)

Most importantly, as a result of these forums, it is recognised that the decision of spatial data holders to modernise their existing datasets should be dependent on readiness of their systems. The modernised datum should provide opportunities for geospatial professionals to create innovative high-accuracy products and applications, but not to force all users into upgrades for which they are not yet prepared. Users are understandably concerned about the cost and risk of transforming or propagating existing datasets, the dearth of existing tools to handle deformation, and the methods by which data from various epochs will be combined for comparisons. The prevalence of value-adding to existing datasets for provision to third parties has been discussed, along with implications of frequent datum updates on data management.

A common theme at all forums is the need for metadata management, and in particular information about when data was captured, and the methodology and estimated precision of the data capture. Many users were surprised to realise that WGS84 was, in practice, a collection of lower-precision datums and that time-stamp metadata was an integral component of expressing coordinates in WGS84 (Donnelly *et al*., 2013).

The provision of an 'authoritative epoch' to which data can be referred was desirable, but the feedback from several forums indicated that users would rather avoid a short-term interim solution of another static, albeit higher accuracy, realisation of the datum. Rather, the general response appears to be that if and when the tools to handle deformation are available, most users would prefer to make a single step directly to a dynamic datum and adopt a new (if necessary) reference epoch.

3 Options for Realisation of the New Datum

A number of optional 'features' are under consideration which would affect the final realisation of a 'next-generation' geodetic datum. These include corrections of known distortions, adoption of the modern ITRF/ATRF, adoption of a deformation model (simple or complex) and therefore a dynamic datum, and choice of reference epoch. It is not the purpose of this paper to select or promote the features of the final realisation, as that authority rests with the ICSM. Instead, this paper presents a discussion of the potential advantages and disadvantages associated with each option, and summarises a few of the likely realisations in Table 1. The following discussion assumes that some form of modernisation *will* be undertaken, the alternative being to accept the status quo and ignore the current issues highlighted in section 1.3.

3.1 Correction between GDA94 and an Homogenous and Rigorous National Adjustment

The most basic form of datum modernisation involves the adoption of modern measurements and adjustment methods to correct any known distortions in the existing GDA94 datum which have been highlighted by modern high precision GNSS baseline measurements and continuous CORS observations (Haasdyk & Watson, 2013). The

advantage of a new simultaneous national adjustment of all available measurements is the provision of improved, rigorous and homogenous coordinates and positional uncertainty across Australia.

These updated coordinates can be compared to existing GDA94 coordinates to characterise and quantify these distortions at a reasonably high density. Figure 2 demonstrates that within NSW for example, the distortions between current GDA94 coordinates and a new national adjustment expressed in GDA94 are quite systematic, with expected coordinate changes of up to 300 mm for this GNSS-only dataset. Of course control stations which are remote or currently have only terrestrial measurements such as directions and distances are likely to have even greater coordinate improvements when re-observed in future with GNSS measurements, with an associated improvement in their positional uncertainty.

Any control stations or spatial data not included in the national adjustment could be easily transformed to the new datum by a dense grid transformation, similar to the NTv2 grids used in AGD-GDA94 transformations (Collier *et al*., 1997; Collier & Steed, 2001). Such a grid transformation would be computed based on the differences between existing GDA94 coordinates and the new adjustment coordinates at common points. There are costs and risks associated with transforming existing datasets, and the benefits of improved accuracy must be weighed up by the data manager. Of course, any spatial data of lower accuracy than the expected coordinate corrections (e.g. worse that 0.5 metre accuracy) need not be transformed.

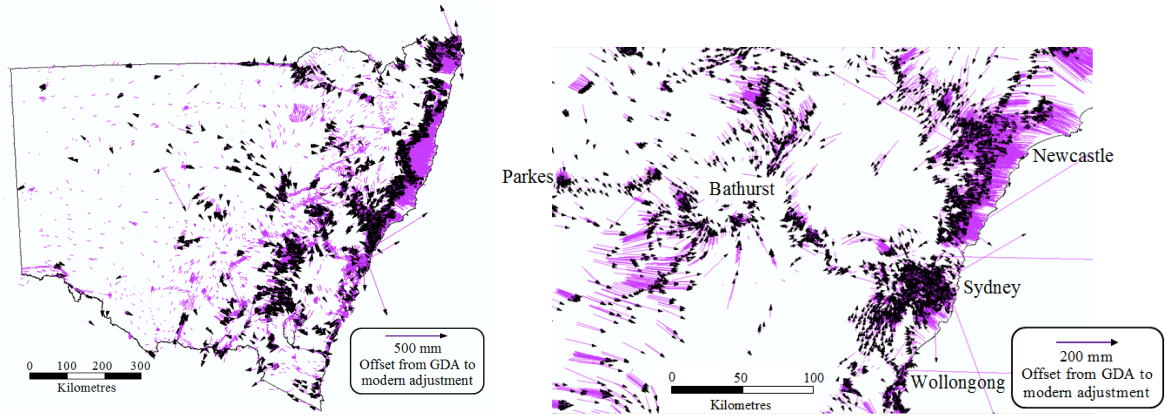


Figure 2: Expected horizontal changes to GDA94 coordinates in NSW Australia after a national adjustment.

3.2 Updating to the latest ITRF

A second option for datum modernisation is to adopt the improvements in accuracy and stability of the latest ITRF, currently ITRF2008. Note that ITRF2013 is currently under preparation and will likely be available by the completion of the Australian national adjustment. ITRF has been updated seven times since ITRF92 resulting in horizontal corrections of less than 4 cm, and a vertical correction of 9 cm (Stanaway & Roberts, 2013). APREF (and eventually ATRF, refer to section 2.1) offers a further densification of the current ITRF in the Australasian region, and is an ideal platform on which to realise the modernised datum.

3.3 Past, Present or Future Reference Epoch

The reference epoch represents a date and time that is conventionally agreed upon, to assist with the meaningful transfer of coordinates, measurements or other parameters. For example, coordinates for a physical feature expressed at reference epoch 2000.0 would be a description of the location of that feature at 00:00:00 01 Jan 2000 in the given datum. To describe that feature at a different point in time requires the application of parameters which describe change with respect to time, i.e. a deformation model. Note that application of a conventional deformation model removes the need to define a reference epoch, however many users (and software) are currently comfortable with the use of reference epochs and will require time to transition to the idea of a fully-dynamic datum.

Intuitively, reference epochs have historically described a ‘snap-shot’ of coordinates at some time in the past. This is particularly salient in that any new spatial data coordinates acquired in WGS84 or ITRF currently need to be transformed and propagated back to the GDA94 epoch of 1994.0. Existing 14-parameter transformation methods may introduce additional uncertainties to the result when propagating across a period of 20 years (Haasdyk & Janssen, 2011). Newer deformation models (e.g. section 2.4) could achieve propagation to a 1994.0 epoch with an accuracy of several millimetres, but may still be subject to undetected deformation.

It is possible, however, to define a reference epoch for the datum at a more contemporary epoch (e.g. 2015.0). The advantages of using such a reference epoch are that any new spatial data acquired with coordinates in the WGS84 or ITRF are, by default, in close alignment with the datum without the need for any data propagation. Unfortunately, tectonic motion will inevitably cause Australian and international datums to diverge at ~7 cm/yr

with resulting differences in spatial data coordinates again reaching 0.5 metres in less than eight years. Any spatial data at higher accuracies will soon again need to be adjusted onto the national datum using appropriate transformation and/or deformation models.

A reference frame with an epoch in the future (e.g. 2020.0) can just as easily be defined using assumptions about future deformation (secular tectonic motion being quite predictable), and computing reference coordinates accordingly. In this way the 'life-span' of the datum would be increased, as the offset with respect to WGS84 or ITRF would first *decrease* by ~7 cm/yr until 2020. This is particularly advantageous for those users who either purposely or inadvertently would not apply the necessary transformation and/or deformation models, such as the current mass-market. Unfortunately these benefits might still be considered to be quite short lived. Even if 20 cm accuracy in future mass-market devices is achieved, then a static datum which deviates by ~7 cm/yr should always be defined with a reference epoch only 2 or 3 years away from the current date.

Some users may have practical reservations about describing or 'predicting' the future location of monuments as localised deformation cannot usually be easily predicted. However, it should be noted that the same issue applies when propagating data to past reference epochs, using imperfect transformation and deformation models. This is especially true for new physical features whose coordinates at the past reference epoch must be inferred, but also applies in areas where past deformation has not been directly measured. Coordinates expressed at any reference epoch are simply a representation (or description) that assists spatial data comparison, and do not necessarily reflect the actual current location of surface points.

Note that regardless of the reference epoch adopted, there will always be a need to transform and propagate spatial data, whether existing and/or newly collected, depending on its spatial accuracy. For example, a change to epoch 2015.0 would necessitate a coordinate shift of ~1.5 metres for all existing GDA94 spatial datasets of metre-level accuracy or better. If mass-market devices do indeed achieve centimetre-level accurate coordinates then manufacturers must either rely on pre-transformed data, or must apply deformation model(s) directly on the device to propagate new data or existing datasets 'on-the-fly'. In either case, a universal improvement in coordinate precision will drive a need for ubiquitous and accurate coordinate propagation and eventually most geospatial software, including on mass-market devices, would ideally apply deformation models commensurate with their location on the globe and their desired positioning accuracy.

3.4 Static vs. Semi-Dynamic vs. Dynamic Datum

3.4.1 Static Datum

As defined above, a static datum is insensitive to any deformation and represents coordinates as fixed in time. The advantages of a static datum are its simplicity and unchanging coordinates, and the minimal overheads in maintaining relationships with other national and international spatial data and datums. Users are accustomed to using such 'fixed' coordinates, but now that positions are increasingly derived from satellite orbits in ITRF (or similar reference frames) which are unaffected by tectonic movements on the Earth's surface, they must now grapple with this extra complication.

The disadvantage is the inability of a static datum to communicate that deformation and tectonic plate motion are measurable and have magnitudes which exceed the positional uncertainty of datum control stations. Coordinates expressed with respect to a static datum can only represent where physical features were at any one point in time.

It is important to note that modern measurements can still be transformed or propagated to a static datum. Consider, for example, how ITRF2008 coordinates in the current epoch are transformed back to GDA94 using the 14-parameter transformation of Dawson and Woods (2010), or how ITRF89 is propagated to ETRF89 by applying a tectonic plate motion model. In general, however, transformation and propagation are easily overlooked by users who may be unaware that modern measurements are at a better accuracy than the underlying spatial data, or simply because the deformation model is not defined *as part of* the datum.

3.4.2 Semi-Dynamic Datum

A semi-dynamic datum has the benefit of reporting coordinates at a single reference epoch, while also incorporating an official deformation model which describes actual movement of physical features between epochs. Coordinates at the reference epoch represent where physical features are/were at a particular point in time. The advantage of a semi-dynamic datum is that it can appear for all intents and purposes as a simple static datum for users of low-precision datasets (e.g. decametre-level), yet still account for deformation in high-accuracy datasets.

Most intuitively, the reference epoch would represent the epoch of a 'snap-shot' of existing physical features (i.e. coordinates at reference epoch 2000.0 describe where physical features were, in both an absolute and relative sense, at epoch 2000.0). Users could apply a deformation model, either a simple tectonic plate model, or a more complex model bases on velocities and 'patches', to retrieve the current coordinates of these physical features. If the deformation model is not applied, then an offset and/or distortions will remain between any newly observed coordinates and any existing spatial data described at the reference epoch. A significant disadvantage is that, in general, spatial data management systems and geospatial software (e.g. GIS applications) are not currently capable

of applying complex deformation models. Although significant work is being undertaken to incorporate dynamics within geospatial software, this currently implies that the semi-dynamic datum is out of reach for all but the most simple deformation models, or the most sophisticated users.

As a compromise, 'reverse-patch' systems have been developed. For example NZGD2000 defines 'reference coordinates' at epoch 2000.0 by starting with the coordinates of physical features in epoch 2000.0 and then distorting them to represent any *relative* non-secular (episodic) deformation between 2000.0 and the current epoch, i.e. deformation due to distinct events such as earthquakes. To retrieve the *true absolute* coordinates in epoch 2000.0 from the 'reference coordinates' in epoch 2000.0, a series of 'reverse-patches' must be applied. (Blick *et al*., 2006; Winefield *et al*., 2010). This model caters well for two users groups: The majority of users are most interested in the current relative geometry between physical features and currently ignore the deformation model entirely. The high-precision geodetic, engineering or scientific user who has the computing power and education to apply the tectonic motion and reverse patches can derive the true coordinate from coordinates or measurements at the current epoch by applying the tectonic motion model and the reverse patches.

3.4.3 Dynamic Datum

A fully dynamic datum simply eschews the concept of reference epoch for the communication of spatial data, and describes coordinates of physical features at any epoch, generally the current epoch. The advantage of this option is that any newly acquired spatial data would be as accurate in the datum as the measurement equipment allows, without the need to apply any transformation and deformation models. Furthermore, this option could be considered to represent what most users of spatial data intuitively want: coordinates which reflect the location of objects today, not coordinates at a reference epoch sometime in the past or future, assuming of course that their map, plan or spatial data are also up-to-date and represented in the same epoch.

However, in order to compare or mix spatial data acquired at different epochs, *all* datasets would need to have the appropriate transformation and deformation models applied, or else they would suffer a commensurate loss of accuracy with respect to the current datum or each other. Further disadvantages include the confusion that could arise by not having an 'official' reference epoch for which all spatial data is referenced. Implicitly, this would require existing spatial data to be propagated to the current epoch (or other agreed convention) for all comparisons and applications, instead of the new spatial data being propagated 'back' in time (or 'forward' to the future) to an official reference datum/epoch. This potentially requires significant computation unless on-demand services become intelligent enough to only transform the data which is of immediate interest. As with the semi-dynamic datum, the fact that most geospatial software packages do not currently make provision for complex deformation models is a significant disadvantage.

Another complexity of a dynamic datum in practice is how local terrestrial measurements (e.g. made by total stations or terrestrial laser scanning) are to be referenced to the correct datum/epoch. For example, if the coordinates of the control stations used for these surveys are expressed in GDA94 and therefore in the reference epoch 1994.0, then the epoch of the surveyed coordinates should also be 1994.0, but could be misconstrued as the date the measurements were gathered.

3.5 Frequency of National Adjustment

The frequent re-adjustment of the ATRF would result in much smaller coordinate changes than those resulting from known tectonic or deformation processes. For example, a future re-adjustment of the national network which includes the latest measurements would improve estimates of positional and relative uncertainty of the control stations in question, and should only correct the coordinates of the control stations within their estimated positional uncertainty. States and territory administrators can shield their users from these 'insignificant' updates by broadcasting coordinate revisions only when significant corrections (due to previous errors or biases) are made.

Additionally, centimetre to millimetre-level dynamics are of increasing importance to earth observation efforts such as by the Global Geodetic Observing System (GGOS) (Plag *et al*., 2009), and could be accounted for by a much more frequent realisation of the datum. Such a concept is being explored for major space geodesy techniques, with the creation of weekly 'epoch reference frames' consistent with ITRF to describe non-linear movements, and to provide updates with a short time delay after deformation events (IAG, 2013). Global dynamic phenomena such as atmospheric and hydrological mass loading occurring on a frequency of hours to years could be incorporated into the datum and/or deformation model for the highest-accuracy users who need to account for them.

Initially it is suggested that from 2015 onward, the national adjustment will be re-run annually to incorporate any new measurements (and possibly retire older, superseded measurements). After that period this frequency could be increased as required, even to the point of weekly or daily re-adjustments to support a fully dynamic datum at the 'current epoch'. Such high-frequency adjustments would only benefit geodetic applications, and could be ignored by most users as high-accuracy 'versions' of the datum with little benefit for centimetre or decimetre-level datasets.

Regardless of the frequency of the national adjustment, states and territories will also continue to compute and publish coordinates and measurements for new ground control stations as required in the intervening period.

4 Comparison of Probable Modernised Datum Realisations

While there are many combinations and permutations of the options presented above, only a few scenarios are likely for the realisation of the modernised datum or ATRF. Several options are almost certain to be adopted, such as the creation of an homogenous national datum using all modern measurements which removes known distortions, and the adoption of the most recent ITRF for the definition of the datum. The frequency of re-adjustment is essentially an academic argument, resulting in millimetre-level changes of importance only to the geodetic user, but still improving the *density and accuracy* of control stations without significantly changing their coordinates. Users will be free to adopt newer adjustments to achieve increased accuracy in their derived datasets, but can, in large part, be shielded from updates of little consequence by correct application of the metadata regarding the version of datum and deformation model applied.

The remaining options of significant concern therefore are the adoption of a 'static', 'dynamic' or 'semi-dynamic' datum, and the chosen reference epoch. The debate revolves around minimising the overall cost and risk to users associated with correctly upgrading existing datasets and obtaining, developing and learning new tools for data manipulation, while still ensuring the highest possible rigour and accuracy required for all user groups.

For the purposes of this paper, the following options for datum realisation are deemed plausible. These are compared in Table 1 and further discussed below:

Option 1) ITRF92(1994.0): Do nothing, maintain (static) GDA94.

Option 2) ITRF92(1994.0): National Adjustment to correct GDA94 distortions only, expressed in ITRF92.

Option 3) ITRF2013(1994.0): National Adjustment to correct GDA94 distortions, expressed in latest ITRF.

Option 4) ITRF2013(2015.0): National Adjustment to contemporary reference epoch.

Option 5) ITRF2013(2020.0): National Adjustment to future reference epoch.

Option 6) ITRF2013(current): National Adjustment to ITRF2013(current) coordinates: A Dynamic Datum.

Option 7) ITRF2013(current) plus ITRF(2013) 1994.0.

Note that either a static or semi-dynamic datum can be defined for each of Options 1 through 5, as each requires the definition of a reference epoch. As previously discussed, decimetre-accurate data must either be gathered within several years of the reference epoch, or have a deformation model applied to propagate the data to the reference epoch, in order to account for the motion of the Australian tectonic plate. Of course, ad hoc adoption of deformation models is possible in a static datum, but the provision of official national deformation model(s) is far more desirable for traceability and consistency. A semi-dynamic datum provides the familiarity for most spatial data consumers, of coordinates which are for all intents and purposes static at a defined epoch, but which also incorporates an official deformation model. For these reasons adoption of a new static datum, as in Option 1, is considered untenable, and the discussion below assumes a semi-dynamic datum for all remaining options except Options 6 and 7.

Options 2 and 3 include a new national adjustment to correct known distortions within GDA94, in order to ensure the highest *relative* accuracy possible. Note that option 3 also accounts for the centimetre-level horizontal bias and decimetre-level height bias between ITRF92 and ITRF2008, while option 2 does not. By retaining the existing GDA94 reference epoch of 1994.0, the main advantage of these options is to minimise the cost and risk of transforming existing GDA94 datasets to a new reference epoch; only data better than approximately 0.5 metre accuracy will need to be transformed. Other benefits to users include a direct agreement between CORS-derived differential positioning and the remainder of the control station network, thus eliminating the need to determine site-specific transformation parameters. Large engineering projects should no longer highlight distortions in the local realisation of the datum, but rather be able to confidently make use of existing control infrastructure. These options do not directly cater for the ~1.5 metre offset due to tectonic motion since 1994.0, which introduces a potential misalignment when new spatial data is acquired using modern GNSS techniques. However, as long as an appropriate deformation model is correctly applied all data could be compared to the 1994.0 epoch at centimetre-level accuracy.

Options 4 and 5 define reference epochs which are either contemporary, or in the future. These options allow the acquisition of new spatial data directly in, or near, the reference epoch. A deformation model is then required to propagate existing data forward from GDA94. This is to cater for new measurements, especially from mass-market users who do not (or cannot) propagate back to epoch 1994.0, or are unaware of datum-related issues. However, if the anticipated decimetre-level precision of new mass-market positioning devices is achieved, specific reference epochs may have unrealistically short life-spans of less than two or three years. Option 5 aims to extend this life-span by allowing the Australian plate to drift towards and then away from its position at the reference epoch, but even then it appears likely that a new reference epoch would need to be proclaimed every five or so years (e.g. 2020, 2025, etc). This would provide users with a set of official common reference epochs in which to collect and compare their data. A disadvantage of these options is that data managers will be forced to undertake the potentially

costly effort of transforming existing datasets to align them to one, or likely several, of the adopted reference epochs. The risk of mis-identification of the data will grow with the adoption of multiple official reference epochs.

Table 1: Comparison of modernised datum realisation options.

Option	**Option 1** ITRF92 (1994.0)	**Options 2,3** ITRF92 or ITRF2013 (1994.0)	**Option 4** ITRF2013 (2015.0)	**Option 5** ITRF2013 (2020.0)	**Option 6** ITRF(current)	**Option 7** ITRF(current) & ITRF2013(1994.0)
Description	Do nothing: retain static GDA94	New national adjustment + deformation model	New national adjustment + new modern ref. epoch + deformation model	New national adjustment + new future ref. epoch + deformation model	New national adjustment + dynamic datum, current epoch + deformation model	Option 6 (e.g. for GNSS, PPP) + Option 3 (for data comparison)
Decimetre distortions in GDA94 corrected	✗	✓	✓	✓	✓	✓
Alignment with existing GDA94 data	✓	✓	✗ ~1.5m offset	✗ ~1.8m offset	✗ ~1.8m offset and growing	✓
New data acquired in WGS84 / ITRF are compatible without further data manipulation	✗ Bias currently greater than metre-level & 9 cm ITRF92 vertical bias	✗ Bias currently greater than metre-level	✓ However bias increases (~7cm/a) from 2015	✓ However bias increases (~7cm/a) from 2020	✓ No bias for new data	✓
Accounts for rotation of tectonic plate without further data manipulation	✗ 7 mm / 30 km baseline from 2015	✗ 7 mm / 30 km baseline from 2015	✓ Bias increases from 2015	✓ Bias increases from 2020	✓ No bias for new data	✓
Datum accuracy if deformation model is applied	cm-level	sub cm-level	mm-level (limited lifespan)	mm-level (limited lifespan)	mm-level	mm-level

Option 6 defines a fully dynamic datum which caters specifically for two user-groups. Existing geodetic users of spatial data are already familiar with the use of dynamic datums such as ITRF. These users require raw, untransformed data, in its original epoch, for the assessment of global dynamic phenomena. They are often pushing the bounds of available accuracy, and may be testing the assumptions and algorithms underlying official deformation models and other un-modelled dynamic phenomena. In terms of emerging spatial data markets, a fully dynamic datum ensures that mass-market devices are positioning implicitly in the Australian datum. However this would require the frequent (and external) transformation of datasets prior to uploading to the device. This may prove to be more troublesome than applying the transformations directly on the device, in order to propagate the new data to a conventional reference epoch as in Option 5.

Finally, Option 7 provides a combination of datum products, both fully and semi-dynamic, to cater for all user requirements. For example, dynamic ITRF (or ATRF) coordinates at the current epoch are determined for any initial GNSS positioning (e.g. PPP), baseline processing, adjustment, Lidar and other remote sensing techniques. All spatial data would be archived without changes in ITRF (or ATRF) at the epoch of acquisition, to facilitate high-precision geodetic analysis and/or propagation to another epoch in the future. An official deformation model (e.g. Stanaway & Roberts, 2013) would be used to propagate coordinates and datasets with sub-centimetre-level accuracy to an updated GDA94 (re-realised as ITRF2013 at epoch 1994.0) Eventually, improvements in GIS

technology will enable 4D coordinates (3D + epoch) to be modelled seamlessly within the GIS, and the need for a fixed epoch will diminish. The mixed option presented here provides the best of both worlds, while minimising cost and risk to nearly all users. It also promotes sound data management principles by encouraging the storage of data in the most accurate reference frame (ITRF or ATRF at the epoch of data acquisition), and treating coordinates for the dataset in the national datum as a derived product to facilitate easy integration of datasets. Importantly, the choice of a single reference epoch at 1994.0 minimises the risk of maintaining multiple similar datasets and therefore reduces the potential for confusion.

In this debate the importance of metadata must be stressed. It is widely recognised that metadata management is rarely approached with appropriate rigour, and that large datasets are often manipulated without the maintenance and propagation of appropriate metadata. The potentially large number of reference epochs, deformation models and adjustments discussed above mean that metadata will become just as important as the data themselves. It is important that all data remain 'traceable' back to its source so as to allow, for example, for future application of updated deformation models if and when they become available. All data should carry metadata about the epoch, datum, method of acquisition, and estimated uncertainty. Where possible the original raw data should be stored unchanged at the epoch of acquisition, and at the very least, by properly recording any transformations or alterations undertaken, the raw data can be re-processed for future use.

5 Concluding Remarks

The development of a modernised datum for Australia – what the authors refer to here as the Australian Terrestrial Reference Frame – is currently underway. However there are still a variety of options to consider regarding the final realisation of the datum. The discussion above concludes that there is obvious benefit in adopting the most accurate and homogenous datum based on the most recent ITRF, a national adjustment of all available measurements, and an accurate deformation model. It is imperative to construct the most rigorous datum possible, designed for the highest-precision geodetic applications, but which will also cater for all other users. Another static realisation of the datum cannot account for recent technological improvements, not to mention the inexorable movement of the Australian tectonic plate.

The major remaining issues are the choice between a semi-dynamic or full-dynamic datum realisation, the selection of reference epoch, and the complexity of the deformation model. Ultimately, the options under consideration each cater for different user groups, and a combination of datum products is the best solution for all users. In particular, the combination of a fully dynamic ATRF defined in the latest ITRF(current epoch), along with a 'derived datum' with all relevant data propagated to the 1994.0 epoch, would provide for the highest accuracy applications as well as the burgeoning but 'uninformed' positioning mass-market, while still minimising the costs and risks associated with data manipulation and metadata mis-management.

Regardless of the option(s) chosen for datum realisation, tools are yet to be developed that correctly apply the highest-accuracy transformation and deformation models. A new variable gridded deformation model for the Australian continent has been proposed but it has yet to be rigorously tested. There will be a need for user-education, at the level of the geospatial practitioner and data-manager, and tools to insulate the mass-market user from unnecessary complications. Although most geospatial software packages can perform simple transformations, they are not explicitly designed to routinely propagate coordinates through time and definitely cannot cater for more complex deformation modelling. However, the market drive to do so is increasing and the Australian geospatial profession has a great opportunity to drive the development of these high-accuracy products.

While the focus of this paper has been on datum development in Australia, by also attempting to cater for changing technologies and complex deformation, the concepts discussed herein are applicable to national and international datum development across the globe.

Acknowledgements

This work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centres Programme.

References

Altamimi, Z., Collilieux, X. and Métivier, L. (2011). ITRF2008: an improved solution of the International Terrestrial Reference Frame, Journal of Geodesy, 85 (8), 457-473.

Blick, G., Crook, C., Grant, D., and Beavan, J. (2006). Implementation of a Semi-Dynamic Datum for New Zealand, *International Association of Geodesy Symposia*, (128), Sapporo, Japan 30 June – 11 July.

Blick, G., Donnelly, N, Jordon, A. (2009). The Practical Implications and Limitations of the Introduction of a Semi-Dynamic Datum – A New Zealand Case Study. *Geodetic Reference Frames*, International Association of Geodesy Symposia, (134), 115-120.

Boucher, C., Altamimi, Z. (2011). Memo: Specifications for reference frame fixing in the analysis of a EUREF GPS campaign, http://etrs89.ensg.ign.fr/memo-V8.pdf (accessed March 2014).

Collier, P., Argeseanu, V., Leahy, F. (1997). Distortion Modelling and the transition to GDA94, *The Australian Surveyor*, Vol. 43, No. 1, March 1998.

Collier, P., Steed, J. (2001). Australia's National GDA94 Transformation Grids, *42nd Australian Surveyors Congress*, Brisbane, 25-28 Sept 2001.

Crook, C., Donnelly, N. (2013). Updating the NZGD2000 deformation model, *125th New Zealand Institute of Surveyors Conference*, Dunedin, 29-31 August 2013, 40-46. http://nzisconference.org.nz/assets/2013-Conference-Uploads/joint-proceedings-nzis-sircnzV1.pdf (accessed Dec 2013).

Dawson, J. and Woods, A. (2010). ITRF to GDA94 coordinate transformations, *Journal of Applied Geodesy*, Vol 4, 189 – 199.

Donnelly, N., Crook, C., Haasdyk, J., Harrison, C., Rizos, C., Roberts, C., Stanaway, R. (2013) Dynamic Datum Transformations in Australia and New Zealand. *Proceedings of Surveying and Spatial Sciences Conference (LOCATE14)*, 7-9 April, Canberra, Australia (in press)

Donnelly, N., Haasdyk, J. (2013) A Dynamic Approach to Datums. https://www4.gotomeeting.com/register/689873399, (video recording, accessed Dec 2013)

Featherstone, W.E. (2013). An updated explanation of the Geocentric Datum of Australia (GDA) and its effects upon future mapping, GA (2012a) Geocentric Datum of Australia (GDA), http://www.ga.gov.au/earth-monitoring/geodesy/geodetic-datums/GDA.html (accessed Dec 2013).

Featherstone ,W.E., Filmer, M.S., Penna, N.T., Morgan, L.M. Schenk, A. (2012). Anthropogenic land subsidence in the Perth Basin: Challenges for its retrospective geodetic detection. *Journal of the Royal Society of Western Australia* (2012) 95(1), 53-62.

Fraser, R., Leahy, F., Collier, P. (2014). DynaNet User's Guide.

Gakstatter, E. (2013a). Reflecting on 2013, from Your Perspective. *GPS World*. http://gpsworld.com/reflecting-on-2013-from-your-perspective/ (accessed Dec 2013).

Gakstatter, E (2013b). Nightmare on GIS Street: GNSS Accuracy, Datums and Geospatial Data. http://geospatial-solutions.com/nightmare-on-gis-street-accuracy-datums-and-geospatial-data/ (accessed Dec 2013).

Geoscience Australia (2013). Asia-Pacific Reference Frame (APREF), http://www.ga.gov.au/earth-monitoring/geodesy/asia-pacific-reference-frame.html (accessed Dec 2013).

Haasdyk, J., Janssen, V. (2011). The many paths to a common ground: A comparison of transformations between GDA94 and ITRF. *Proceedings of IGNSS Symposium 2011 (IGNSS2011)*, Sydney, Australia, 15-17 November.

Haasdyk, J., Roberts, C. (2013). Monitoring station movement using a state-wide simultaneous 'adjustment of everything' − Implications for a next-generation Australian datum. *Proceedings of IGNSS Symposium 2013 (IGNSS2013)*, 16-18 July, Gold Coast, Australia.

Haasdyk, J., Roberts, C., Janssen, V. (2010). Automated monitoring of CORSnet-NSW using the Bernese software, *Proceedings of XXIV FIG International Congress 2010*, Sydney, Australia, 11-16 April.

Haasdyk, J., Watson, T. (2013). Data-mining in NSW: Working towards a new and improved Australian datum, *Proceedings of the 18th Association of Public Authority Surveyors Conference (APAS2013)*, 12-14 March, Canberra, Australia, 85-102.

IAG (2013). Strategies for Epoch Reference Frames. (http://erf.dgfi.badw.de/) (accessed Dec 2013).

ICSM (2006). Geocentric Datum of Australia technical manual, version 2.3, http://www.icsm.gov.au/gda/gdav2.3.pdf (accessed Dec 2013).

ICSM (2013a). Geocentric Datum of Australia 1994 (GDA94), http://www.icsm.gov.au/gda/index.html (accessed Dec 2013).

ICSM (2013b). Adjustment of the combined state and territory geodetic networks, http://www.icsm.gov.au/gda/spine.html (access Dec 2013).

ICSM (2013c) Standards for the Australian Survey Control Network, (SP1), version 2.0 http://www.icsm.gov.au/geodesy/sp1.html (accessed Dec 2013).

Janssen, V., Haasdyk, J., McElroy, S. (2011). CORSnet-NSW Network RTK: Same look and feel... only better. *Proceedings of Association of Public Authority Surveyors Conference (APAS2011)*, 6-7 April, Bathurst, Australia, 39-54.

Leahy, F.J., Collier, P.A. (1998). Dynamic network adjustment and the transition to GDA94, *The Australian Surveyor*, 43(4), 261-272.

LINZ (2013). NZGD2000 Deformation Model Format, http://apps.linz.govt.nz/ftp/geodetic/nzgd2000_deformation_20130801_full.zip (accessed Dec 2013).

NCI (2013). Raijin, http://nci.org.au/nci-systems/national-facility/peak-system/raijin/ (accessed Dec 2013).

Ng, A H-M. Ge, L. Rizos, C. (2010). Horizontal and Vertical Movement Estimation for Underground Mining using ALOS PALSAR, Engineering Geology, 143-144, 18-27.

Ordnance Survey (2013). A guide to coordinate systems in Great Britain, http://www.ordnancesurvey.co.uk/docs/support/guide-coordinate-systems-great-britain.pdf (accessed March 2014).

Pearson, C., Snay, R. (2013). Introducing HTDP 3.1 to transform coordinates across time and spatial reference frames, *GPS Solutions*, Vol 17(1), 1-15.

Petit, G., Luzum, B. (eds.) (2010). IERS Conventions (2010), IERS Technical Note 36.Verlagdes BundesamtsfürKartographie und Geodäsie, Frankfurt am Main, Germany.

Plag, H. P., Gross, R., Rothacher, M. (2009). Global geodetic observing system for geohazards and global change. Geoscience BRGM Journal of Sustainable Earth (9), 96-103.

Rizos, C., Janssen, V., Roberts, C., Grinter, T. (2012). Precise Point Positioning: Is the Era of Differential GNSS Positioning Drawing to an End?, *FIG Working Week*, Rome, Italy 6 – 10 May.

Simons, M. and Rosen, P. (2007). Synthetic Aperture Radar Geodesy, *Treatise on Geophysics*, 3, 391-446.

Stanaway, R., Roberts, C. (2009). A Simplified Parameter Transformation Model from ITRF2005 to any Static Geocentric Datum (e.g. GDA94), *Proceedings of IGNSS Symposium 2009 (IGNSS2009)*, Surfers Paradise, Australia, 1-3 December.

Stanaway, R., Roberts, C., Blick, G. (2011). Realisation of a Geodetic Datum using a gridded Absolute Deformation Model (ADM), Earth on the Edge: Science for a Sustainable Planet. *IAG Symposium Series* (139), *Proceedings of the IAG General Assembly*, Melbourne, Australia, 28 June-2 July, 2011, Springer.

Stanaway, R., Roberts, C., Blick, G., Crook, C. (2012). Four dimensional deformation modelling, the link between international, regional and local reference frames, *Proceedings of FIG Working Week 2012*, Rome, Italy, 6-10 May.

Stanaway R., Roberts, C. (2013). A High-Precision Deformation Model to support Geodetic Datum Modernisation in Australia, Proceedings *of IAG Scientific Assembly, Potsdam, Germany, 2013* (to be published).

Tregoning, P. (2003). Is the Australian Plate deforming? A space geodetic perspective, Geological Society of Australia Special Publication, 22 and Geological Society of America Special Publication, 372, 41-48.

Tregoning, P., Burgette, R., McClusky, S.C., Lejeune, S., Watson, C.S., McQueen, H. (2013). A decade of horizontal deformation from great earthquakes. *Journal of Geophysical Research: Solid Earth* (118), 1–11.

Winefield, R., Crook, C., and Beavan, J. (2010). The Application of a Localised Deformation Model after an Earthquake, *Proceedings of XXIV FIG Congress 2010*, Sydney, Australia, April 11-16.

Vertical accuracy assessment of LiDAR ground points using minimum distance approach

Seyedhossein Pourali* Colin Arrowsmith* Nicholas Chrisman* Aliakbar Matkan**

*School of Mathematical and Geospatial Sciences, Royal Melbourne Institute of Technology (RMIT University), GPO Box 2476, Melbourne, 3001.
(hossainpourali@yahoo.com, colin.arrowsmith@rmit.edu.au, nicholas.chrisman@rmit.edu.au)

**Centre of Remote Sensing and GIS, Shahid Beheshti University, Daneshju Blvd, Tehran 1983963113, Iran.
aa_matkan@sbu.ac.ir

Abstract

Light Detection And Ranging is now being widely used to provide accurate digital elevation model (DEM). One common method used to determine the accuracy of LiDAR - Light Detection And Ranging - vertical accuracy is compared LiDAR-derived DEM elevation with survey-derived ground control points (GCPs). However, because the DEM elevations are generalised to the areas covered by one cell, inherent errors when compared to the elevation of a GCP are evident. This paper presents a method based on a minimum distance approach using the so called first law of geography to assess the accuracy of LiDAR ground point dataset. The result has shown that the tested LiDAR ground point dataset is suitable for applications that do not need a vertical accuracy better than 0.5m.

1 Introduction

The Victorian Government in Australia has embarked upon a program, called "Future Coast", which involves acquiring vertical heights along the Victorian coast using Airborne Light Detection and Ranging, or LiDAR. The intent of this program is to construct detailed digital elevation models (DEMs) for the littoral zone from the high tide water level inland up to an elevation of ten metres. Therefore the inland horizontal extent of this data set varies due to the irregularities in the coastal topography, involving venturing only several metres inland on cliffs and stretching up to several kilometres on low-lying land. The metadata that comes with the dataset claims a vertical accuracy of 0.1 metres (DSE, 2010).

Aguilar et al. (2010) state that the best achievable realistic vertical accuracy in open terrain is around 0.15 metres. A similar result has been reported by Hodgson and Bresnahan (2004). However, this level of accuracy can rarely be achieved (Aguilar et al.,2010). Although LiDAR has been acquired for the specific needs of the Future Coast project in Victoria, there are demands for using just LiDAR data for various applications, ranging from research to public resource management (Hodgson et al.,2005). There is a need to convey the limitations of the LiDAR data to end users. This is a global issue, not limited to Victoria, Australia.

An accurate topographic dataset is important for supporting decisions in water resource management. Local governments, who are responsible for local flood management, can use LiDAR for drainage design in the context of stormwater and overland flood management. Whilst the accuracy of the LIDAR acquired heights has been reported by the Future Coast project, there are concerns regarding the vertical accuracy of LiDAR dataset in different regions.

The accuracy of LiDAR depends on the full workflow from data acquisition through to the final derived DEM. During acquisition, the accuracy of positioning using the Inertial Measurement Unit (IMU) located on the aircraft, and the on-board Global Positioning System receiver (GPS) will inform the accuracy of the point clouds acquired. These devices establish the position of the platform, and then the lasers measure distances at different angles. Multiple returns are possible from objects in the landscape, including vegetation, buildings and bare ground. Errors or bias can be introduced during LiDAR point cloud segregation when a filtering process to separate ground and non-ground points is applied. Most of these effects introduce systematic errors that are non-

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

normally distributed and exhibit kurtosis and skewness in their distribution. According to Aguilar and Mills (2008) the nominal stated accuracy for LiDAR elevations quoted by vendors assumes a normal distribution and are typically derived from limited checks. It would seem more prudent to present accuracy with a lower and upper Root-Mean-Square-Error (RMSE) at a specific confidence level.

This present is a study which examines the vertical accuracy of LiDAR ground point data using survey permanent marks. The accuracy is determined using the Root Mean Square Error (RMSE). However, RMSE is sensitive to the outliers, and the skewness and the kurtosis effect of non-normal error distribution (Zandbergen,2008). Therefore, a robust measure is preferred where error distribution is not normal (Höhle and Höhle,2009). The present study uses the statistical method which includes the upper limit of the RMSE value in order to account for the skewness and kurtosis effects on the final accuracy assessment.

The remainder of this paper is organised as follows. Firstly, a brief overview of airborne LiDAR technology and a short background about LiDAR data vertical accuracy assessment is presented. The remainder focuses on a case study area and the sources of available elevation data. The methodological section introduces the used datasets for this study, the workflow used to compare GCPs and LiDAR ground point dataset, and the statistics have been used to assess LiDAR vertical accuracy.

2 Airborne LiDAR 3D point cloud

The basic operation of a LiDAR survey involves some platform (airplane, helicopter or terrestrial vehicle) which carries the LiDAR instrument that sends out a laser pulse and records a distance and (optionally) reflected intensity at each angle. Therefore, LiDAR technology is considered an active remote sensing which actively transmits pulses toward target areas as shown in Figure1(Wehr and Lohr,1999).

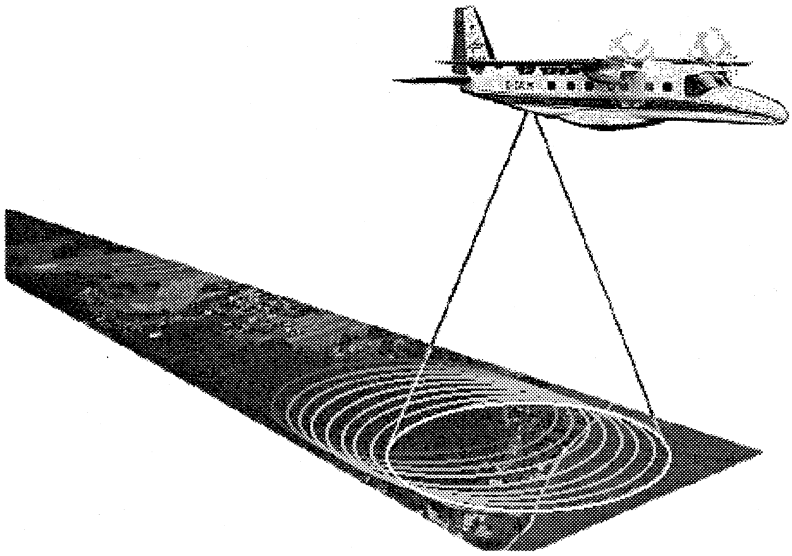


Figure 1. LiDAR footprints (Wehr and Lohr,1999)

The footprint characteristics of the LiDAR reflected pulses are controlled by beam divergence and flying height (distance from target). The usual commercial beam divergence is between 0.2 and 0.8mrad giving a footprint size between 0.2m and 1.10m at a one kilometre flying height (Lemmens,2011). Ground locations for LiDAR are obtained from post processing of Differential Global Navigation Satellite System (DGNSS) data or directly in real-time, through the integration of LiDAR data with GPS data (Habib et al.,2005). Data is collected in terms of latitude, longitude, and ellipsoidal height based on the reference ellipsoid,WGS1984. Then the acquired data needs to be transformed to the formally accepted regional or national horizontal and vertical datum before further use. The vertical or height datum represents the ortho-metric height value while GPS height is the ellipsoidal height value(Liu et al.,2007).

Acquiring LiDAR reflected pulses can be either discrete or in full waveform. Discrete pulse returns are captured at two or more vertical levels depending upon the ground characteristics. Crowns of trees can return one pulse at one level, whilst the ground will return a pulse at a lower elevation.

Since 2008, Multiple LiDAR Pulse in Air (MPiA) LiDAR has been used, whereby a larger ground area can be covered, or lower flying heights can be used, without reducing spatial resolution. Flying at lower altitudes

results in lower air turbulence. Therefore MPiA enables LiDAR data to be acquired more cheaply than traditionally acquired LiDAR data (Lemmens et.al (2011); Mallet and Bretar (2009). The result of these surveys is termed a point cloud, since it takes further processing to turn the raw results into a surface model for a particular purpose.

3 Background into determining LiDAR vertical accuracy

The accuracy, and in particular vertical accuracy, of LiDAR is highly dependent on errors of parameters such as flying height, location (from on-board GPS) and inertial measurement units (IMU) errors, distance to ground station and LiDAR post-processing (Hodgson and Bresnahan,2004; Hodgson et al.,2005).The common method used to determine the relative vertical accuracy of LiDAR derived elevations is to compare the LiDAR-derived DEM against ground control points (GCP) (Maune et al. 2007).

The GCP points can be permanent survey marks or GPS points. Besides the common accuracy assessment using LiDAR-derived DEM (Maune et al.,2007), GPS points can be surveyed at exact location of interest LiDAR points according the approach suggested by Hodgson and Bresnahan (2004). Furthermore, it is possible to extract GCP values for points of interest using all surrounding LiDAR points. The latter approach was used to develop a tool by Webster and Dias (2006). Re-measuring the LiDAR points by accurate GPS is time consuming and labour-intensive while using the tool developed by Webster and Davis (2006) is fast and user-friendly. Survey marks which have already been established as a basis for determining the coordinate position of spatial datasets are used to reduce time and effort. Furthermore, using a sufficient number of check point with proper spatial distribution across study area is required to undertake a LiDAR data accuracy assessment. Thirty points are deemed to be the minimum requirement for each prominent land cover across a study area according to the American Society for Photogrammetry and Remote Sensing (ASPRS) and ICSM (Inter-Governmental Committee on Surveying and Mapping) (ICSM,2008). Hodgson et al. (2003) claim that LiDAR data accuracy varies for different land cover conditions. It is also claimed that LiDAR data accuracy in flat areas are twice as accurate as those in steeper landscapes with forest coverage (Hodgson et al.,2003). ASPRS (Flood,2004) and ICSM (ICSM, 2008) guidelines have separated the vertical accuracy assessment into three land categories in order to accurately assess LiDAR vertical accuracy. Three major categories for accuracy assessment include open terrain suitable to support fundamental vertical accuracy (FAV), areas covered by various land cover to support the supplemental vertical accuracy (SVA), and combined land cover area to support consolidated vertical accuracy (CVA). It is assumed that FAV has the highest accuracy in estimating terrain surface height value due to there being limited barriers for LiDAR pulses. LiDAR vertical accuracy assessment can be reported at 95% confidence level by 1.96(RMSE) for FAV and 95th percentile for SVA and CVA. Regarding to report accuracy assessment, the ICSM (2008) has recommended that the 95th percentile value of accuracy should be adhered to for FAV, SVA, and CVA. However, Liu (2011) claimed that the accuracy value reported by of 95th percentile overestimates vertical accuracy in compare with vertical accuracy reported by 1.96(RMSE).

4 The study area

The test study was conducted for the Bass Coast Shire council area lying approximately 130 kilometres to the south-east of Melbourne, Victoria, Australia. This area is experiencing unprecedented population growth due to many people wanting to retire close to the sea, as well as general regional pressures. The Bass Coast Local Government Area had the highest growth rate in Victoria between 2001 and 2006 of around 21 precent (Buxton, 2008). Growth has highlighted the need for appropriate development in what is a predominantly a low lying coastal plain. The area is exposed to rainfall from frontal systems from the Southern Ocean. The coastal area is also fed overland by mountainous terrain to the north. All drainage water for this area flows into Bass Strait (Figure 2). Developers' requests to subdivide new estates within Bass Coast Shire Council's jurisdiction are increasing. Two important policy documents control water resources management for the Bass Coast Shire Council. Land use control is managed through the Bass Coast Flood Management Plan and the Integrated Water Management plan (formerly the Stormwater Management Plan).

Figure 2: Inverloch test site in Bass Coast Shire Council (Source: Bass Coast Shire Council)

The digital representation of a terrain surface which is the core dataset in hydrological modelling and flood management, is available through elevation data resources in Victoria, Australia. VICMAP, a digital resource of natural and cultural features covering the entire state of Victoria, provides the most comprehensive coverage of elevation data, DEM. Each DEM file has a 20m by 20m cell size, with the cell value interpolated from digital contours at 20 meters spacing on a map series at 1:25000. One source of elevation is the LiDAR dataset. However, LiDAR is not provided for the whole state. The LiDAR dataset as a reliable method for acquiring 3D data is useful because of its higher resolution as well as its vertical accuracy. LiDAR-derived DEMs applications are various, for example, they are suitable for extracting detailed overland flow paths, refined catchment boundaries and providing a better estimation of hydrographs for hydrological modelling.

5 Methodology

The goal of this work is to assess the vertical accuracy of the Future Coast LiDAR bare-ground points dataset in the given study area. This requires the collection and preparation of observed elevations and the GCPs. Common comparisons can be made using an interpolation technique for transforming sample elevation points into a DEM (gridding method), then comparing the cell value to a corresponding GCP such as survey permanent marks (PMs). The gridding method introduces error due to the influence of averaging elevations over a given cell.

Interpolating elevations derived by LiDAR, to determine unknown heights, will ultimately depend upon the interpolation method used, including the input parametres such as neighbouring distance and the effect of direction (anisotropy).

An alternative approach is to undertake a one-to-one comparison between the LiDAR ground points height and the elevation derived from GCPs. This approach was used by Hodgson et al. (2003). This approach is suitable to accurately compare heights because errors introduced through gridding are eliminated. However, the method is time-consuming.

The third approach is use a proximal point algorithm suggested by Webster and Dias (2006). In this method the user determines a search radius around target GCPs. This automatically produces minimum, maximum and other statistical values relating to the GCPs and the LiDAR dataset comparison. The values for the heights of the GCPs are compared to the results from interpolation of LiDAR points for a defined buffer around the GCPs. Although the proximal approach facilitates the comparison between the LiDAR dataset and the GCPs, it uses an interpolation technique to estimate GCPs height values and compare on a one-to-one basis between the LiDAR and the GCP elevation values. Furthermore, there are no rules to limit the search distance.

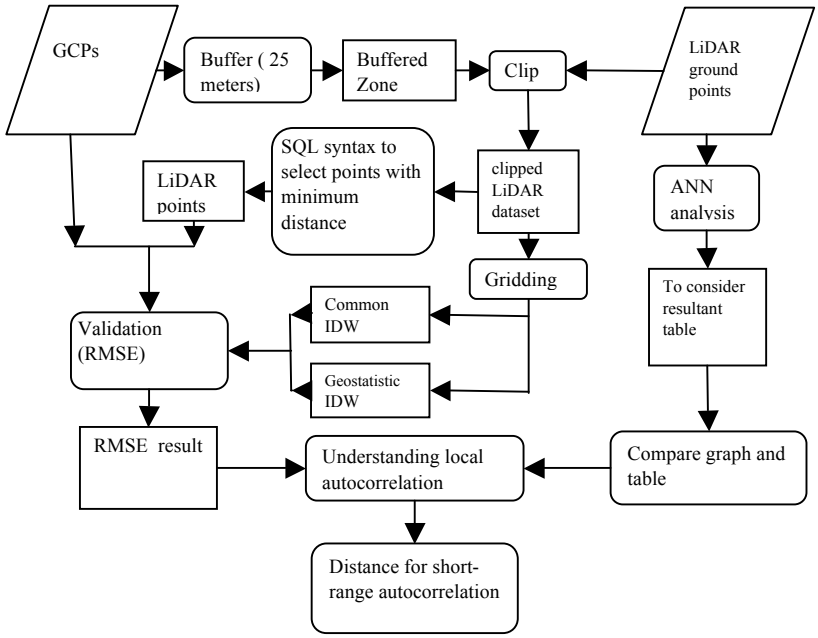
In order to overcome the search radius problem and avoid imposing gridding errors into the procedure, the solution is to undertake a one-to-one comparison between the GCPs and the LiDAR sample points using a autocorrelation distance (minimum distance) based on Tobler's first law of geography (Tobler,1970). This can be used to assess vertical accuracy of LiDAR dataset. Therefore, the LiDAR ground points within a spatially autocorrelated distance (closer features are similar) should represent similar height values as the GCPs. This approach has been used in this study.

Although the search distance can be manually defined by user, increasing the search distance will result in the number of elevations being compared and therefore relevant processing time will increase. It also results different RMSE values. One metre distance has been chosen for proximity analysis in this study. One metre selected due to the common pixel size used in the LiDAR-derived DEM. Within the one meter search distance, those points which are located at the first minimum distance from the checkpoints selected are compared, then the points that are located at the second, third and fourth sequential minimum distances have been chosen and compared using a separate comparison procedure.

The present study also takes the autocorrelated distance using average nearest neighbour analysis (ANN) and compared derived distance with the distance, which shows the lowest RMSE, extracted from proposed method developed for this study, the sequential minimum distance.

Another comparison has been done using GCPs and DEM derived from LiDAR ground points at 25m around each GCPs using common Inverse Distance Weigh (IDW) and geostatistic IDW interpolation techniques to define the influence of interpolation methods on LiDAR DEM vertical accuracy.

Flow chart 1 shows the work flow steps used in this study. Flowchart 1 also shows the process from which it is possible to find the relation between the difference in value and short-range autocorrelated distance. Short-range autocorrelation distance represents the distance over which it is expected that LiDAR points have the lowest difference to GCPs.



Flowchart 1. Present study workflow

6 LiDAR dataset

The test data was selected for an area around the township of Inverloch. This area was selected because new residential development is permitted in the urban fringe area, the availability of data, and the importance of the area in a local flood management plan. The test data was sourced from the Future Coast program provided to the Bass Coast Shire Council by the Victorian Government Department of Environment and Primary Industry (DEPI). The data consisted of LiDAR points covering a two kilometre by two kilometre area. The vertical accuracy (DSE, 2010) is stated as +/- 0.10m with 68% confidence, with the positional accuracy being +/- 0.35m.

Data was projected on the Geocentric Datum of Australia 1994 (GDA94) and elevations defined by the Australian Height Datum (AHD). The data had been captured in 2007 and has been accessible to users since 2009.

7 Ground Control Points (GCPs)

Ground control points such as survey permanent marks (PMs), that have previously been established for elevation and position, using conventional surveying and levelling techniques, were used to assess the accuracy of LiDAR ground point dataset. The suitability of PMs in Victoria was discussed in Liu (2011). Liu (2011) stated that 80% of PMs in rural area have vertical accuracy better than +/- 0.20m and 90% of PMs in Township area have vertical accuracy better than +/-0.03m due to the greater accuracy requirements and therefore greater survey accuracies applied for urban PMs. PMs have a density of 0.63 marks per square kilometre in Victoria. This is around ten times the density for other Australian states (Liu,2011). Figure 3 shows the geographic location of 85 PMs that were available throughout the study area.

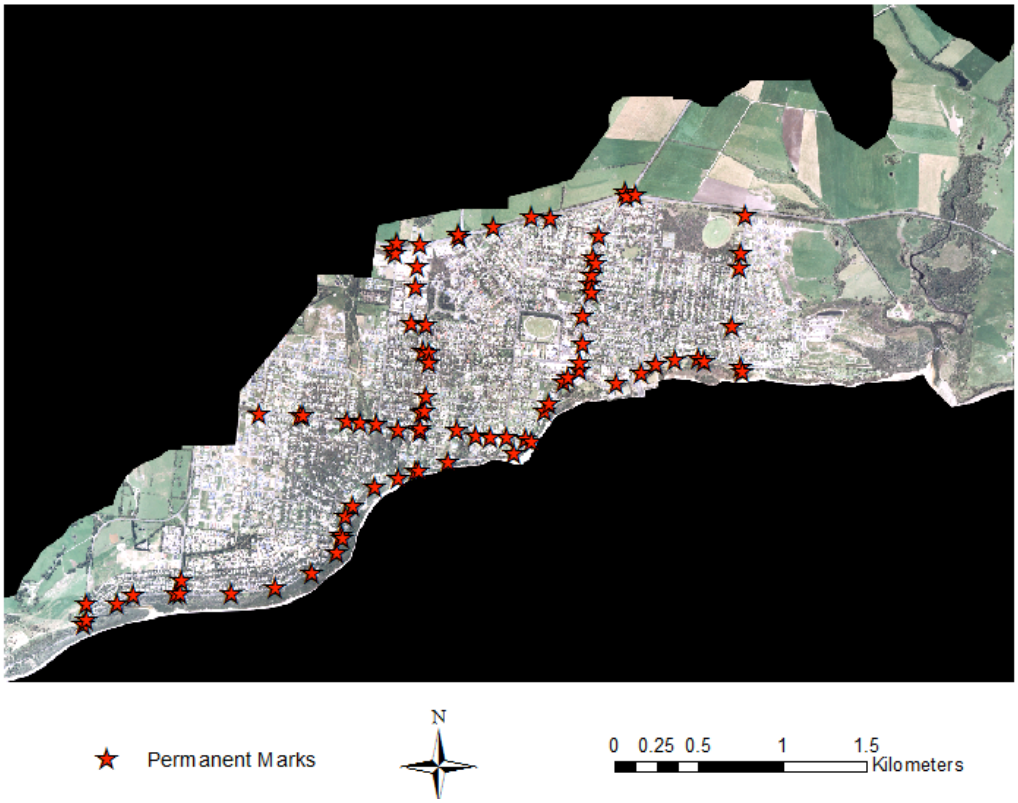


Figure 3: Permanent Survey Marks location in Inverloch, Victoria, Australia (Source: Bass Coast Council's GIS database)

8 Statistics for vertical accuracy assessment

Root Mean Square Error (RMSE) is a frequently used measure to understand overall accuracy assessment (Carlisle,2005). Furthermore, Aguilar and Mills (2008) have suggested using the upper and lower RMSE error for showing the range of error instead of representing only the average error value by RMSE. The upper and lower limits show the influence of kurtosis and skewness of error distribution. This gives the end user an appreciation of the range of error each vertical value may have. This means the effect on non-normal distribution of error can be included in accuracy assessment. Equations 1 to 3 show RMSE, RMSE upper (worst accuracy) and RMSE lower.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Z_{\mathrm{model}(i)} - Z_{check(i)})^2}{n}} \quad (1)$$

$$RMSE_{lower} = \sqrt{mse - \frac{\frac{\gamma_{2mse}+2}{\gamma_{1mse}} - \sqrt{(\frac{\gamma_{2mse}+2}{\gamma_{1mse}})^2 + \left(4\left(\frac{t_\alpha\sqrt{(\gamma_{2mse}+2)(\gamma_{2mse}+2-\gamma_{1mse}^2)}}{|\gamma_{1mse}|}+1\right)\right)}}{2}\sigma_{mse}} \quad (2)$$

$$RMSE_{upper} = \sqrt{mse + \frac{\frac{\gamma_{2mse}+2}{\gamma_{1mse}} - \sqrt{(\frac{\gamma_{2mse}+2}{\gamma_{1mse}})^2 + \left(4\left(\frac{t_\alpha\sqrt{(\gamma_{2mse}+2)(\gamma_{2mse}+2-\gamma_{1mse}^2)}}{|\gamma_{1mse}|}+1\right)\right)}}{2}\sigma_{mse}} \quad (3)$$

Where mse is mean square error , σ_{mse} is the mse standard deviation, γ_{1mse} shows mse skewness and γ_{2mse} is mse standardised kurtosis, t_α is statistical t-student test. For further information on equation parameters, see Aguilar and Mills (2008); Aguilar (2005); Aguilar and Aguilar (2007).

9 Analysis

For the test LiDAR data in Inverloch, each PM was buffered by 25m, then LiDAR ground points were clipped inside each buffer to reduce the LiDAR points to be processed(Figure 4). Distance to each PM was calculated for all LiDAR ground points within 1m around PMs. RMSE values were estimated based on individual comparing between elevation values of LiDAR points located in first, second, and subsequent minimum distances from each PMs. RMSE value for each minimum distance considered to identify the distance where RMSE start to increases due to reduces of similarity. Short-range autocorrelation distance, alternatively, can be determined by average nearest neighbour (ANN) analysis.

The RMSE value was then estimated by including all LiDAR points located in 1m from PMs in order to identify the influence of averaging technique occurring in gridding process. The one metre distance was chosen due to the typical cell size in the existing LiDAR-derived DEM for the study area. Finally, RMSE values were estimated between LiDAR DEMs - derived using Inverse Distance Weight (IDW) and Geostatistic IDW interpolation methods - and PMs using the suggested method by ASPRS (Flood,2004) and ICSM (ICSM, 2008) to identify the influences of used interpolation techniques.

Figure 4: The LiDAR dataset in each buffer around PMs

10 Results

Using the procedure described in section 9, the accuracy assessment between LiDAR points and PMs. Table 1 shows the results that have been extracted through the analysis.

Table 1: Resultant RMSE for the minimum distance at different distances around PMs. N^i shows i^{th} minimum distance

	Average distance (m)	RMSE	Upper-RMSE
N^1_Minimum	0.366	0.241	0.377
N^2_Minimum	0.625	0.257	0.410
N^3_Minimum	0.980	0.259	0.413
N^4_Minimum	1.252	0.270	0.428
Combination of N^1 to N^4	1	0.259	0.373
IDW	1	0.228	0.364
Geostatistic IDW	1	0.249	0.398

The first column in Table 1 shows the average distance from each PMs. The second and third columns show the RMSE and Upper-RMSE (worst accuracy) for each individual comparisons. Results in Table 1 show that the closest LiDAR ground points to the PMs (0.366m) have the lowest value of RMSE (0.241). The RMSE increases from the second minimum distance. Combining the value for all LiDAR points with a minimum distance to PMs and calculating the RMSE value shows a 0.259 for the RMSE. Using two Inverse Distance Weight (IDW) and geostatistic IDW interpolation techniques to estimate the value for PMs has shown that using the simple common IDW technique represents the lowest value RMSE (0.228) while the geostatistic IDW method has shown a RMSE equal to 0.249. The result of this research shows that the geostatistic IDW-derived DEM with one square metre spatial resolution gives a better estimate of elevation for the study area than common IDW, which leads to underestimate derived elevation error. The inherent error in the LiDAR ground point dataset expected to be increased in developed DEM.

Due to error normal distribution (Figure 5), vertical accuracy is estimated using 1.95*RMSE at a 95 percent confidence level (Flood,2004). Normality checks on the error distribution has been conducted using the Kolmogorov-Smirnov test. Assuming a value of 0.259 for RMSE, the vertical accuracy value for LiDAR ground points in the test area is around 0.5m. Furthermore, this study shows that the vertical accuracy in the common IDW-derived DEM with one metre by one metre pixel size is around a 0.444m, which is more accurate than the original LiDAR dataset with 0.5m vertical accuracy.

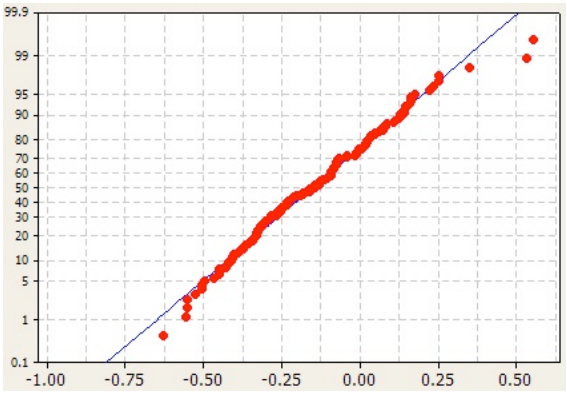


Figure 5: The sample of Q-Q plot to show error normal distribution

As mentioned earlier, PMs were used as a control for determining the accuracy for observed LiDAR elevations. However, it has been pointed out by Liu (2011) that determining the corresponding elevations for PMs from LiDAR data is critical for accuracy assessment. It is largely accepted that elevation is a geographic phenomenon that adheres to Tobler's first law of geography (Tobler,1970), that is, "Everything is related to everything else, but near things are more related than distant things." This is often referred to as "autocorrelation". However, a geographic distance, or "neighbourhood distance" from those control points, that can be used for estimating the accuracy of LiDAR observations, needs to be determined. Lag size is required to determine long-range autocorrelation based on an empirical semivariogram for a given dataset. However, proper lag size is also needed to determine the short-range autocorrelation.

Comparing the result of the ANN (Table 2), it has been shown that ANN can be used to determine an appropriate distance to identify autocorrelated neighbour LiDAR points. Table 1 shows that the RMSE value between 0m and 0.366m gives the lowest amount the RMSE. It can be concluded that the highest similarity between LiDAR point and PMs occurs within 0m-0.625m (average in 0.366m) around each PMs. Resultant outcomes from the ANN analysis (Table 2) shows that the lowest RMSE occurs between 0m and 0.531m distance from PMs due to the short-range autocorrelation, which is similarity between LiDAR points and PMs. Thus, instead of a comparison between LiDAR sample points and PMs in sequential minimum distances distance, ANN can be applied to show the short-range autocorrelation distance between LiDAR ground points and PMs.

The study area is influenced by anisotropy in long-range autocorrelation (Figure 6), the short-range autocorrelation within 0.531 meters has been assumed to be isotropic.

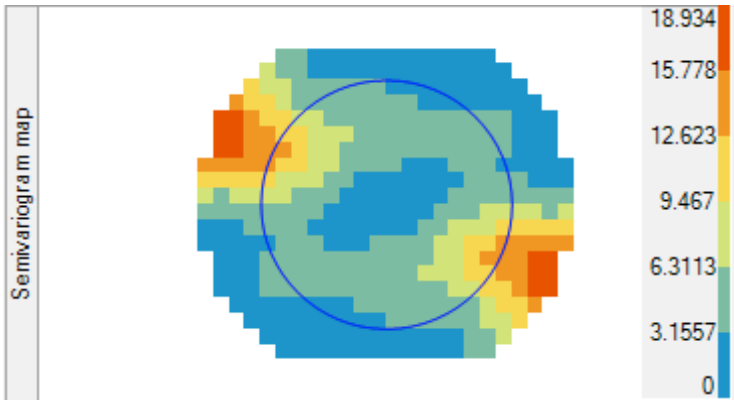


Figure 6: Anisotropy influences on derived empirical semivariogram using 500 random stratified derived sample points in a given LiDAR dataset.

Figure 6 shows the southwest-northeast directional effect, calculated from a sample of 500 points drawn as a stratified random sample of the LiDAR dataset. The noted direction (pathway) is because of the orientation of

the coastline and the change from flat areas to more rugged surface in the opposite direction, southeast-northwest.

Table 2: Statistical results of the ANN analysis for given LiDAR dataset

Ratio	*P*-value	Expected(m)
1.262	<0.0001	0.531

Table 2 shows the result of the ANN analysis. Short-range autocorrelation can be represented in the lag-size being equal to expected distance. A ratio score more than 1 shows that the distribution of sample points are not clustered, therefore the resultant vertical accuracy is valid for the LiDAR ground points which are not within 0.531m around PMs. Furthermore, P-value less than 0.0001 shows 99% confidence level in estimated vertical accuracy.

11 Conclusion

This paper has presented an alternative method of determining the accuracy of a LiDAR dataset. The approach was based upon a comparison between LiDAR ground points and PMs in sequential minimum distance. This approach enables LiDAR vertical accuracy assessment studies to avoid gridding influence on the final vertical assessment. The Table 1 shows that 0.366m from PMs results the lowest RMSE.

Furthermore, autocorrelation distance has been defined using an ANN analysis. The analysis shows that 0.531m around each LiDAR point is expected the similarity distance for each LiDAR point. The 0.531m is similar with the distance where the value of the RMSE changes from 0.241 to 0.257 in Table 1.

To determine the extent to which vertical accuracy is subject to interpolation methods, LiDAR DEMs derived from common IDW and geostatistic IDW were considered. Result shows that LiDAR DEM derived from geostatistic IDW presents elevation close to the real ground elevation in study area.

The vertical accuracy resulted from comparing between LiDAR ground point and PMs in autocorrelated distance is around 0.5m at 95% confidence level. Therefore, the LiDAR ground point dataset can be used for applications that do not need the vertical accuracy to be better than half a metre in the given study area.

Acknowledgement

The authors thank Iranian Ministry of Science, Research and Technology for support of this research project. We would like to acknowledge RMIT mathematical and geospatial sciences staffs' kind support. Our special thanks go to Mr. Paul Lennox, GIS team and all kind staff at Bass Coast Shire council for all their supports. We also would like to acknowledge anonymous reviewers for their comments and suggestions.

References

Aguilar, F. and Aguilar, M. (2007). Accuracy assessment of digital elevation models using a non-parametric approach. International Journal of Geographical Information Science **21**(6), 667-686.

Aguilar, F. J., Aguera, F., Aguilar, M. A. and Carvajal, F. (2005). Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. Photogrammetric engineering and remote sensing **71**(7), 805.

Aguilar, F. J. and Mills, J. P. (2008). Accuracy assessment of lidar-derived digital elevation models. Photogrammetric Record **23**(122), 148-169.

Aguilar, F. J., Mills, J. P., Delgado, J., Aguilar, M. A., Negreiros, J. G. and Pérez, J. L. (2010). Modelling vertical error in LiDAR-derived digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing **65**(1), 103-110.

Buxton, M. (2008).Planning Sustainable Futures for Melbourne's Peri-urban Region, RMIT University.

Carlisle, B. H. (2005). Modelling the spatial distribution of DEM error. Transactions in GIS **9**(4), 521-540.

DSE (2010).2008-9 Greater Melbourne Urban LiDAR Project. S. I. Infrastructure. The state of Victoria, DEPARTMENT OF SUSTAINABILITY & ENVIRONMENT.

Flood, M. (2004). American Society for Photogrammetry and Remote Sensing Guidelines–Vertical Accuracy Reporting for Lidar Data. ASPRS, Bethesda, Maryland.

Habib, A., Ghanma, M., Morgan, M. and Al-Ruzouq, R. (2005). Photogrammetric and LiDAR data registration using linear features. Photogrammetric Engineering and Remote Sensing **71**(6), 699-707.

Hodgson, M. E., Jensen, J. R., Schmidt, L., Schill, S. and Davis, B. (2003). An evaluation of LIDAR-and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. Remote Sensing of Environment **84**(2), 295-308.

Hodgson, M. E. and Bresnahan, P. (2004). Accuracy of airborne lidar-derived elevation: empirical assessment and error budget. Photogrammetric engineering and remote sensing **70**(3), 331-340.

Hodgson, M. E., Jensen, J., Raber, G., Tullis, J., Davis, B. A., Thompson, G. and Schuckman, K. (2005). An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions. Photogrammetric engineering and remote sensing **71**(7), 817.

Höhle, J. and Höhle, M. (2009). Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS Journal of Photogrammetry and Remote Sensing **64**(4), 398-406.

ICSM. (2008).ICSM Guidelines for Digital Elevation Data Version 1.0. Canberra,Australia, Inter-Governmental Committee on Surveying and Mapping (ICSM)**:** 49.

Lemmens, M. (2011).Airborne Lidar Geo-information. Netherlands, Springer **5:** 153-170.

Liu, X., Zhang, Z., Peterson, J. and Chandra, S. (2007). LiDAR-derived high quality ground control information and DEM for image orthorectification. GeoInformatica **11**(1), 37-53.

Liu, X., Zhang, Z. and Peterson, J. (2009).Evaluation of the performance of DEM interpolation algorithms for LiDAR data. Proceedings of the Surveying and Spatial Sciences Institute Biennial International Conference (SSC 2009), Surveying and Spatial Sciences Institute.

Liu, X. (2011). Accuracy assessment of LiDAR elevation data using survey marks. Survey Review **43**(319), 80-93.

Mallet, C. and Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing **64**(1), 1-16.

Maune, D., Black, T. and Constance, E. (2007). DEM user requirements. Digital elevation model technologies and applications: the DEM users manual, 2nd edn. American Society for Photogrammatry and Remote Sensing, Bethesda, 449-473.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography **46**, 234-240.

Webster, T. L. and Dias, G. (2006). An automated GIS procedure for comparing GPS and proximal LiDAR elevations. Computers & geosciences **32**(6), 713-726.

Wehr, A. and Lohr, U. (1999). Airborne laser scanning—an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing **54**(2), 68-82.

Zandbergen, P. A. (2008). Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy. Transactions in GIS **12**(1), 103-130.

EDMCAL: Processing EDM Calibrations in NSW

Volker Janssen
NSW Land and Property Information
Bathurst NSW 2795, Australia
Volker.Janssen@lpi.nsw.gov.au

Tony Watson
NSW Land and Property Information
Bathurst NSW 2795, Australia
Tony.Watson@lpi.nsw.gov.au

Abstract

The Surveyor General of New South Wales is a verifying authority under the National Measurement Act 1960 and responsible for ensuring that surveyors use verified measuring equipment. According to the Surveying and Spatial Information Regulation 2012, surveyors are required to verify their Electronic Distance Measurement (EDM) equipment in relation to an Australian standard of measurement of length at least once a year. For this purpose, Land and Property Information (LPI) provides and maintains several EDM baselines across the state. LPI is currently in the process of improving this infrastructure by upgrading existing baselines and building new baselines for the calibration of EDM instruments. This paper presents the current status of EDM baseline infrastructure in NSW and outlines the data processing performed by LPI in regards to EDM calibrations. The EDMCAL software currently employed by LPI is described and compared to a spreadsheet calculation generated by the University of New South Wales. Finally, LPI's new online EDM baseline booking system is introduced. This online system should now be used by surveyors to book access to all EDM baselines in NSW in order to allow efficient and effective use of existing and future baseline infrastructure.

Keywords: EDM calibration, EDMCAL, baseline infrastructure, online booking system, legal metrology.

1 Introduction

Legal metrology covers all measurements carried out for any legal purpose, including measurements that are subject to regulation by law or government decree. The National Measurement Act 1960 provides the legal basis for a national system of units and standards of measurement of physical quantities (Australian Government, 2014a). This Act is administered by the National Measurement Institute (NMI), which may in turn appoint organisations as verifying authorities under the provisions of Regulation 73 of the National Measurement Regulations 1999 (Australian Government, 2014b). As such, the office of the Surveyor General of New South Wales (NSW) has been appointed as a verifying authority for length measurement standards.

Practising surveyors in NSW are subject to the Surveying and Spatial Information Act 2002 (NSW Legislation, 2014a) and the Surveying and Spatial Information Regulation 2012 (NSW Legislation, 2014b). The latter states, among other things, that a surveyor must not use any Electronic Distance Measurement (EDM) equipment unless it is verified against the state primary standard of measurement of length by using pillared baselines, at least once every year and immediately after any service or repair. This instrument verification establishes traceability of its measurements to the national standard.

In this context, it is important to explain the difference between the terms verification and calibration. The verification of an EDM baseline is carried out periodically with precise EDM instrumentation carrying a current Regulation 13 certificate issued by NMI (the associated meteorological equipment is also calibrated against industry standards). This process determines the 'true' inter-pillar distances and establishes traceability because the EDM

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

baseline becomes a subsidiary standard of the International Metre. The calibration of an EDM instrument on a verified baseline determines the corrections that need to be applied to the instrument in order to obtain the 'true' inter-pillar distances, thereby establishing traceability of its measurements to the national standard.

In order to assist the surveying profession in meeting these requirements, the Surveyor General has established several EDM baselines throughout New South Wales. On behalf of the Surveyor General, Land and Property Information (LPI) is currently in the process of improving this infrastructure by upgrading existing baselines and building new baselines for the calibration of EDM instruments.

This paper briefly presents the current status of EDM baseline infrastructure in NSW and outlines the data processing performed by LPI in regards to baseline verifications and EDM calibrations. The EDMCAL software currently employed by LPI is described and compared to a spreadsheet calculation generated by the University of New South Wales. Finally, a new online EDM baseline booking system is introduced to allow efficient and effective use of the baseline infrastructure in NSW.

2 Current status of EDM baseline infrastructure in NSW

The Surveyor General has established several EDM baselines consisting of between four and seven concrete pillars throughout NSW. Current best practice has established that EDM baselines should consist of at least five (and preferably six or seven) pillars to increase the number of distances observed, thereby allowing a more reliable determination of the instrument correction. As a result, LPI is in the process of rationalising and improving its EDM baseline infrastructure by upgrading existing baselines to include more pillars and building new 7-pillar baselines.

Figure 1 illustrates the location of the 16 EDM baselines presently maintained in NSW. The new 7-pillar Seaham baseline (constructed in December 2013) replaces the 4-pillar baseline at Newcastle, which will cease to be maintained by LPI in August 2014 (but will continue to be used for teaching purposes at the University of Newcastle). It is planned to upgrade the 4-pillar Armidale baseline to include seven pillars, and contracts have been signed to establish a new 7-pillar baseline at Coffs Harbour (replacing the 4-pillar baseline at Grafton). In addition, efforts are underway to establish new 7-pillar baselines at Wollongong (replacing the existing 4-pillar baseline) and the South Coast (replacing the 4-pillar baselines at Nowra and Bega).

Figure 1: Location of current EDM baselines in New South Wales.

All EDM baselines in NSW follow the Heerbrugg design (also known as the Schwendener design), which features an almost equal distribution of the distances measured in all combinations over the baseline length as well as over the unit length of the EDM and permits the detection of all distance-dependent errors, including cyclic errors (e.g. Schwendener, 1972; Rüeger, 1996). For a detailed description of the substantial issues that need to be considered in the design and construction of a state-of-the-art EDM baseline, the reader is referred to Ellis et al. (2013). Depending on the location of the baseline, additional environmental aspects may have to be considered in some cases (Janssen, 2012).

LPI verifies these baselines on a 2-yearly basis and makes the current measurement reports available on the LPI website (LPI, 2014a). In accordance with the appointment as a verifying authority for length measurement standards, the least uncertainty quoted for the verified inter-pillar distances is currently 0.5 mm + 1.3 ppm at the 95% confidence level. The field procedures prescribed for EDM calibrations in NSW are documented in Surveyor General's Direction No. 5: Verification of Distance Measuring Equipment (LPI, 2009). It should be noted that the accurate observation of meteorological data is essential for a reliable EDM calibration. An error in the measurement of 1ºC in temperature or 3 millibars in atmospheric pressure will cause a corresponding error in the reduced distance of approximately 1 part per million (ppm).

3 The EDMCAL software

Naturally, the processing of EDM calibrations can be performed in different ways and with different tools. The standard mathematical methods involved have been described in various textbooks (e.g. Rüeger, 1996; Harvey, 2009). Most Australian jurisdictions have adopted an EDM calibration software package developed by Landgate in Western Australia. While this software, known as Baseline, is slowly progressing towards becoming the nationally preferred EDM calibration software, it will not be discussed in this paper.

In NSW, different software called EDMCAL has been used for many years by LPI for this purpose. Alternatively, several surveyors utilise a spreadsheet calculation developed at the University of New South Wales (UNSW). This section describes the EDMCAL software, and section 4 compares EDMCAL to the UNSW spreadsheet in regards to processing outcomes.

3.1 History

The program EDMCAL determines the additive constant (also known as instrument/reflector constant) and scale factor of EDM instruments using the parametric method of a rigorous least squares adjustment. This adjustment includes:

- Data snooping after Baarda (1968) to enable the detection of likely gross errors, i.e. a multidimensional test on the 'a posteriori / a priori' variance factors (test 1) and a one-dimensional test on the ratios of 'residual / a priori standard deviation of residual' (test 2).
- A one-dimensional similarity transformation in which the solution of pillar distances from the calibration adjustment is transformed to previously determined pillar distances.

The original program was written in the FORTRAN programming language by J.D. Love as part of an undergraduate student project for the Bachelor of Surveying degree at UNSW under the supervision of Dr J.M. Rüeger. The project report was submitted in April 1978. Since then, EDMCAL has undergone several modifications at LPI. These include:

- Ensuring compatibility with modern operating systems.
- Improving system performance, data structure and output format.
- Providing the possibility to input the EDM's modulation frequency, carrier wavelength and unit length as an alternative to the first velocity correction parameters.
- Using the mean of forward and reverse distance observations as the distance measurement between pillars if available.
- Output of individual corrections applied to the slope distances, i.e. instrument/reflector constant, atmospheric correction, slope correction, height (or datum) correction and chord-to-arc correction.
- Generation of a baseline database containing verification data on all baselines in NSW.
- Additional optional output to confirm the results, i.e. scale factor computation by linear regression and output of a HAVOC (Horizontal Adjustment by Variation Of Coordinates – see LPI, 2011) input file that can also be used to compute the scale factor.

The current version 5.1 of EDMCAL was created in December 2013. The mathematical and statistical procedures have been thoroughly tested over many years. During the most recent update, the mathematical algorithms used to determine the atmospheric corrections were improved to make them more readable and provide clear reference to their origin in the source code. Continuing EDMCAL usage and feedback from users may result in additional modifications to the statistical output in order to further improve the interpretation of results.

3.2 Operation

A sample EDMCAL input file is shown in Figure 2. It includes the following information:

- Label for the 'test number' of the program run.
- Baseline number and name.
- Observation date.
- Institution/company and operator name.
- Instrument make, model and serial number.
- Reflector make, model and serial number.
- Approximate additive constant.
- Instrument thermometer make, model, serial number and its correction.
- Reflector thermometer make, model, serial number and its correction.
- Instrument barometer and its correction (if barometers are used at the instrument and reflector, this entry should include information on both barometers and the mean barometer correction).
- Instrument standard deviation (generally set to 1 mm + 1 ppm for EDM calibrations) and 'f' factor for variance test (dependent on the number of distance observations and the number of pillars).
- First velocity correction parameters for the EDM (i.e. reference refractive index *VC1* and instrument pressure factor *VC2*) and partial water vapour pressure (default: 15 mb).
- Optional: Modulation frequency (Hz), carrier wave length (nm) and unit length (m) of the EDM.
- Optional: Linear regression option selection.
- Optional: HAVOC input file generation option selection.
- Forward observations including slope distance, height of instrument, height of target, mean atmospheric pressure and mean temperature.
- Reverse observations including slope distance, height of instrument, height of target, mean atmospheric pressure and mean temperature.

```
00 13DS1B13.cal                                  (File Name format: ##=year,AS=your initials,1=edmcal No.1,b##=Baseline code No.)
01 13 NOWRA                                      (Baseline code No. & Name)
02 13/09/13                                      (Date)
03 LPI                                           (Owner)
04 D.Sluys                                       (Operator)
05 TRIMBLE S3 DR                                 (Instrument Make & Model)
50 91210131                                      (Inst. Serial No.)
06 Leica 8812249                                 (Reflector Make, Model & Serial No.)
07 +0.000                                        (Approx. Additive Constant)
71 Thies S/No 17723                              (Thermometer used at instrument)
72 +0.0                                          (correction to thermometer)
73 Thies S/No 17723                              (Thermometer used at reflector)
74 -0.0                                          (correction to thermometer)
75 Negretti & Zambra S/No 932                    (Barometer used at instrument)
76 0.0                                           (correction to Barometer)
08 1.00   1.00                  1  2.200         (Instrument Std Deviation ; 2.200 ="f" factor for variance test)
09 278.3  80.7 15.0                              (Ist Velocity Parameters for inst.; "15" = Partial Water Vapour Press.)
11 1 0.241 2 0.242 150.6548    1 1016.3 13.9     (11=Fwd Obs, Pillar No Inst, HI, Pillar No Ref, HR, Slope Dist, Flag,
12 1 0.240 2 0.240 150.6542    1 1016.5 14.2         Press(mm or mb)(Inst&Ref Mean),Temp (Inst&Ref Mean))
11 1 0.241 3 0.241 317.2696    1 1016.5 13.8     (12=Rev Obs, Pillar No Ref, HI, Pillar No Inst, HR, Slope Dist, Flag,
12 1 0.240 3 0.241 317.2694    1 1016.7 14.6         Press(mm or mb)(Inst&Ref Mean),Temp (Inst&Ref Mean))
11 1 0.241 4 0.240 580.8092    1 1016.9 14.0
12 1 0.240 4 0.240 580.8084    1 1016.4 14.2     *** NOTE:   "HI"   MUST ALWAYS be in columns 6-10
11 2 0.240 3 0.241 166.6788    1 1017.0 13.7                 "HR"   MUST ALWAYS be in columns 14-18
12 2 0.240 3 0.242 166.6786    1 1017.2 14.4
11 2 0.240 4 0.240 430.3018    1 1016.9 14.4
12 2 0.240 4 0.242 430.3012    1 1016.9 14.8
11 3 0.240 4 0.240 263.6996    1 1017.0 14.8
12 3 0.240 4 0.241 263.6992    1 1017.0 15.0
99
```

Figure 2: Sample EDMCAL input file.

The input file is first checked for correctness of the number of pillars and number of observations. Several corrections are applied to the observed slope distances in order to reduce these to horizontal distances at the height of the lowest pillar: approximate additive constant (input by the user), atmospheric correction, slope correction, height (or datum) correction, and chord-to-arc correction. The standard deviations of the measured distances are then computed. This is followed by the formation and solution of the normal matrix derived from the observations and the output of the results of calibration and data snooping according to Baarda (1968). The 'null' hypothesis is tested for acceptance or rejection.

The reduced distances are used to form observation equations for a least squares adjustment (e.g. Harvey, 2009). The adjustment parameters are the distances from the first pillar to each of the other pillars and the correction to the additive constant, leading to the determination of the additive constant of the instrument/reflector pair used. In order to determine the scale factor, a one-dimensional similarity transformation is carried out using the calibrated distances as coordinates. The program output is terminated with a brief summary that also provides the differences

between the 'known' verified baseline distances and the adjusted distances determined with the EDM instrument under investigation (Figure 3).

```
                              SUMMARY
                              -------

     Test number                  :  13DS1B13.cal
     Baseline used                :  NOWRA
     Verification date            :  08 AUG 12
     Date of observation          :  13/09/13
     Name of owner                :  LPI
     Name of operator             :  D.Sluys
     Instrument manufacturer/model:  TRIMBLE S3 DR
     Instrument serial number     :  91210131
     Refl manufacturer,model & S/N:  Leica 8812249
     Approx. inst/refl constant   :   0.0000
     Inst thermometer S/N         :  Thies S/No 17723
     Refl thermometer S/N         :  Thies S/No 17723
     Barometer serial number      :  Negretti & Zambra S/No 932

  No. of pillars on base line:  4  No. of distances observed: 12

      Corrected instrument/reflector constant = -0.0351 metres

             Scale factor = 1.0000014891 (    1.5 ppm )

                  Calibrated Distance   Adjusted Distance  Difference

Pillars 1 to 2          150.4696            150.4695       -0.0001

Pillars 1 to 3          317.0879            317.0862       -0.0017

Pillars 1 to 4          580.7327            580.7321       -0.0006
```

Figure 3: Sample EDMCAL output summary.

3.3 EDMCAL mathematics

This section outlines the mathematics involved in the EDMCAL program. The equations stated are based on those found in Rüeger (1996). For a detailed description of least squares adjustments and statistics related to surveying applications, the reader is referred to Harvey (2009). As mentioned earlier, the following corrections are applied to the observed slope distances d_1 in order to reduce these to 'horizontal' distances d at the height of the lowest pillar:

$$d = d_1 + c_{AC} + c_{atm} + c_{slope} + c_{height} + c_{chord2arc} \tag{1}$$

where c_{AC} = approximate additive constant (input by user)
c_{atm} = atmospheric correction
c_{slope} = slope correction
c_{height} = height (or datum) correction
$c_{chord2arc}$ = chord-to-arc correction

The additive constant is valid for a particular combination of instrument and reflector only, accounting for the distance measurement reference points of the EDM instrument and the reflector not being coincident with the vertical axes at either end of the distance. In this case, the approximate additive constant stated in the input file (e.g. sourced from a previous EDM calibration) is utilised. If unknown or not specified, a zero entry is used. Obviously, the final additive constant will be determined through a least squares adjustment later on.

The basic principle of EDM instruments is the indirect determination of the travel time of a wave of light from the instrument to the reflector and back. While the speed of light in a vacuum is well known, in practice measurements are (of course) not carried out in a vacuum. The EDM measurements must therefore be corrected for the ambient atmospheric conditions because the velocity of visible and infrared waves changes with temperature, pressure and relative humidity (for light waves, the humidity is usually ignored). The atmospheric correction (also known as first velocity correction) c_{atm} is calculated according to:

$$c_{atm} = \left[VC1 - \frac{VC2\ p}{(273.15 + t)} + \frac{11.27\ PWVP}{(273.15 + t)} \right] 10^{-6}\ d_1 \tag{2}$$

where $PWVP$ = partial water vapour pressure (mb)
t = dry bulb temperature (°C)
p = atmospheric pressure (mb)
$VC1$ = reference refractive index (specific to EDM instrument)
$VC2$ = instrument pressure factor (specific to EDM instrument)
d_1 = observed slope (wave path arc) distance (m)

It should be noted that the second velocity correction (accounting for the fact that the light wave does not follow a circular path between the two pillars) is more important for microwaves than for light waves and ignored for EDM calibrations because it is insignificant over such distances. Also insignificant over such distances is the reduction from wave path arc distance to wave path chord distance (i.e. d_1 to d_2 in Figure 4).

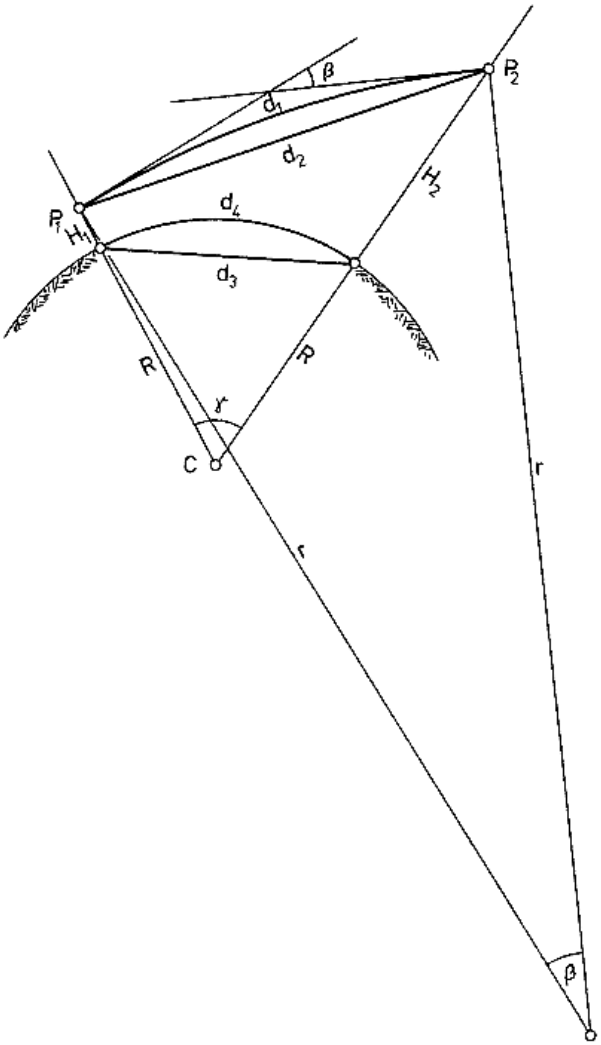


Figure 4: Reduction of wave path length d_1 to ellipsoidal arc length d_4 with the aid of the elevations (H_1 and H_2) of the terminals P_1 and P_2. The wave path chord and the ellipsoidal chord lengths are denoted by d_2 and d_3, respectively. The angle γ is subtended by the ellipsoid normal of radius R and through the terminals of the line. Similarly, the angle β is subtended by the normal to the wave path with radius r (Rüeger, 1996).

The slope correction reduces the slope wave path chord to a horizontal distance at the mean elevation of the two pillars involved:

$$c_{slope} = -\frac{\Delta H^2}{2\,d_2} - \frac{\Delta H^4}{8\,d_2^3} - \frac{\Delta H^6}{16\,d_2^5} \tag{3}$$

where ΔH = height difference between instrument and reflector (m)
d_2 = wave path chord distance (m)

The height (or datum) correction (also known as sea level correction) reduces the horizontal distance at the mean elevation of the two pillars to the chord distance referred to the datum plane through the lowest pillar elevation:

$$c_{height} = -\frac{H_M}{R}\,d_2 + \frac{H_M\,\Delta H^2}{2\,d_2\,R} + \frac{H_M\,\Delta H^4}{8\,d_2^3\,R} + \frac{H_M\,\Delta H^6}{16\,d_2^5\,R} \tag{4}$$

where H_M = mean height of instrument and reflector above the lowest pillar (m)
R = radius of curvature of the ellipsoid along the line (m), here assuming R = 6,370,100 m for New South Wales

The chord-to-arc correction converts the chord distance at the lowest pillar elevation d_3 to the datum arc distance d_4 (this correction is generally zero over distances used for EDM calibrations):

$$c_{chord2arc} = \frac{d_3^3}{24\,R^2} \tag{5}$$

The additive constant is then determined together with the distances between pillar 1 and the remaining pillars from the solution of the normal equations of a standard least squares adjustment. In order to determine the scale factor (and its ppm equivalent), a one-dimensional similarity transformation is carried out using the calibrated distances as coordinates.

4 Comparison of EDMCAL and UNSW spreadsheet

An alternative tool for the calculation of EDM instrument calibrations was developed by Dr B.R. Harvey at the University of New South Wales (UNSW) in form of an Excel spreadsheet (Harvey, 2014). Initially created for teaching purposes in 2006, the spreadsheet is now used by several surveyors for their EDM calibration calculations. The advantage of this spreadsheet is that all equations are visible to the user (rather than hidden in source code) and calculations can be customised for special cases if desired. However, this also means that users are at risk of inadvertently changing calculations. Currently, the spreadsheet allows for the determination of additive constant and scale factor (via a least squares adjustment solving for only these two parameters) on baselines consisting of between four and eight pillars. A sample calculation of the cyclic error is also included.

It is useful to investigate whether the UNSW spreadsheet and EDMCAL provide comparable results in regards to EDM calibrations. Several datasets observed by LPI legal metrology staff during EDM baseline verifications are used for this purpose. The comparison is mainly based on seven verification datasets for each of three EDM baselines: Wollongong (600 m, 4 pillars), Wagga Wagga (535 m, 5 pillars) and Dubbo (765 m, 6 pillars). In the absence of a history of observations on a particular baseline consisting of seven pillars, the comparison also incorporates four datasets recently collected on the 7-pillar baselines at Kingscliff (K, 721 m), Eglinton (E, 849 m) and Lethbridge Park (L, 984 m). The 25 datasets used in this study are summarised in Table 1. In order to provide maximum redundancy, all possible inter-pillar distances were observed. Most datasets were collected using a Leica TCA2003 total station, while a Leica TS30 total station was used for all datasets from 2011. Both instruments are similar in regards to the precision stated by the manufacturer (1.0 mm + 1 ppm and 0.6 mm + 1 ppm, respectively).

Table 1: EDM baseline verification datasets used in this study.

Dataset	4 pillars (Wollongong)	5 pillars (Wagga Wagga)	6 pillars (Dubbo)	7 pillars (various)
1	Jun 2000	Dec 2000	Jun 2000	Mar 2010 (K)
2	Jun 2002	Jun 2002	May 2002	Nov 2012 (E)
3	May 2004	May 2004	Jun 2003	Jul 2013 (K)
4	Nov 2005	May 2006	Oct 2004	Oct 2013 (L)
5	May 2007	Apr 2009	Sep 2006	–
6	May 2011	May 2011	Aug 2008	–
7	Jun 2013	Jun 2013	Aug 2012	–

In this context, it should be noted that the UNSW spreadsheet requires the user to input the relative humidity for each measured line in the calculation of the first velocity correction, while this quantity is ignored in EDMCAL. However, LPI staff routinely observe and record these values in the field during baseline verifications, allowing the comparison to be undertaken without the need for estimated or externally sourced values. It should also be noted that the UNSW spreadsheet requires meteorological data to be corrected before it is entered, while EDMCAL is able to accept the raw observations and apply thermometer and barometer instrument corrections during processing.

The results of the comparison between the additive constants and scale factors calculated using the UNSW spreadsheet (no iteration performed) and EDMCAL are illustrated in Figures 5 and 6, respectively. It should be noted that the first Dubbo dataset (June 2000) was identified as an outlier and removed from the analysis. The results indicate good agreement between the two calculation tools. The additive constants generally agree within 0.3 mm for the 4-pillar (Wollongong) and 5-pillar (Wagga Wagga) baselines, while larger differences of up to 0.8 mm are obtained for the 6-pillar baseline at Dubbo. Much better agreement of 0.1 mm or better is achieved at the 7-pillar

baselines, although it should be noted that three of these datasets represent the first verification after construction of the baseline and therefore use their own results as 'known' distances in the processing. The scale factors agree within about 0.5 ppm for the 4-pillar, 5-pillar and 7-pillar baselines and about 1.5 ppm for the 6-pillar baseline.

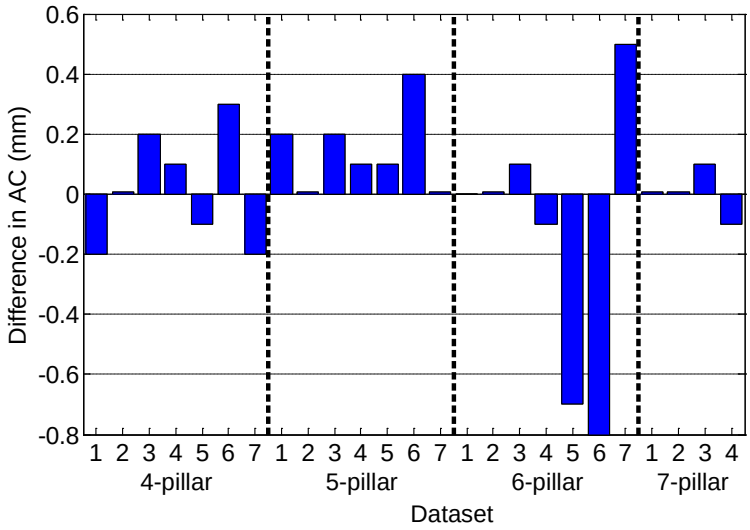


Figure 5: Difference in additive constant (AC) between UNSW spreadsheet and EDMCAL output (mm).

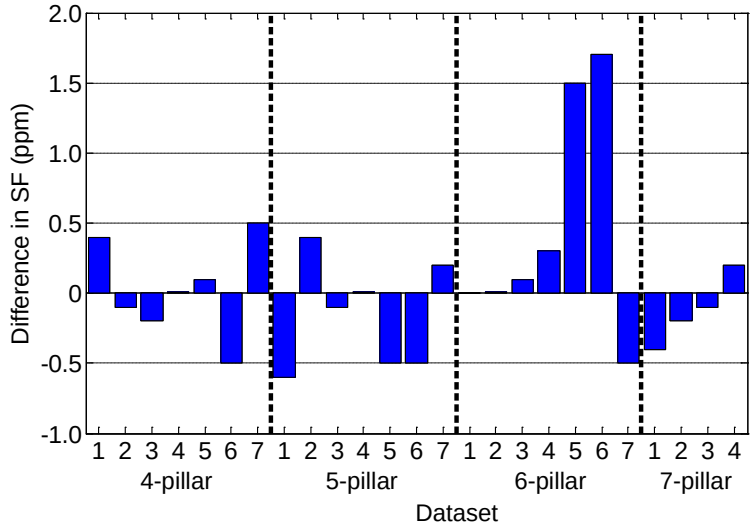


Figure 6: Difference in scale factor (SF) between UNSW spreadsheet and EDMCAL output (ppm).

It is also useful to investigate the differences in the distances relative to the first pillar after the calculated additive constant and scale factor have been applied. As already mentioned, the June 2000 dataset at Dubbo was identified as an outlier and therefore removed from the analysis. Descriptive statistics on the comparison between the output of the UNSW spreadsheet and EDMCAL are summarised in Table 2. A high level of agreement between the two calculation tools is evident, and the stated differences can be assumed negligible for most EDM calibrations in practice.

Table 2: Descriptive statistics of the differences in distances from pillar 1 between UNSW spreadsheet and EDMCAL (all values in mm).

	4 pillars	**5 pillars**	**6 pillars**	**7 pillars**
Min.	-0.2	-0.4	-0.7	-0.6
Max.	0.2	0.7	0.6	0.4
Range	0.4	1.1	1.3	1.0
Mean	0.01	0.03	-0.07	-0.04
RMS	0.08	0.20	0.33	0.19

These results show that the UNSW spreadsheet provides results comparable to EDMCAL processing for general practical purposes. However, it is important to note that recently verified distances must be used as 'known' inter-pillar distances in order to obtain reliable results. Using the latest verification results (i.e. 2012 or 2013) as 'known' distances for the processing of all datasets investigated in this study provided considerable differences not only in the resulting scale factors due to slight pillar movement but also negatively affected the relative comparison between the two calculation tools.

5 Online EDM baseline booking system

In the past, EDM baselines in NSW were not subject to a booking requirement. Surveyors were generally able to visit a baseline at any time, provided no baseline specific access requirements were in place (e.g. prior approval and/or keys from baseline host required). Particularly at popular baselines, this can result in several surveyors attempting to use the baseline at the same time, thus negatively affecting the surveyors' productivity and time management (particularly if considerable travel time to the baseline is involved).

In order to avoid these disadvantages, LPI has developed the EDM Baseline Booking System. This free online booking system is now available via the LPI website (LPI, 2014b) and allows registered users to reserve a particular time slot at the desired baseline in advance. The booking process is simple and straightforward, comparable to booking a hotel room online. A help page with instructions on how to use the system and the opportunity to make enquiries or provide feedback are also included.

A screenshot of the booking system's main page is shown in Figure 7. The process consists of the following three simple steps:

1. Select a booking date.
2. Select an EDM baseline.
3. Select an available booking time.

Once the booking is finalised, a confirmation will be sent by email, also outlining the general and baseline specific conditions of use that had to be accepted during the booking process. The user is required to carry a printout of this booking confirmation with them at all times when on the baseline site. This will provide proof of approved access to the baseline for the specified time period.

The EDM Baseline Booking System was launched on 29 October 2013 and should now be used by surveyors to book access to all EDM baselines in NSW in order to allow efficient and effective use of existing and future baseline infrastructure. The booking system will also assist LPI in monitoring the frequency of use of each baseline, thereby allowing more informed decision making in regards to the state's EDM baseline infrastructure in the future.

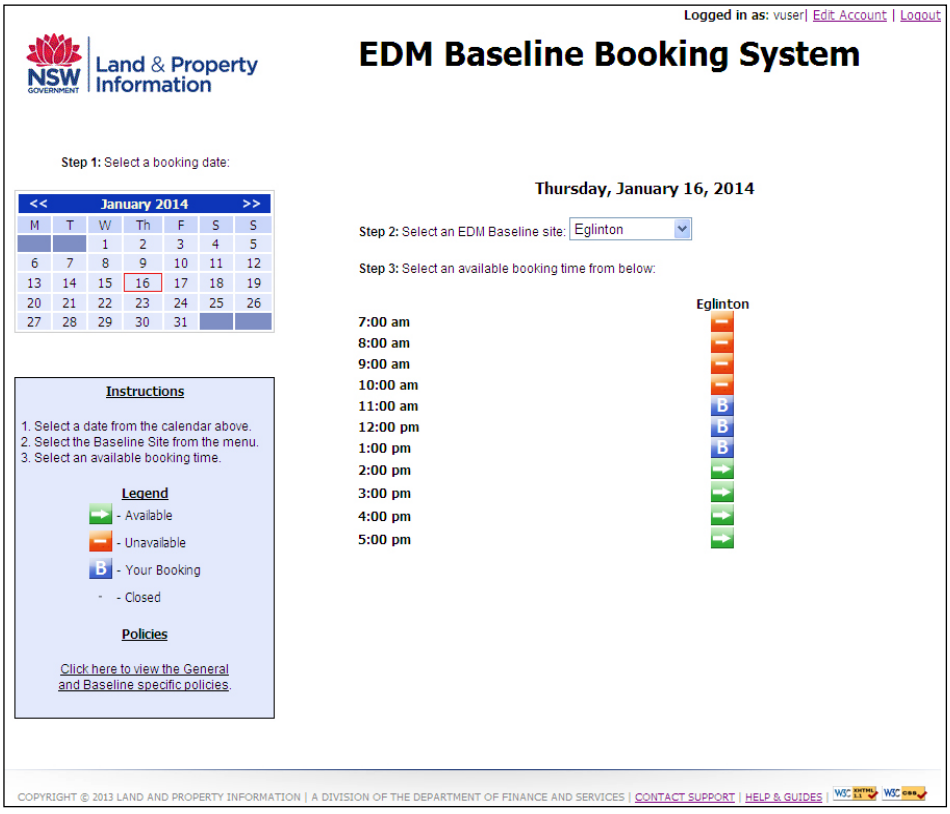


Figure 7: Main page of the online EDM Baseline Booking System (http://lpi.nsw.gov.au/edmbooking).

6 Conclusion

This paper has briefly described the status of EDM baseline infrastructure in NSW, which is currently being rationalised and improved by upgrading existing baselines to include more pillars and building new 7-pillar baselines. The EDMCAL software used by LPI for the processing of EDM baseline verifications and EDM calibrations was outlined in detail. Based on 25 datasets collected on baselines consisting of between four and seven pillars, it was shown that the EDM calibration spreadsheet developed by the University of New South Wales provides additive constants and scale factors comparable to the EDMCAL output.

Further research could investigate using slope distances in a 3D solution instead of distances reduced to the horizontal at the height of the lowest pillar. It would also be useful to examine the use of verified distances and their uncertainties in a Bayesian solution rather than holding these inter-pillar distances fixed. This is expected to yield better statistics but may not cause any significant change in the parameters.

It is important to note that calibrating an EDM instrument in prism mode does not calibrate the reflectorless EDM laser. These two modes generally have different additive constants and scale factors within any one instrument, i.e. testing in reflectorless mode must be performed separately. It should also be noted that if surveyors measure any distances that are longer than the longest line on their EDM calibration baseline, they should consider the reliability of the extrapolation of their calibration parameters.

Finally, LPI's new EDM baseline booking system was introduced. This free online system should now be used by surveyors to book access to all EDM baselines in NSW in order to allow efficient and effective use of existing and future baseline infrastructure. By allowing LPI to monitor the frequency of use of each baseline, the booking system will also assist LPI in making more informed decisions regarding the state's EDM baseline infrastructure.

Acknowledgements

LPI's ICT Technical Services in general and Sam Osborne in particular are gratefully acknowledged for their assistance in the design of the online booking system.

References

Australian Government (2014a). National Management Act 1960 (Commonwealth), available at http://www.comlaw.gov.au/Details/C2013C00612 (accessed Feb 2014).

Australian Government (2014b). National Measurement Regulations 1999 (Commonwealth), available at http://www.comlaw.gov.au/Details/F2013C00709 (accessed Feb 2014).

Baarda, W. (1968). *A testing procedure for use in geodetic networks*. Publications on Geodesy, 2(5), Netherlands Geodetic Commission, Delft, 97pp.

Ellis, D., Janssen, V. and Lock, R. (2013). Improving survey infrastructure in NSW: Construction of the Eglinton EDM baseline. *Proceedings of Association of Public Authority Surveyors Conference (APAS2013)*, Canberra, Australia, 12-14 March, 187-201.

Harvey, B.R. (2009). *Practical least squares and statistics for surveyors*. Monograph 13, 3rd edition, School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia, 332pp.

Harvey, B.R. (2014). EDM calibration spreadsheet, available at http://www.surveying.unsw.edu.au/ls/ (accessed Feb 2014).

Janssen, V. (2012). Indirect tracking of drop bears using GNSS technology. *Australian Geographer*, 43(4), 445-452.

LPI (2009). Surveyor General's Direction No. 5: Verification of distance measuring equipment, available at http://www.lpi.nsw.gov.au/surveying/publications/surveyor_generals_directions (accessed Feb 2014).

LPI (2011). HAVOC, available at http://www.lpi.nsw.gov.au/surveying/geodesy/havoc (accessed Feb 2014).

LPI (2014a). EDM baseline certificates, available at http://www.lpi.nsw.gov.au/surveying/surveying_services/edm_baseline_certificates (accessed Feb 2014).

LPI (2014b). EDM baseline booking system, available at http://lpi.nsw.gov.au/edmbooking (accessed Feb 2014).

NSW Legislation (2014a). Surveying and Spatial Information Act 2002, available at http://www.legislation.nsw.gov.au/viewtop/inforce/act+83+2002+cd+0+N (accessed Feb 2014).

NSW Legislation (2014b). Surveying and Spatial Information Regulation 2012, available at http://www.legislation.nsw.gov.au/viewtop/inforce/subordleg+436+2012+cd+0+N (accessed Feb 2014).

Rüeger, J.M.R. (1996). *Electronic distance measurement: An introduction*. 4th edition, Springer, Berlin, 300pp.

Schwendener, H.R. (1972). Electronic distancers for short ranges: Accuracy and checking procedures. *Survey Review*, 21(164), 273-281.

The new digital orthometric elevation model of Kilimanjaro

Pascal Sirguey
National School of Surveying
University of Otago
Dunedin, New Zealand
pascal.sirguey@otago.ac.nz

Nicolas J. Cullen
Geography Department
University of Otago
Dunedin, New Zealand
nicolas.cullen@otago.ac.nz

Jorge Filipe Dos Santos
Vale Technological Institute
Brazil
jorge.filipe@itv.org

Abstract

Kibo, the highest of three peaks of Kilimanjaro, has not benefited from a medium to large scape topographic mapping in about 50 years. The rapidly changing topography associated with the glacier retreat and the fact that the slopes of Kibo attract about 40,000 climbers each year thus justify the need to develop a new topographic survey of this outstanding landmark, designated a UNESCO World Heritage Site in 1987. In this context, the application of the photogrammetric principles to the latest generation of very high resolution space-borne optical sensors (VHRS) offers new surveying opportunities by enabling the topographic mapping of remote and hardly accessible areas at large scale with unprecedented spatial resolution. This paper illustrates the potential of a space-borne photogrammetric survey technique by reporting on the last effort to map the topography of Kibo from GeoEye-1 stereo imagery, which has led to the creation of a new 50cm resolution Digital Elevation Model (DEM), namely KILISoSDEM2012. Furthermore, this new model is combined with the refined local geoid model KILI2008 generated from gravimetric observations captured by the international team that completed the last survey of the orthometric height of Kibo in October 2008. This paper shows that the new digital orthometric elevation model exhibits a 35% and 25% improvement in planimetric and elevation accuracy, respectively, compared to the specifications of GeoEye-1 Precision products.

1 Introduction

In 1912, German explorer Eduard Oehler and glaciologist Fritz Klute completed the first topographic survey of Kibo, the highest of three peaks of Kilimanjaro, using the emerging photogrammetric technique (Klute, 1920, 1921). This led to a 1:50,000 scale map being produced, the quality of which should be praised given the complexity of the terrain and the technical limitations of the emerging surveying technique at the time.

The mapping of Kilimanjaro at a scale of 1:50,000 was not repeated for another 50 years. A photogrammetric survey was conducted in January 1962 for the main mountain, following a survey in March 1958 for its surroundings (Directorate of Overseas Surveys, 1964; Young and Hastenrath, 1987; Shirima, 2013). Although more recent

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

research projects have used some limited survey data in selected areas of the volcano, such as to characterise the demise of glaciers on Kibo (Cullen et al., 2013), the massive volcano has not benefited from an updated and more detailed survey in 50 years. The rapidly changing topography associated with the glacier retreat and the fact that the slopes of Kibo attract about 40,000 climbers each year (International Mountaineering and Climbing Federation, UIAA, 2013) justify the need to develop a new topographic survey of this outstanding landmark, designated a UNESCO World Heritage Site in 1987.

In this context, the application of the photogrammetric principles to the latest generation of very high resolution space-borne optical sensors (VHRS) offers new surveying opportunities by enabling the topographic mapping of remote and hardly accessible areas at large scale with unprecedented spatial resolution. Recent hardware and software advances now allow dense point clouds to be generated, thus making the use of VHRS stereo imagery a viable technique to complete a large topographic survey at a small pecuniary and logistical cost. Thus, 100 years after Klute and Oehler completed the first ground based photogrammetric survey of Kibo, and 50 years after the most recent aerial photogrammetric survey, this paper illustrates the potential of a spaceborne photogrammetric survey technique by reporting on the last effort to map the topography of Kibo from GeoEye-1 stereo imagery. This has led to the creation of a new 50cm resolution Digital Elevation Model (DEM), namely KILISoSDEM2012. Furthermore, orthometric heights are obtained by combining this new DEM with the refined local geoid model KILI2008 generated from gravimetric observations captured by the international team that completed the last survey of the orthometric height of Kibo in October 2008 (KILI2008, 2009).

2 Data and methods

2.1 Satellite imagery

The GeoEye-1 sensor belongs to the latest generation of very high spatial resolution optical sensors on the civil market. It was launched on 6 September 2008 by GeoEye Inc (now merged with Digital Globe) and supports the capture of imagery at 1.65 m in four multispectral bands (MSI, visible and near infrared) and 0.41 m in the panchromatic band (PAN) although data is sold at 2 m and 50 cm resolution due to US government regulation. A GeoStereo product was ordered over an area of about 100 km2 centred on Reusch Crater (see Figure 1). Because of the persistent cloud cover on Kibo, the minimum cloud-free requirement could not be met despite the multiple acquisition attempts. This led to the acquisition and delivery of five bundle multispectral (MSI) and panchromatic (PAN) stereo pairs that, when considered together, provided almost a cloud-free coverage of the entire area (Table 1). Only two areas remained obscured in all pairs, namely north-west of the Great West Breach (a.k.a. Western Breach) and south-west of the Breach Wall (see Figure 1). In order to provide terrain elevation data for those gaps, a 15 m resolution Level 1A stereo image of the area, which had been acquired on 19 August 2004, was obtained from the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER).

Table 1: Specifications of ASTER and GeoEye-1 imagery.

Date of each stereo pair (GMT)	Pixel size [m]	Cloud [%]	View Azim.	View Elev.
19 Aug 2004, 7:54	15.0	0	270	81.4
07:55	15.0	0	186.5	32.9
20 Sep 2012, 7:58	0.50	25	226.6	74.3
07:57	0.50	30	354.5	61.6
9 Oct 2012, 07:49	0.50	65	33.8	61.3
07:50	0.50	52	147.1	74.1
20 Oct 2012, 07:50	0.50	59	33.8	64.4
07:51	0.50	54	155.8	73.3
23 Oct 2012, 07:59	0.50	68	345.7	61.6
08:00	0.50	57	258.2	75.5
24 Jan 2013, 07:48	0.50	22	68.7	71.8
07:49	0.50	23	156.7	64.4

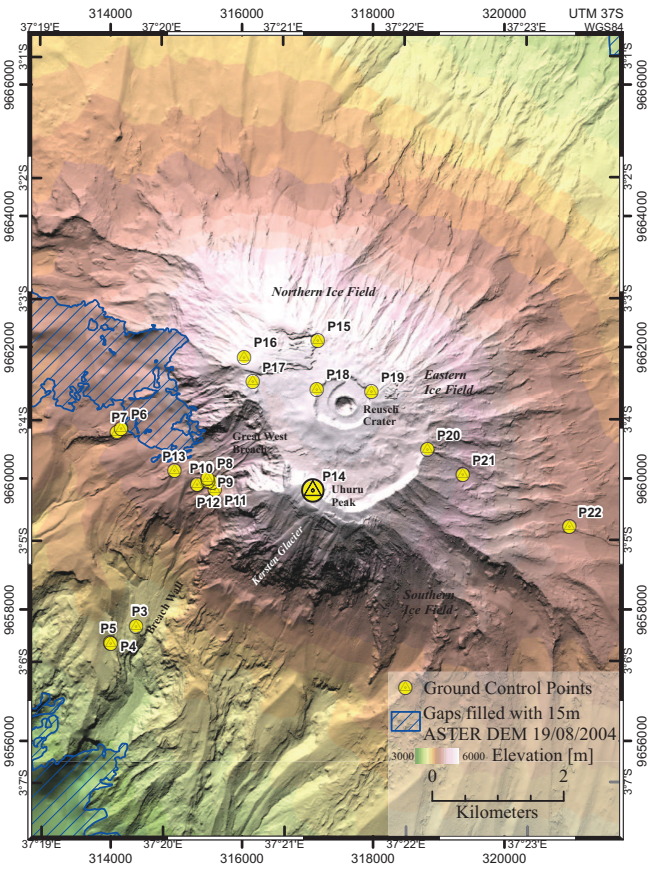


Figure 1: Colourized shaded relief of KILISoSDEM2012. Note the position of the 20 Ground Control Points used in the triangulation, as well as the areas with remaining obscuration by clouds in the GeoEye images that were filled with 15m resolution elevation points derived from an ASTER stereo image on 19/08/2004.

2.2 Ground control

Twenty Ground Control Points (GCPs) were collected on Kibo over 20-26 September 2012, thus matching or close to the capture dates. Ground features such as huts, a helicopter pad, prominent rocks, and well identified ice boundaries and ground discoloration marks were used. The WGS84 coordinates of these GCPs were measured using a Leica GS20 GPS L1 Code receiver in static mode with an occupation time of 2 minutes. The GPS data were post-processed via differential correction using the MAL2 IGS permanent GPS site located at Malindi, Kenya about 270 km away (great circle distance). Despite the relatively long baseline and short occupation time, post-processing yielded mean position quality measures RMS_{xy}=0.29 m and RMS_{z}=0.48 m. This was considered of suitable accuracy to support the triangulation of the 50 cm PAN stereo pairs.

2.3 Image block and triangulation

The triangulation of the satellite image stereo-pairs was conducted in ERDAS LPS 2013. A single image block based on the GeoEye-1 Rational Polynomial Coefficient (RPC) model was formed with the 10 panchromatic images corresponding to the five stereo-pairs. Processing the multi-date acquisition of GeoEye-1 images in a single block permitted a triangulation of all images together via a single bundle adjustment that enabled multiray photogrammetric processing. This allowed every pixel in numerous image overlaps to be processed, thus yielding redundancies that could be used to increase the accuracy of the point cloud via statistical filtering, or to increase the point density (Leberl et al., 2010).

About 300 tie points (TPs) were collected in a semi-automatic manner. Most TPs were automatically measured in LPS in multiple images and checked subsequently. The de-correlation between image pairs from different dates yielded a substantial amount of wrong points and points not being identified in all images. Those were re-

measured manually when wrong, as well as transferred to images where they were not automatically identified. All twenty GCPs were converted to UTM37S cartographic projection, measured in all images where they appeared, and used as full control. Because of the uncertainties associated with the local geoid and the Tanzanian Vertical Datum (TVD) (see Saburi et al., 2000; KILI2008, 2009), heights above the reference WGS84 ellipsoid (HAE) were initially used. Any customized adjustments to other vertical datums could thus be easily processed subsequently to the triangulation and Digital Elevation Model production.

The ASTER stereo image was triangulated separately using the ASTER orbital pushbroom sensor model, 28 TPs, and 20 GCPs. Fifteen were GCPs collected during fieldwork that could be identified in the 15m ASTER images. Advantage was taken of the triangulated GeoEye image block to support the collection of five additional control points well distributed around Kibo with a sub-metre accuracy comparable to that of the GPS points.

2.4 DEM generation

The dense point cloud (PC) was generated with the enhanced Automatic Terrain Extraction (eATE) of LPS 2013 in a pseudo multiray approach. First, each of the four triangulated stereo pairs acquired in 2012 was considered separately to support most of the PC. The imagery acquired on 24 January 2013 was initially disregarded because of the substantially later and transient snow on ice surfaces. All 45 overlap combinations from all 10 images were then considered in order to generate more points in the non-glaciated stable areas with lower point density, such as those affected by repeated cloud obscuration or steep relief. The normalized cross correlation feature matching method was used with a relatively low threshold (0.65) and low contrast settings to generate numerous 3D points at the expense of a relatively large number of blunders (about 5%). The raw PC was generated in about one week with three parallel jobs on an Intel©i7-2600K equipped with 16 GB RAM; this yielded about 270 million points (MPts), with substantial redundancy in some areas.

The PC was thinned via median filtering within 50 cm grid cells, thus providing a first level of blunder removal. Further cleaning was achieved using a statistical outlier removal from the Point Cloud Library (PCL) (Rusu and Cousins, 2011), while remaining blunders were deleted via manual editing. Similarly, a PC was generated from the ASTER stereo pair from which about 13,000 points were used to fill gaps in the GeoEye PC. The cleaned gap-filled PC accounted for about 181 MPts at typically 50 cm spacing, meaning that more than 40% of the final 50 cm resolution raster DEM (437 Mcells) was supported by measured 3D points. Finally, the meshed PC was smoothed using a Laplacian operator and the resulting PC was interpolated to a 50cm resolution raster DEM using the ANUDEM thin plate smoothing spline terrain interpolator in ArcGIS 10 (Hutchinson, 1989).

2.5 Local Geoid model KILI2008

2.5.1 Data acquisition

In October 2008 an international project called KILI2008, involving 19 researchers from institutions of six different countries, aimed at accurately measuring the orthometric height of Mount Kilimanjaro (KILI2008, 2009). The measurements involved the combined use of GNSS double frequency receivers (Trimble R8) and two gravimeters (Scintrex CG3-M and Scintrex CG5) illustrated in Figure 2 (a) and (b), respectively. The gravimetric observations were necessary to estimate a local geoid with sufficient accuracy to convert the ellipsoidal heights from GNSS observations to orthometric heights.

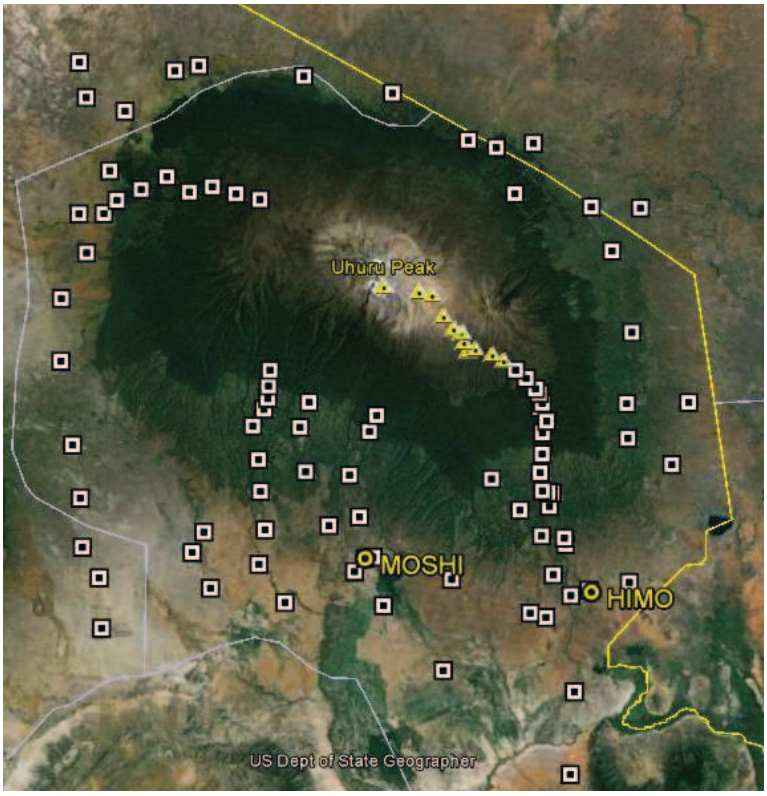
Ninety nine gravimetric and GNSS data were acquired over 10 days at locations approximately distributed in a grid centered on Kibo, depending on logistic constraints and existing routes (Figure 2). GNSS static observations of 30 to 45 minutes duration were collected together with the gravimetric measurements. Two GNSS reference stations located in Moshi and Himo were used for coordinate correction. Seven additional GPS points were acquired along the Manrangu route between 3200 meters height and the summit (Uhuru Peak).

2.5.2 Data processing

The KILI2008 geoid was calculated based on the gravimetric measurements using the Standard Residual Terrain Modeling strategy (Forsberg and Tscherning, 1981). The long wavelengths of the geoid undulation were determined using degree 0-360 of the Earth Gravitational Model 2008 (EGM2008) (Pavlis et al., 2012). The effect associated with the mass of the mountain was computed with Least Squares Collocation (see KILI2008, 2009, for more details) and using topographic information obtained from Shuttle Radar Topography Mission (SRTM) Digital Terrain Elevation Data (DTED) Level 1 (Farr et al., 2007). The calculated KILI 2008 geoid surface is

(a)

(b)



(c)

Figure 2: Instruments used during the KILI2008 project, namely (a) GNSS receiver Trimble R8 and (b) Scintrex CG3-M and Scintrex CG5 gravimeters. White squares in (c) indicate locations where both GNSS and gravimetric observations were collected, while only GNSS observations are identified as yellow triangles. GNSS reference stations are represented by yellow circles.

illustrated in Figure 3. Differences between KILI2008 and EGM2008 are -12 cm in Moshi, -12 cm in Himo and +21 cm at Uhuru Peak.

The GNSS data collected at the two reference stations were processed using the GIPSY-OASIS II software package in order to compute positions up to sub-centimeter level with respect to the ITRF2005 reference frame. Those two stations were later used as reference to compute the coordinates of the other points using Trimble Business Center.

The final orthometric heights obtained from the combination of GNSS survey and the KILI2008 geoid are referred to as global mean sea level elevation. A reference was needed to convert those heights to the national vertical datum of Tanzania. There was only one existing benchmark in the region with known height in that local datum situated close to Moshi. Using GNSS observations on that point it was possible to detect a 1.28 m offset between the global and the local vertical datum.

3 Results

3.1 Accuracy assessment

The triangulation results are shown in Table 2. Given the relatively small number of GCPs, the accuracy was independently assessed using a leave-one-out cross validation protocol (LOOCV), whereby each GCP was used as an independent check point in turn and the block re-triangulated. The residuals associated with each GCP were collected to quantify the quality of the triangulated block. The relative accuracy of the ASTER triangulation can be explained by the accurate collection of GCPs at a sub-pixel level being supported by the interpretation from the very high resolution GeoEye-1 triangulated image block. The consistency between the dependent and the independent LOOCV residuals further demonstrates the robustness of both triangulations.

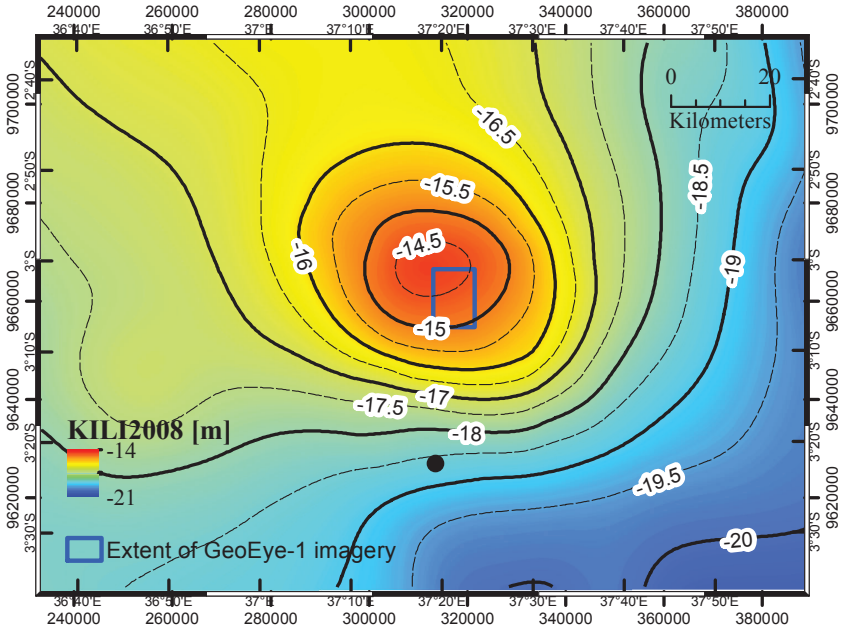


Figure 3: The KILI2008 geoid surface. The benchmark of the Tanzanian vertical datum is indicated by the black dot.

Table 2: Results of the image triangulation.

Image blocks	Pixel size [m]	Image RMSE [px]	Residuals of the control points [m]		
			RMS_x	RMS_y	RMS_z
GeoEye-1	0.50	0.27	0.42	0.61	1.09
Leave-one-out cross validation:			0.45	0.67	1.19
ASTER	15.0	0.21	1.92	1.89	4.40
Leave-one-out cross validation:			2.59	2.25	6.24

The propagation of Gaussian errors between the LOOCV residuals (Table 2) and the uncertainty of the GPS survey (Section 2.2) supports the accuracy specification of the final DEM product shown in Table 3. The latter exhibits a 35% and 25% improvement in planimetric and elevation accuracy, respectively, compared to the specifications of GeoEye-1 Precision products.

In Table 3, $RMSE = \sqrt{RMS_x + RMS_y}$ denotes the root mean square planimetric error. $CE90 = 1.5175 \times RMSE$ (Circular Error of 90%) is commonly used for quoting and validating geodetic image registration accuracy. A CE90 value is the minimum diameter of the horizontal circle that can be centred on all photo-identifiable Ground Control Points (GCPs) and also contain 90% of their respective twin counterparts acquired in an independent geodetic survey (FGDC, 1998, pg. 3-21). A Linear Error of 90% (LE90) is commonly used for quoting and validating DEMs. $LE90 = 1.6449 \times RMS_z$ (Linear Error of 90%) and represents the linear vertical distance that 90% of control points and their respective twin matching counterparts acquired in an independent geodetic survey should be found from each other (FGDC, 1998, pg. 3-21). $NMAS$ is the approximate map scale equivalencies based on the United States National Map Accuracy Standard and is defined as $1/NMAS = 1181 \times CE90$ (FGDC, 1998, pg. 3-21).

Table 3: Accuracy of the final DEM product.

RMSE [m	CE90 [m]	LE90 [m]	NMAS
0.86	1.31	2.12	1:1,600

3.2 Vertical datum

The photogrammetric block and subsequent DEM were produced initially in terms of height above the WGS84 ellipsoid: KILISoSDEM2012$_{\text{HAEWGS84}}$. In order to obtain the orthometric DEM, the geoid undulation KILI2008 was resampled to 50 cm using cubic convolution interpolation. KILISoSDEM2012$_{\text{KILI2008}}$ was then obtained by subtraction between the DEM in ellipsoidal height and the geoid separation as follows:

$$KILISoSDEM2012_{KILI2008} = KILISoSDEM2012_{HAEWGS84} - KILI2008. \quad (1)$$

The new KILISoSDEM2012$_{\text{KILI2008}}$ is illustrated in Figure 1, while Figure 4 shows the current topography of Kersten glacier. Figure 5 illustrates the level of detail of the new DEM by comparing 3D scenes with corresponding photographs of the Northern and Eastern Ice Field.

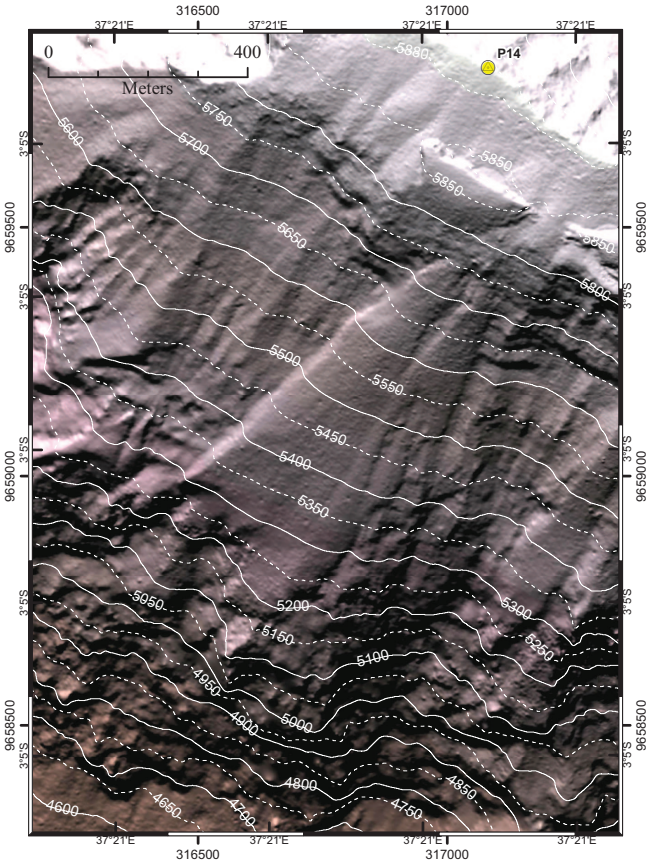
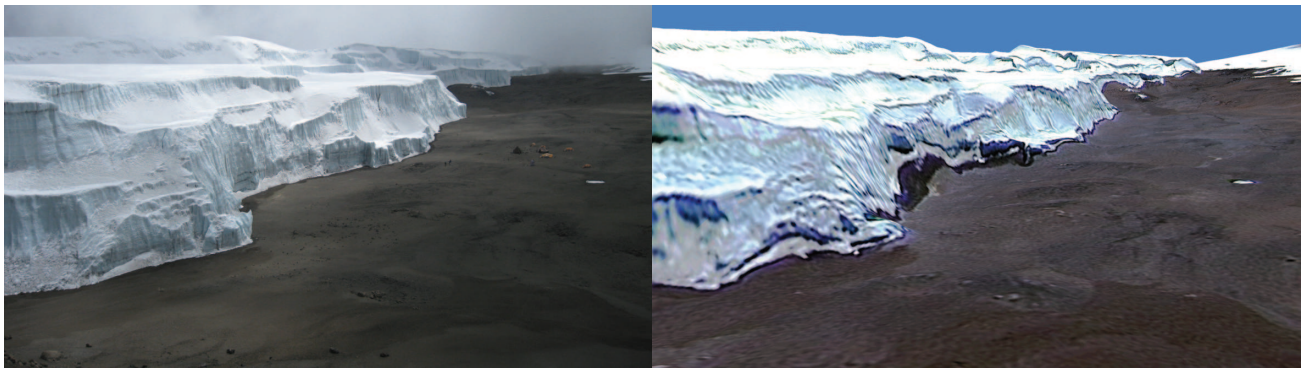


Figure 4: Colour shaded relief of KILISoSDEM2012 illustrating Kersten glacier. Contour lines indicate elevation in metres according to the Tanzanian Vertical Datum. Note the position of Uhuru peak at point P14.

3.3 About the height of Uhuru Peak

This study would not be complete without a note about the height of the highest point of Africa. The height of Uhuru Peak is most commonly known to be 5895 m amsl since the British Ordnance Survey in 1952 (Pugh, 1954). Before that, Klute (1921, ph. 149) reported on earlier estimates of what was named Kaiser Wilhelm Peak by



(a)

(b)

Figure 5: Left: Photo of (a) the Northern ice field seen from the top of the Western Breach approximately from P17 (Figure 2) eastward; (b) the Eastern ice field approximately from P19 location eastward (credit: Nicolas Cullen, 25 September 2012). Right: 3D visualisation of the KILISoSDEM2012 DEM at both corresponding locations: (a) GeoEye true-colour orthoimage of 20 September 2012 draped on the DEM; (b) shaded relief of the DEM.

Hans Meyer in 1889 (Meyer, 1891, pg. 154). Meyer estimated the height at 6010m using an aneroid barometer (Meyer, 1891, pg. 375-378). Klute (1921) also refers to the Anglo-German boundary expedition (1904-1906) which estimated Kibo to reach a height between 5888 and 5892 m although it is suspected that surveyors could not have seen the summit from their trigonometric point. Klute (1921) finally provided an estimated height of 5930 m from his photogrammetric survey Figure 1), although this figure is tarnished by the fact that the altitudes of the photogrammetric stations were determined on the basis of uncertain barometric measurements, rather than trigonometrically (Gillman, 1923).

In 1999, an accurate GPS survey was conducted at Uhuru Peak which led to a measured an ellipsoidal height of 5875.50 m (Saburi et al., 2000). This corresponded to the orthometric height which was estimated to be 5891.77 m based on the EGM96 geoid model that indicated a separation of -16.27 m at this location. Finally, a mean shift of 0.59 m was found with the Tanzanian height datum, thus yielding the final estimate of 5892.37 m. In 2008, the KILI2008 team (KILI2008, 2009) revised the geoid separation to be -14.48 m at Uhuru Peak. Given the ellipsoidal height of 5875.43 m (solution from the GIPSY software), the orthometric height above the KILI2008 geoid was found to be 5889.91 m ± 0.25 m. The additional departure from the TVD of 1.28 m at the Moshi benchmark yielded the most recent and assumedly most accurate estimate of the height of Uhuru Peak to be 5891.19 m 0.25 m.

A GPS point was collected at Uhuru Peak for this study yielding a relatively inaccurate ellipsoidal height of 5872.90 m. However, the triangulation of the GeoEye image block allowed this error to be partially mitigated as the precise targeting of the Kilimanjaro summit yielded a height of 5874.4 m, just one metre less than Team KILI2008, and within the specifications of the final product (see Section 3.1). However, the final HAE from the smoothed, interpolated KILISoSDEM2012 at Uhuru Peak (317082.1E; 9659820.4N) is found to be 5873.6 m ± 2.1 m, while the orthometric height above KILI2008 is 5888.1 ± 2.1 m assuming a 2σ error of 0.30 m in the KILI2008 undulation (KILI2008, 2009). KILISoSDEM2012 exhibits a relatively higher point at 317078.7E; 9659817.7N, HAE = 5874.2m ± 2.1 m, Orthometric height = 5888.7m±2.1m which is however not significantly higher given the uncertainty. The lower value can likely be attributed to the smoothing and interpolation process given the proximity of Uhuru peak to the edge of the crater rim.

4 Conclusion

This study documents KILISoSDEM2012$_{\text{KILI2008}}$, the new 50 cm resolution orthometric DEM of Kibo, the highest peak of Kilimanjaro. It is derived from multiray photogrammetry applied to five GeoEye-1 stereo pairs combined with the refined geoid model KILI2008. Triangulation results and independent accuracy assessment based on a leave-one-out cross validation protocol show that the KILISoSDEM2012 meets an accuracy level that is substantially better than that specified for the GeoEye Precision products. This product can therefore be used to create new topographic maps of this important landmark at much larger scales than what exist today. This new topography will also support the characterization of the rapid demise of glaciers on Kibo (e.g., Sirguey et al., 2013). Finally, this study provides a practical example of how space-borne sensors can now be used to support surveying applications with relatively strict accuracy requirements.

Acknowledgements

This research was funded by the National School of Surveying and the Department of Geography, University of Otago, New Zealand. ASTER images were obtained with support from the Global Land Ice Measurement from Space (GLIMS). The authors thank Stephen Shirima, principal land surveyor and assistant director of the mapping division at the Ministry of Lands, Housing and Human Settlements Development, Dar es Salaam, Tanzania, for his useful insight. We also thank the KILI2008 Team coordinated by Rui M. S. Fernandes, John Msemwa, and Machiel Bos. The two anonymous reviewers are also thanked for their constructive comments on the manuscript. This research would not have been possible without the support from the following Tanzanian authorities (COSTECH, KINAPA, TANAPA, TAWIRI).

References

Cullen, N. J., P. Sirguey, T. Mölg, G. Kaser, M. Winkler, and S. J. Fitzsimons (2013, March). A century of ice retreat on Kilimanjaro: the mapping reloaded. *The Cryosphere 7*(2), 419–431.

Directorate of Overseas Surveys (1964). *Kilimanjaro, East Africa 1:50,000 (Tanganyika)*. D.O.S. 422, series Y742, sheet 56/2, Edition 1. Directorate of Overseas Surveys for the Tanganyika Government.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf (2007, May). The shuttle radar topography mission. *Review of Geophysics 45*(2), RG2004–.

FGDC (1998). Part 3: National standard for spatial data accuracy. In *Geospatial Positioning Accuracy Standards*, Number FGDC-STD-007.3-1998 in Geospatial Positioning Accuracy Standards, pp. 28. Reston, Virginia: Federal Geographic Data Committee.

Forsberg, R. and C. C. Tscherning (1981, September). The use of height data in gravity field approximation by collocation. *Journal of Geophysical Research 86*(B9), 7843–7854.

Gillman, C. (1923, January). Dr. Klute's map of Kilimanjaro. *The Geographical Journal 61*(1), 70.

Hutchinson, M. F. (1989). A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology 106*, 211–232.

International Mountaineering and Climbing Federation, UIAA (2013). Safety and success on kilimanjaro. retrieved from http://www.theuiaa.org/kilimanjaro.html on September 27, 2013.

KILI2008, T. (2009, 3-8 May). Precise determination of the oorthometric height of Mt. Kilimanjaro. In *Proceedings of FIG Working Week*, Eilat, Israel, pp. 11pp.

Klute, F. (1920). *Ergebnisse der Forschungen am Kilimandscharo, 1912*. Berlin: D. Reimer (E. Vohsen).

Klute, F. (1921). Die stereophotogrammetrishe aufnahme der hochregionen des kilimandscharo. *Zeitschrift der Gesselschaft Erdkunde zu Berlin 56*, 144–151.

Leberl, F., A. Irschara, T. Pock, P. Melxner, M. Gruber, S. Scholz, and A. Wlechert (2010, October). Point clouds: Lidar versus 3D vision. *Photogrammetric Engineering and Remote Sensing 76*(10), 1123–1134.

Meyer, H. (1891). *Across East African glaciers: an account of the first ascent of Kilimanjaro*. London, UK: George Philip & son. Translated from the German by E. H. S. Calder.

Pavlis, N. K., S. A. Holmes, S. C. Kenyon, and J. K. Factor (2012). The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research 117*(B4), B04406–.

Pugh, K. T. (1954). *Height determination of Kilimanjaro*, Volume XII of *Empire Survey Review*, pp. 194–206. Crown Agents' for Overseas Governments and Administration.

Rusu, R. B. and S. Cousins (2011, May 9-13). 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, pp. 1–4.

Saburi, J., N. Angelakis, R. Jaeger, M. Illner, P. Jackson, and K. Pugh (2000, October). Height measurement of Kilimanjaro. *Survey Review 35*(278), 552–562.

Shirima, S. (2013, September 25). personal communication.

Sirguey, P. J., N. J. Cullen, T. Mölg, and G. Kaser (2013, 7-13 December). Ice volume assessment of the northern ice field of kilimanjaro: the photogrammetry strikes back. In *Proceedings of the American Geophysical Union (AGU) fall meeting*, San Francisco, CA, USA, pp. C42A–05. AGU.

Young, J. A. T. and S. Hastenrath (1987). *Glaciers of the Middle East and Africa - Glaciers of Africa* (Richard S. Williams and Jane G. Ferrigno ed.)., pp. G49–G70. Satellite image atlas of glaciers of the World. U.S. Geological Survey.

Emerging data challenges for next-generation spatial data infrastructure

Benjamin Adams Mark Gahegan

Centre for eResearch
The University of Auckland, New Zealand
{b.adams, m.gahegan}@auckland.ac.nz

Abstract

The landscape of spatial data infrastructures (SDIs) is changing. In addition to traditional authoritative and reliably sourced geospatial data, SDIs increasingly need to incorporate data from non-traditional sources, such as local sensor networks and crowd-sourced message databases. These new data come with variable, loosely defined, and sometimes unknown provenance, semantics, and content. The next generation of SDIs will need the capability to integrate and federate geospatial data that are highly heterogeneous. These data comprise a vast observation space: they could be represented in many forms, will have been generated by a variety of producers using different processes and will have originally been intended for purposes that may differ markedly from their later use. There are several discriminative dimensions along which we can describe the properties of the data found in SDIs, such as the data structure, the spatial framework (e.g., field, image, or object-based), the semantics of the attributes, the author or producer, the licensing, etc. These dimensions define a universe of model possibilities for data in an SDI, known as a *model space*. A core research challenge remains to recognise and resolve - to the degree possible - a comprehensive set of *model dimensions* that will enable us to characterise the many possible models by which geospatial data can be represented. A second challenge is to describe the transformations within and between models, and the ways in which these transformations change aspects of the underlying model. Despite recent movement toward semantically described services for SDIs, the scope and range of descriptive dimensions for geospatial data are underspecified. In this paper we present a diverse set of important dimensions that point to a series of challenges for data integration and then describe how both traditional and emergent datasets can be characterised within these dimensions, and point to some interesting differences.

1 Introduction: The evolving role of the SDI

National and regional spatial data infrastructures (SDI) were originally conceived to be centralised geospatial data repositories containing data that came largely from authoritative sources (Masser, 1999; Groot and McLaughlin, 2000; Jacoby et al., 2002). The advent of local sensor webs, web 2.0 and so-called volunteered geographic information (VGI) - i.e.,

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

voluminous geospatial data that are made available from a multitude of sources of varying quality and that are often uncontrolled in terms of their: creation process, representation and content- has changed the landscape (Goodchild, 2007; Budhathoki et al., 2008). While not produced by authoritative agencies, these data can represent better coverage of specific geospatial phenomena or be more timely due to their distributed and unconstrained methods of generation (Coleman et al. (2009)). Thus in many cases they are, in fact, of higher value, e.g., for time critical tasks in emergency response. However, their domain content (or the tasks to which they can be usefully applied) may not be known in advance and may require post-processing to extract. What then is the role of the spatial data infrastructure in such an environment, given that many of the data that analysts and policy makers will find useful may come from such widely varying and incompatible sources? And how can its users understand the utility and reliability of the data products derived from mashing up such heterogeneous datasets? It is a challenging problem because in order to effectively match data to a specific application need we must consider several aspects of the data at once, including not only the spatial framework of the data but also the provenance, semantics, context of authorship, access rights, etc. (what we might call the *pragmatics* of the data to differentiate it from the geospatial semantics of the data) (Pike and Gahegan, 2007; Gahegan et al., 2009). Recent work in merging VGI with SDI has advocated for better semantic representation, using formal languages from the semantic web, and while this is a good step we argue that a more holistic approach is necessary (Janowicz et al., 2010). This is not a new research problem, but to date the relevant research in GIScience has focused on piecemeal solutions to specific strands of the problem, tackled in isolation that do not work together in the orchestrated way that would be necessary to build a more advanced SDI.

In the following section we will introduce our vision for a next-generation SDI. In section 3 we present a diverse set of important dimensions for next-generation SDI. We follow with example for how data transformation can be represented in terms of those dimensions and show examples for both authoritative and other less formal data. Finally, we conclude with a summary of why we think this is an important time to reconsider the role of SDI as a vital component in a connected approach to the science process: i.e., linked science or eScience (Hey and Trefethen, 2005; Mäs et al., 2011).

2 Next generation spatial data infrastructure

In a typical GIS problem-solving workflow, we typically encounter distinct steps such as the following:

1. Locate, gain access to and – to some extent – understand the limitations of each dataset we intend to use. Currently, SDI and specifically their data catalog and search tools can sometimes help here.
2. Transform the datasets we will use into a consistent form (model), for example by re-projecting, converting from raster to vector or harmonising the semantics. The decisions we make here can have profound implications for the quality of the data.
3. Combine the datasets via an analytical workflow of some kind.
4. Assess the accuracy and reliability of the result and (possibly) publish it back into the SDI.

The geospatial datasets that we might wish to combine could be highly heterogeneous. They will be represented in many forms, will have been generated by a variety of producers using different processes and may have originally been intended for purposes that are different from their present use. Each of these ideas, and others, form the dimensions along which we can describe the properties of a dataset found in an SDI, such as the data structure, the spatial framework (e.g., field, image, or object-based), the semantics of the attributes, the author or producer, etc. These dimensions define a universe of model possibilities, or *model space*, for datasets that the SDI interacts with. In order to capitalise on the value of such a wide variety of data, we need to detail the many ways in which we might integrate or transform data that reside at different points in model space. A core research challenge, therefore, is to recognise and resolve - to the degree possible - a comprehensive set of characteristic model dimensions[1] that enable us to characterise transformations within and between data models. These dimensions provide us with a conceptual framework to understand the ways that data are transformed and are made fit-for-purpose.

A data source, such as the Landsat 7 sensor or a crime logging system has the potential to create a series of datasets, so a source can be represented in this model space similarly to an individual dataset. But rather than being represented as one point in model space, a data source may be represented as ranges along certain dimensions, describing the potential values that a specific dataset may inherit. For example, each Landsat 7 dataset will have a unique timestamp and a spatial footprint drawn from a set of possibilities defined by the orbital characteristics. But the spatial framework will always be an image and the data will always be packaged into a raster data structure.[2]

[1] The term 'dimension' is used loosely here in a cognitive sense and does not imply an ordering of values, as in the mathematical sense of the word.

[2] Interestingly, the error characteristics of the datasets change over time as the sensor picks up damage, so this too is a range rather than a point.

Moving a dataset from one point in model space to another point will incur a series of costs related to: the work done, changes in accuracy or resolution, changes in semantics, etc. We routinely DO move data in model space but we typically do not account for all the changes that ensue. We aim to address this shortcoming by representing the model space and describing (as richly as we can) what happens to datasets that are transformed from one point in this space to another. We can assign a cost function for each dimension (i.e., a distance metric) that allows us to account for the cost of transforming data from one model to another. A data transformation is represented as a function that takes one or more datasets and their associated models and returns a tuple consisting of a new dataset and its location within the model space. Finally, each model in the space has its own sets of behaviours, translators (to other models), constraints, and supported data structures.

An important difference between the traditional SDI and next-generation SDI is that because of the heterogeneity of producers, unlike the traditional model of a centralised repository, the next-generation SDI will be distributed and federated. It will thus need to incorporate data from disparate sources that are not controlled from within the SDI, in line with the paradigm of linked data (Bizer et al., 2009; Schade et al., 2010). Perhaps more importantly, the idea of SDI as simply an ingester of data will change. Geospatial data sources such as sensor networks and social media feeds are increasingly real-time and configurable. For example, a sensor network may be able to sample some phenomenon every day, hour or minute and may be able to report the value in a variety of different units (e.g. Celsius or Fahrenheit). An SDI may also need to *negotiate* with its data sources on behalf of the user. This means that the tasks that the SDI performs are not just search and integration (i.e. pulling) but also communication and requests for re-configuration and new information (i.e. pushing). From the perspective of describing the characteristics of data, therefore, the purpose is not only for publishing, sharing, and integration of data from static sources but also for communicating "what the user wants" back to sources, so they can better meet the need. Developing this capability will become essential because we can easily imagine that a universal data harmoniser might require a combinatorial explosion of pairwise translators. It might be much easier just to ask the source again for the data to match the user's needs![3]

Figure 1 provides a schematic view of one way such a next-generation SDI might be architected. At the heart of the proposed infrastructure are three layers of functionality shown in shades of green. An outer **Federation and Analysis Layer**, in which all supported datasets are descriptively rich, interoperable, and can be readily combined in analysis. This federation layer serves as the shell in which all data known to the infrastructure can be discovered, queried, analysed or shared. A **Mediation Services Layer** comprising of software services to transform geospatial data sources with the supported conceptual models in the Kernel. Specific mediation services to harmonise a given data source are shown as jigsaw pieces in the figure. The services in this layer are created and maintained by geospatial knowledge engineers, and are used by domain experts to create the required mediation services. A **Knowledge Representation Kernel**, used to describe and create rich descriptions of geospatial knowledge and the conceptual models of geospatial information that underpin the various exchange formats and analysis methods, along with their description semantics. This layer also includes the fit-for-purpose reasoning, which uses rich descriptions to create a 'recommender system' for geospatial data selection.

3 Dimensions of model space

The volume, variety and velocity of geospatial data create an extremely large model space that must be navigated smartly to facilitate meaningful discovery and analysis activities (Cavoukian and Jonas, 2012). Any fruitful paths taken through the space in order to transform the data to make it commensurate will be dictated by the application contexts. From a theoretical point of view then, the core research problems require us to identify the useful morphisms[4] that map from one part of the model space domain to another. The difficult research challenges arise because geospatial data is often highly contextual (interpreted), and there are many models in use, often with missing or implied dimensions such as semantics, accuracy and authority.

The creation of harmonised information from heterogeneous datasets presents us with several research challenges. Table 1 summarises some of these challenges. Most are current research themes that are usually studied separately in the GIScience literature that will need to be: (i) extended where needed, then (ii) integrated together. For the purposes of this paper, we have chosen six different domains along which we can define characteristic dimensions for geospatial data: 1) spatio-temporal frameworks, 2) semantics, 3) access and licensing, 4) provenance, 5) authority, and 6) quality. These domains are not entirely separable, as values in one may have a bearing on others in many cases, but it serves as a useful hierarchical organisation for the myriad dimensions that can describe geospatial data. The first three challenges

[3] Of course it makes sense to avoid the need for pairwise translators by favouring a small number of models with well-understood paths between them, as we currently see in most GISystems.

[4] A morphism is a structure-preserving mapping between two abstract conceptual or mathematical structures, such as is used in set theory and various description algebras.

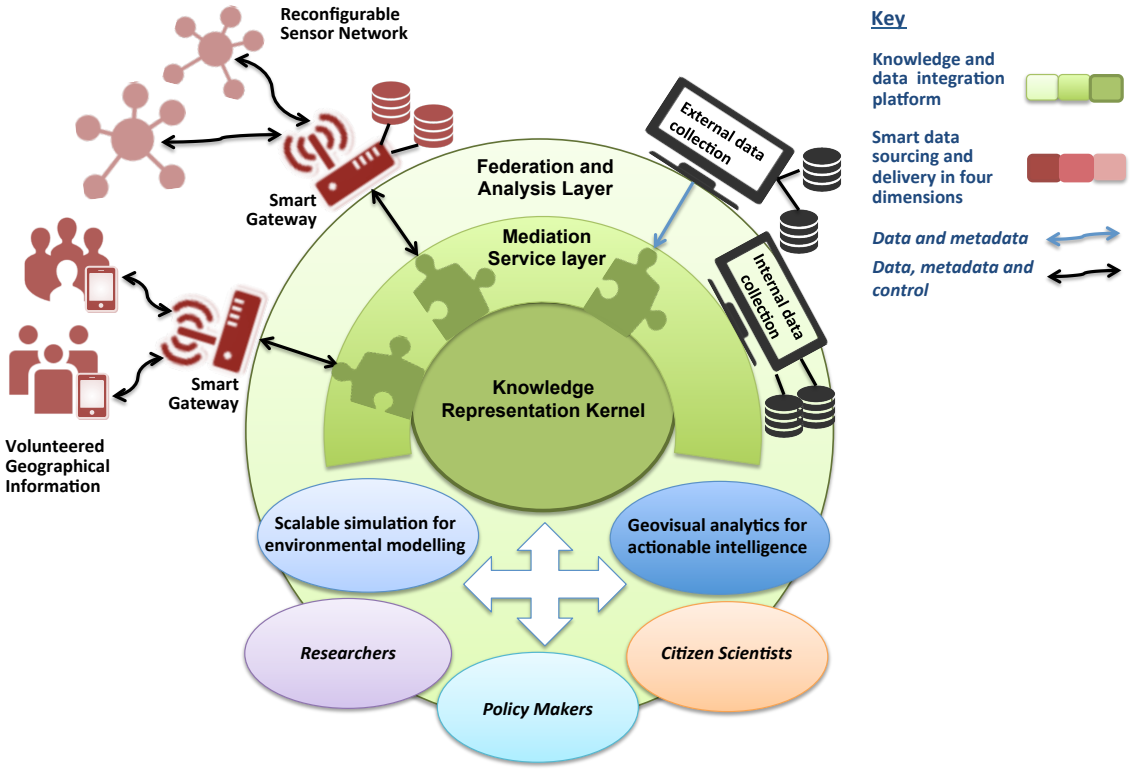


Figure 1: An holistic view of how a next-generation SDI might be architected to harmonise heterogeneous data for multiple purposes.

address data harmonisation, the latter three expand to issues of governance and lead to a holistic measurement of *fitness-for-purpose* (Georgiadou et al., 2006; Devillers et al., 2007). Of course, many other domains could be chosen, these are not an exhaustive set, and new domains may emerge in the future (for example to support epistemology to go with ontology).

We consider the dimensions described here are important to most contemporary SDIs. The actual computational solutions for measuring along these dimensions will to some degree depend on the needs of the system being developed. Thus, we are not proposing a universal framework that will work for all data and every SDI, but rather a series of "best-practices" combined with a highlighting of what we see as the most pertinent research challenges to advance SDI. We fully anticipate that the structure of these dimensions will become more nuanced as we develop computable frameworks that can reason over all of them in concert.

3.1 Spatio-temporal frameworks

Spatio-temporal frameworks have been extensively studied in GIScience over that last couple of decades. These spatiotemporal models (which we term here spatial frameworks after Worboys and Duckham (2004)) concern the how the structure of space and time are represented in the data. While the format of the data (i.e., the syntactic description of the data) will often imply a specific spatial framework, that is not always the case – the same format can be used for data described by different models and vice versa. Some of the challenges for data harmonisation include the following aspects of the spatial and temporal framework in which the data reside. The tessellation of the space can be continuous, a regular grid, or an irregular grid. Worboys and Duckham (2004) describe the common bifurcation between field-based and object-based representations, though alternate models are proposed, including image-based, which has characteristics of both, as well as non-spatially explicit models, which are becoming more prevalent in research on place-based representation (Gahegan, 1996; Winter and Truelove, 2013). Other important spatiotemporal dimensions concern the projection, scale, and resolution of the data. Finally, the representation of time can be continuous or discrete.

Dimensions of *model space*	Challenges for data harmonisation	Possible solutions
Spatio-temporal frameworks (Worboys and Duckham, 2004; Gahegan, 1996)	• Tessellation of the space (continuous, regular grid, irregular grid), Field-, image-, object-based, or non-spatially explicit • Projection, scale, and resolution • Time	Devise a formal, holistic conceptual framework to encompass the variety of models along with the required tools to move data between these models.
Semantics of attributes (Egenhofer, 2002; Bishr and Kuhn, 2007; Brodaric and Gahegan, 2007; Gahegan et al., 2009; Adams and Janowicz, 2011; Janowicz et al., 2013)	• Measurement scale (ratio, interval, ordinal, categorical, unstructured, tuple) • Implied or missing semantics • Ontology alignment	Develop ontology creation and alignment tools, using both formal (top down) and informal (bottom-up, via use-cases) approaches.
Access and licensing (Onsrud et al., 1994; Miller et al., 2008; Cavoukian and Jonas, 2012; Hosking and Gahegan, 2013)	• Access rights to the data, according to purposes • Rights to update or propagate changes	Research suitable security models for use in a distributed setting.
Provenance (Clarke and Clark, 1995; Bose and Frew, 2005; Simmhan et al., 2005; Ludäscher et al., 2006; Belhajjame et al., 2013)	• Author and source • Workflow used to generate data	Extend current provenance research to explicitly represent key aspects of geospatial information processing.
Authority (Gahegan and Pike, 2006; Flanagin and Metzger, 2008; Coleman et al., 2009; Bishr and Kuhn, 2013)	• Authoritativeness of the source (top down) • Trustworthiness of the contributing individual or organisation (bottom up)	Develop models for digital governance that can encompass both *imposed* and *earned* authority
Quality (Chapman, 2005; Pike and Gahegan, 2007)	• Confidence in intended semantics • Confidence in the process used to generate data • Propagation of uncertainty through the analysis workflow	Extend and integrate spatial data accuracy methods to work within this context.

Table 1: Summary of some of the complex dimensions that comprise the model space for geospatial data and some of the related harmonisation challenges.

3.2 Semantics of attributes

Representing the meaning of attributes, i.e., the non-spatial data associated with features represented in a geographic dataset, presents a significant challenge to the next-generation of SDI. When present these semantics are often communicated informally, e.g., as table column labels, which poses a problem for building an SDI designed to do automated data transformation and harmonisation. Movement toward more formal representation of attribute semantics using the languages of the semantic web has been advancing but much VGI data are described in less structured ways (e.g., folksonomic tags) (Egenhofer, 2002; Bishr and Kuhn, 2007; Janowicz et al., 2013). Thus, while the semantics of the attributes are ideally well-understood by the creators of data, they are often missing when data are communicated. It is also the case that meanings of geographic concepts vary not only across but within communities (Brodaric and Gahegan, 2007). Toward the goal of calculating the "fitness-for-purpose" of a dataset, it will not always be the case that an estimation of attribute meaning can be made through formal reasoning and ontology alignment but rather will rely on other fuzzier rules and patterns (derived either through data mining and machine learning or via use-cases) to suggest better or worse fitness-for-purpose (Gahegan et al., 2009; Adams and Janowicz, 2011).

3.3 Access and licensing

With a more distributed and federated structure and subsequently less control over many data sources, the data processed through an advanced SDI will likely be governed by multiple and at times conflicting access and licensing restrictions. Despite the academic community's interest in open licensing for linked data, much geospatial data that we want to make accessible through SDI will be restricted in terms of use (Miller et al., 2008). This includes derived products from analyses of crowdsourced data collected through commercial applications such as Twitter[5]. We will need to develop ways of characterising the access and licensing models for secondary datasets that are derived from multiple, external sources and published through the SDI (Hosking and Gahegan, 2013). The SDI must be able to negotiate between a user profile model that describes access rights and the access model for the data. Another important aspect of access management is the need to build privacy-preserving mechanisms into the data by design as much as possible (Onsrud et al., 1994; Cavoukian and Jonas, 2012).

3.4 Provenance

In addition to modelling the state changes in the data model, a next generation SDI will also maintain descriptions of the provenance semantics of the data. These aspects of provenance include information about the author and source as well as the workflow used to generate the data. Knowing the path that a dataset has taken through the entire model space would provide very useful insight into its likely accuracy and utility for a specific task. Research on provenance representation in eScience and scientific workflow systems will be extended to represent key aspects of geospatial information processing operations and how they relate to the model space (Clarke and Clark, 1995; Bose and Frew, 2005; Simmhan et al., 2005; Ludäscher et al., 2006; Belhajjame et al., 2013).

3.5 Authority

Authority refers to a characterisation of the data producer in terms of its status within a Community of Practice (Gahegan and Pike, 2006; Coleman et al., 2009). Formal integration of authority models into SDI are increasingly needed now as we move away from the architecture of a centralised repository. We can describe authority of the source top down, or rather represent it bottom up in terms of trustworthiness of the contributing individual or organisation. Community and trust in VGI is often an emergent phenomenon where contributors gain trust through community interaction and past behaviour (Flanagin and Metzger, 2008). In absence of direct feedback on trust, proxies such as the spatial location of the contributor can be useful indicators (Bishr and Kuhn, 2013).

3.6 Quality

An important set of dimensions for describing data include measures of data quality that are independent of the task (Chapman, 2005). We distinguish measures along these dimensions of quality from evaluations of *fitness-for-purpose*, which we see as an outcome of all the dimensions described in this paper and based on a situational context (Pike and Gahegan, 2007). Examples of quantitative representations of quality are confusion (or error) matrices in a land-cover layers or surveying errors. Other measures of quality are more qualitative, e.g. confidence in the intended semantics of the attributes or confidence in the process that was used to generate the data.

[5] `https://twitter.com/`

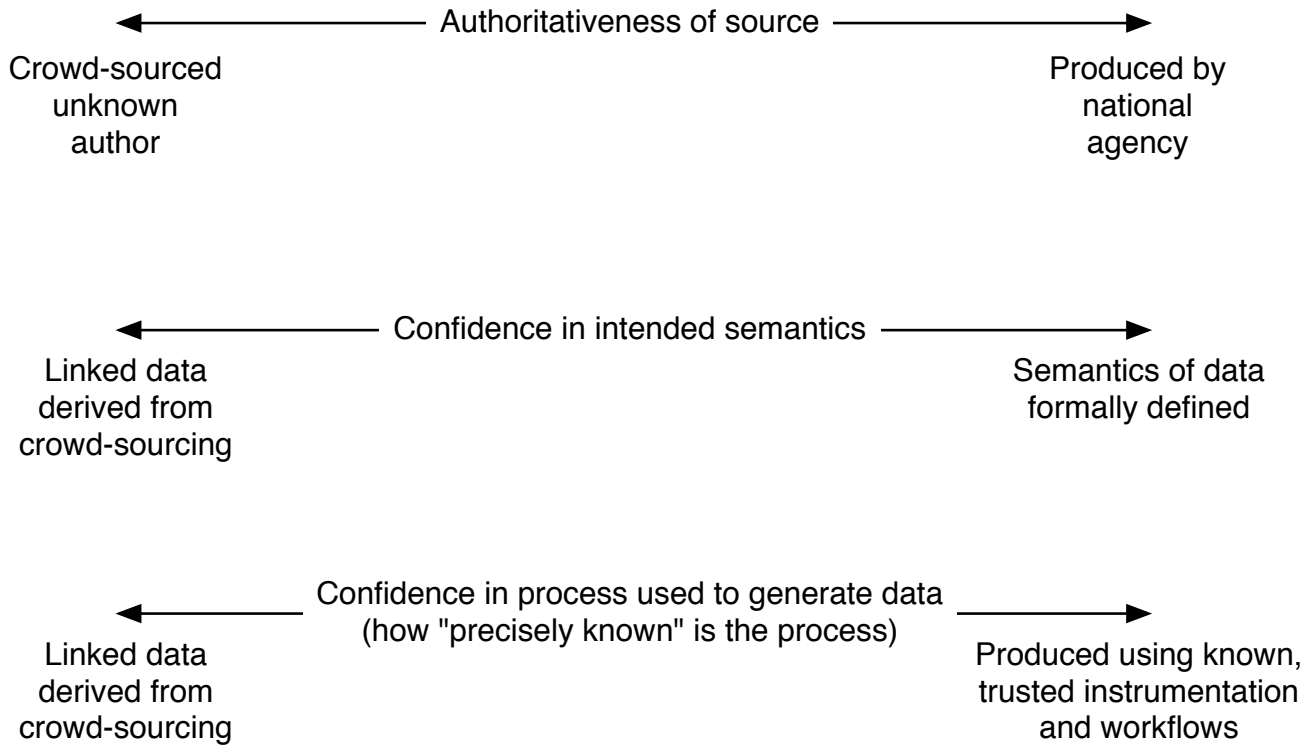


Figure 2: Some dimensions of authority and quality contributing to fitness-for-purpose.

The following are some examples of the many different factors that can influence the positioning of a dataset in model space along Quality and Authority dimensions:

1. Field guide describing how data are collected
2. Recorded workflow
3. Stamp of authority from a certifying organisation, which implies a rigorous process
4. Tiers of compliance
5. Peer approval ranking (trust)

Figure 2 illustrates three sample quality and authority dimensions of the model space, along which a dataset can be described. The first dimension is a measure the authoritativeness of the data source, which is defined within the context of a Community Of Practice. The second dimension characterises the semantics of the data; i.e., what they are intended to mean by the original source. The third dimension characterises how precisely known is the process by which the data came to be in its present form prior to being incorporated into the SDI. As with others, these dimensions might be correlated in certain contexts.

4 User Interaction

We envision that the primary function of a spatial data infrastructure will be to enable users to transform, combine, and fit geospatial data to the task at hand – i.e. to provide data that are fit-for-purpose, where the purpose is defined by the application context (Frank et al., 2004). In terms of the model space, this entails the following. First, we must identify where in the model space the user wants to be. Then the system needs to do one of the following: 1) transform data that exist in different models to the one the user wants (Figure 3a), 2) if the user is flexible in terms of the model, identify a set of candidate data from similar models (Figure 3b), or 3) communicate to a data source the need for *new* data that will match the needs of the user. In addition to a rich description of the data in model space, an essential component of an advanced SDI is a user profile that represents preferences in relation to the model space. User templates based on common categories of users and which learn based on previous behaviour of similar users will add efficiency to this functionality. To quote

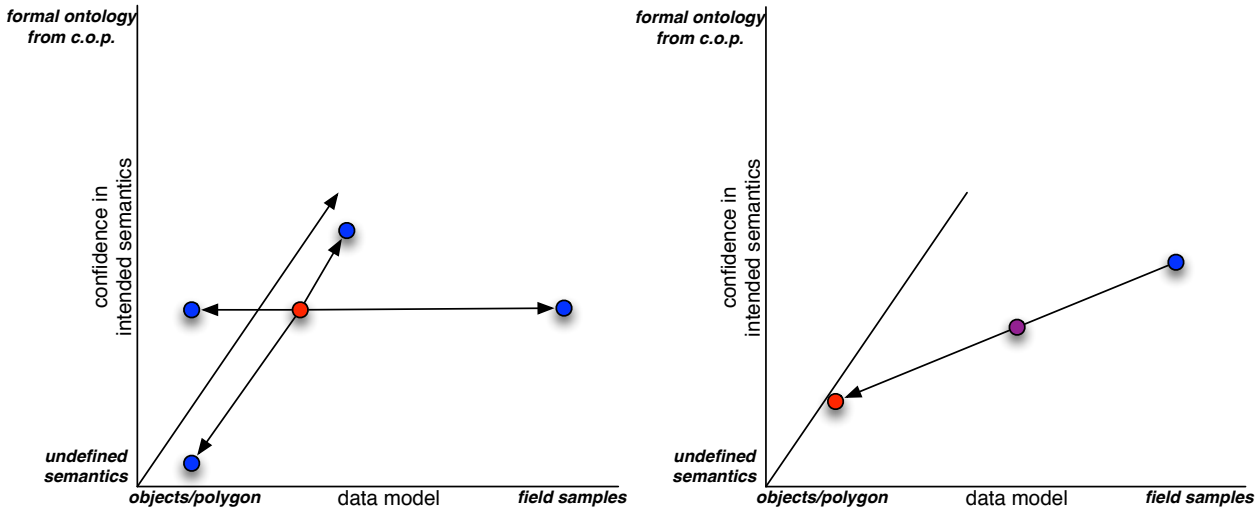


Figure 3: This figure is a schematised representation of example transformations in the model space. The red dot indicates the model desired by the user and the blue dots represent models of data that are available. (a) In the left picture, the system provides a set of data objects (shown in blue) that have similar models but which vary with respect to the dimensions along which they are different. (b) In the right picture, the data object is transformed to fit the desired model through a series of state changes.

the *Hendler hypothesis*[6] we anticipate that "a little semantics will go a long way" to help constrain the context for finding, transforming, and harmonising data. For example, knowing something of the scientific domain in which a user works, such as geodesy or criminology, can help determine where in the model space is best for providing useful data. Likewise, if a user is known to specialise on kinds of features found in specific localities (e.g., a vulcanologist) the SDI can favour data that match the appropriate scale and projection for those locations. The user profile might also provide restrictions on what data are available, based on, e.g., the access model of the data.

Because we describe the model space abstractly in terms of conceptual dimensions, it is a research challenge to develop methods to communicate these dimensions so that a user understands where in the model space they are and where they want to be. We propose that an exemplar-based system that provides visual cues to the user by attaching recognisable icons to common points in the model space will reduce the cognitive load somewhat. Figure 4 shows an example of icons that might be used to reference models that vary along dimensions for the spatio-temporal framework and the measurement scale of the attribute domain. These visual exemplars provide a frame of reference for the user.

4.1 Transforming data

Every GIS operation that transforms data can be more formally represented as movement along a state transition diagram. The state changes shown in Figure 5a are an example of the kinds of transformations that are possible between data represented with different spatial frameworks and attribute measurement semantics. State changes will incur information loss (or possibly gain) depending on the kind of operation done, which will contribute to the measure of quality (accuracy). The process of data transformation will be recorded as a workflow to enable users to reconstruct the steps used in analyses. Every state change is modelled as a binary property change in the data model, e.g., transforming from a continuous value to a categorical value, though richer descriptions may eventually be possible. The dashed connections between the discrete space / field models and continuous space / object models have a step in between, such as discrete space / object model. Rarely, the data is stored in these intermediate representations, it is more typical to use them only as a temporary step. For example, a raster to vector operation that converts integer-valued ordinals will first identify objects in the image space based on connected components with the same value (DS-O-Ord). Then, the space is made continuous by converting the edges of the regions to vector form (CS-O-Ord).

Figure 5b shows an example of how the transformation of sensor measurements of rainfall to polygonal regions can be modelled using this state diagram. The dataset begins as a set of rainfall measurements collected by a sensor network in the environment. The data are field-based measurements that are continuous-valued (e.g., in mm). In step 2, an operation

[6] `http://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html`

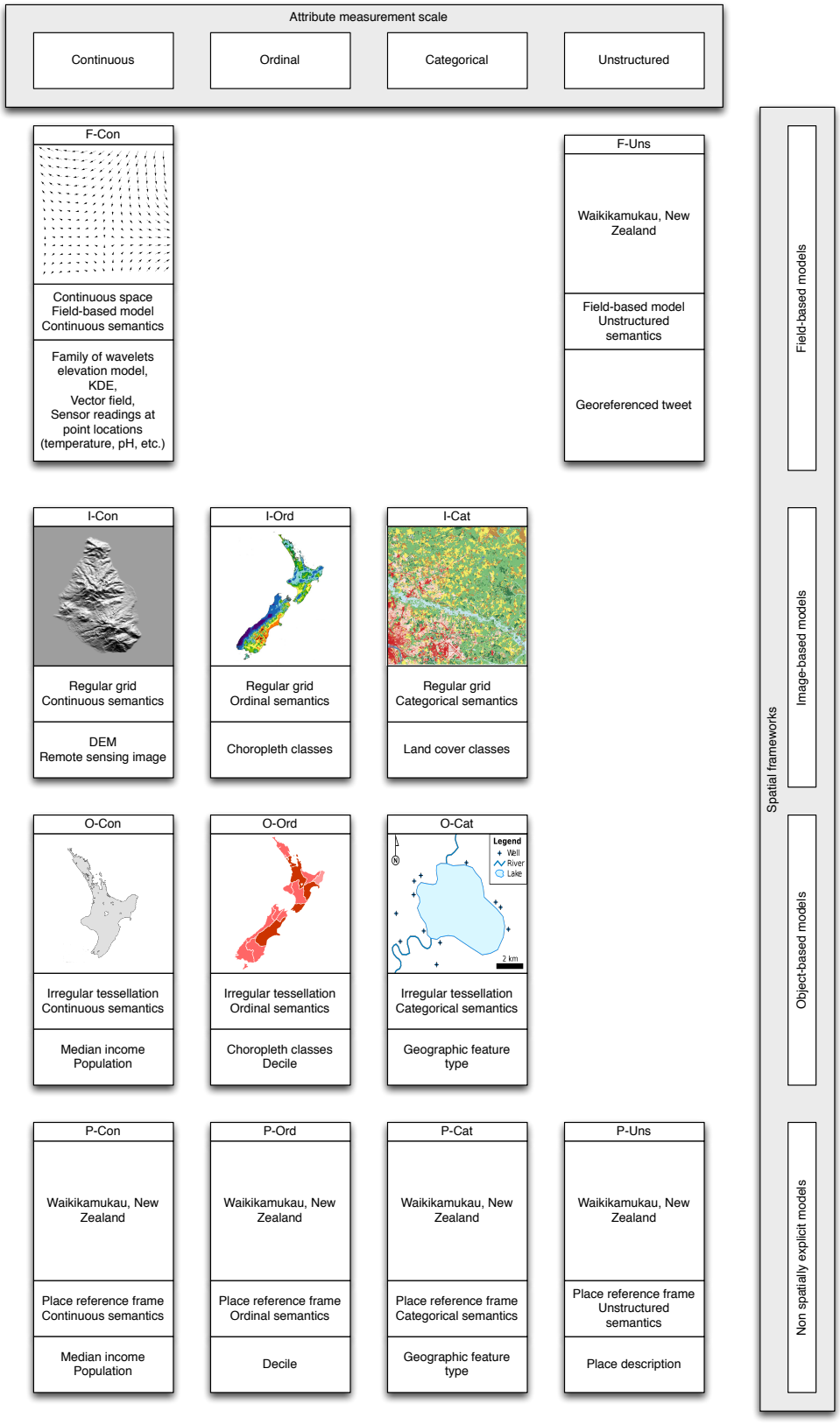


Figure 4: Example of conceptual data model icons that can be used by an SDI to communicate common points in the model space to data consumers. Visual exemplars provide a frame of reference for the consumer.

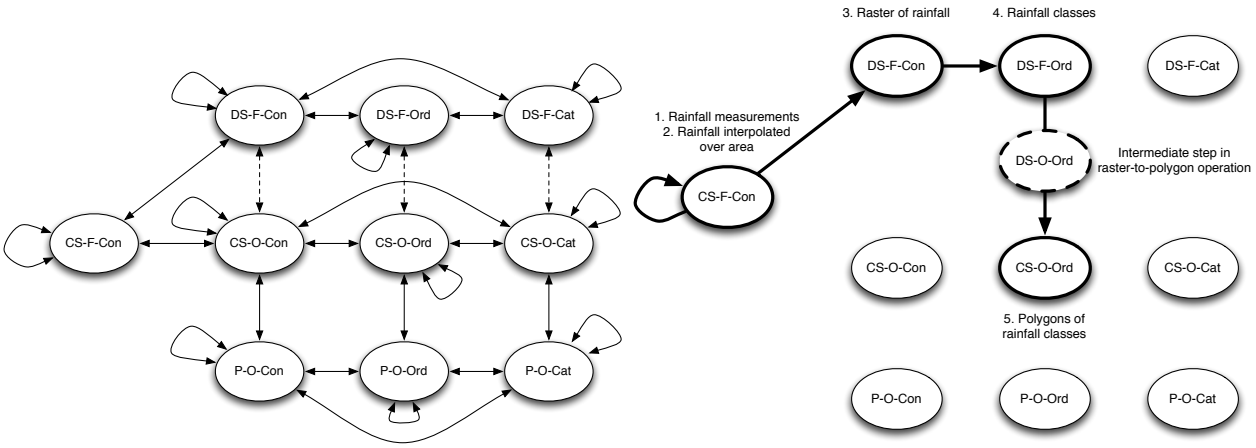


Figure 5: Data transformation as a movement in a state transition diagram: (a) Data model state diagram showing data transformation on the edges; (b) Example of state transformation for rainfall data.

interpolates values for locations not being observed, and the result is converted to an image-space representation in step 3. The continuous values in the raster pixels are binned into ordinal classes in step 4, and a raster-to-polygon operation transforms the field-based view into discrete objects (represented by polygons) in step 5.

In order to combine data from different sources meaningfully in an analysis, it is often necessary that they be transformed to share the same spatio-temporal framework. Note of course that just because datasets share the same spatial framework they may still be incommensurable due to, e.g., different categorisation schemes, and so harmonisation will require transformation along other dimensions, such as semantics.

5 Summary and conclusions

If a researcher is able to clearly state their needs in terms of the dimensions of the model space, then the system can more clearly provide recommendations of data to use (or not use). The next-generation SDI we are proposing is moving away from the model of a centralised data warehouse toward a federated data network model, where the SDI acts as a mediator that harmonises data generated from heterogeneous sources and provides data to a user. New data sources, such as sensor networks and crowd sourcing, that provide streaming and high-resolution data in real-time will become important parts of the SDI. The data from these sources will be emergent and dynamic. At the edge of the system architecture these data will need to be translated to the model space adopted by the SDI. Despite the challenges of harmonising data from so many different kinds of sources, the dynamic nature of these sources provides opportunity to significantly expand the "control" functionality of the SDI to push data requests out to the sensors, thereby allowing a user to interact directly with the data providers. By specifying a data-need (in terms of location in model space) at the user side, an advanced SDI should be able to push a request for data that fits the specified parameters out to the data sources, which will reconfigure and acquire the needed data.

Formalising this control mechanism is an important research challenge and will require the development of "smart gateways" with the appropriate communication protocols that 1) broadcast the range of models available to the SDI, 2) reconfigure components as needed, 3) wrap the physical data output with a description vis à vis the model space, and 4) handle data requests from the SDI. For example, for a sensor network requests might involve reconfiguring to sample at higher temporal resolution. We can imagine these smart gateways working for human sensor networks as well. For example, spatially-referenced social media data has been shown to be a valuable data source for health-related information (Paul and Dredze, 2011). With a formal representation that maps Twitter hash tags to health-related concepts, a well-designed smart gateway could – based on the statistics of tags in the Twitter feed – communicate to an SDI that it is able of producing flu-related datasets on request, at certain spatial and temporal resolutions. A related challenge exists for the traditional SDI data catalog: how are the range of possibilities for such a dataset to be represented to the user? More generally, a number of usability challenges exist, including communicating the dimensions of the model space to consumers of data and facilitating interaction between consumers and producers.

Perhaps the biggest challenge to the current status quo concerns how we represent geospatial features and sets of features (datasets). Most data exchange formats, whether proprietary or open, do not support the inclusion of the kind of rich semantics and pragmatics described here. However, all that is required is that geospatial data carries with it a persistent

link back to the rich descriptions formulated and maintained by the advanced SDI, such as a SPARQL endpoint. There is some movement towards richer descriptions in open exchange formats, The Open Geospatial Consortium (OGC) initiatives already support the notion of legends, and the beginnings of workflow and domain semantics to go with its feature and map exchange standards[7].

References

Adams, B. and K. Janowicz (2011). Constructing geo-ontologies by reification of observation data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 309–318. ACM.

Belhajjame, K., J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, and J. Zhao (2013). PROV-O: The PROV ontology. Technical report.

Bishr, M. and W. Kuhn (2007). Geospatial information bottom-up: A matter of trust and semantics. In *The European information society*, pp. 365–387. Springer.

Bishr, M. and W. Kuhn (2013). Trust and reputation models for quality assessment of human sensor observations. In *Spatial Information Theory*, pp. 53–73. Springer.

Bizer, C., T. Heath, and T. Berners-Lee (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS) 5*(3), 1–22.

Bose, R. and J. Frew (2005). Lineage retrieval for scientific data processing: a survey. *ACM Computing Surveys (CSUR) 37*(1), 1–28.

Brodaric, B. and M. Gahegan (2007). Experiments to examine the situated nature of geoscientific concepts. *Spatial Cognition and Computation 7*(1), 61–95.

Budhathoki, N. R., B. C. Bruce, and Z. Nedovic-Budic (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal 72*(3-4), 149–160.

Cavoukian, A. and J. Jonas (2012). Privacy by design in the age of big data. *Office of the Information and Privacy Commissioner*.

Chapman, A. D. (2005). *Principles of data quality*. GBIF.

Clarke, D. G. and D. M. Clark (1995). Lineage. *Elements of spatial data quality*, 13–30.

Coleman, D. J., Y. Georgiadou, and J. Labonte (2009). Volunteered geographic information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research 4*(1), 332–358.

Devillers, R., Y. Bédard, R. Jeansoulin, and B. Moulin (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science 21*(3), 261–282.

Egenhofer, M. J. (2002). Toward the semantic geospatial web. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, pp. 1–4. ACM.

Flanagin, A. J. and M. J. Metzger (2008). The credibility of volunteered geographic information. *GeoJournal 72*(3-4), 137–148.

Frank, A. U., E. Grum, and B. Vasseur (2004). Procedure to select the best dataset for a task. In M. J. Egenhofer, C. Freksa, and H. J. Miller (Eds.), *GIScience*, Volume 3234 of *Lecture Notes in Computer Science*, pp. 81–93. Springer.

Gahegan, M. (1996). Specifying the transformations within and between geographic data models. *Transactions in GIS 1*(2), 137–152.

Gahegan, M., J. Luo, S. D. Weaver, W. Pike, and T. Banchuen (2009). Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure. *Computers & Geosciences 35*(4), 836–854.

[7] `http://www.opengeospatial.org/standards/wmc`

Gahegan, M. and W. Pike (2006). A situated knowledge representation of geographical information. *Transactions in GIS 10*(5), 727–749.

Georgiadou, Y., O. Rodriguez-Pabón, and K. T. Lance (2006). Spatial data infrastructure (SDI) and e-governance: A quest for appropriate evaluation approaches. *URISA-WASHINGTON DC- 18*(2), 43.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal 69*(4), 211–221.

Groot, R. and J. D. McLaughlin (2000). *Geospatial data infrastructure: concepts, cases, and good practice*. Oxford university press Oxford.

Hey, T. and A. E. Trefethen (2005). Cyberinfrastructure for e-Science. *Science 308*(5723), 817–821.

Hosking, R. and M. Gahegan (2013). The effects of licensing on open data: Computing a measure of health for our scholarly record. In *The Semantic Web–ISWC 2013*, pp. 432–439. Springer.

Jacoby, S., J. Smith, L. Ting, and I. Williamson (2002). Developing a common spatial data infrastructure between state and local government–an australian case study. *International Journal of Geographical Information Science 16*(4), 305–322.

Janowicz, K., S. Schade, A. Bröring, C. Keßler, P. Maué, and C. Stasch (2010). Semantic enablement for spatial data infrastructures. *Transactions in GIS 14*(2), 111–129.

Janowicz, K., S. Scheider, and B. Adams (2013). A geo-semantics flyby. In S. Rudolph, G. Gottlob, I. Horrocks, and F. van Harmelen (Eds.), *Reasoning Web. Semantic Technologies for Intelligent Data Access*, pp. 230–250. Springer.

Ludäscher, B., I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao (2006). Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience 18*(10), 1039–1065.

Mäs, S., M. Müller, C. Henzen, and L. Bernard (2011). Linking the outcomes of scientific research: Requirements from the perspective of geosciences. In T. Kauppinen, L. C. Pouchard, and C. Keler (Eds.), *LISC*, Volume 783 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Masser, I. (1999). All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science 13*(1), 67–84.

Miller, P., R. Styles, and T. Heath (2008). Open data commons, a license for open data. In *LDOW*.

Onsrud, H. J., J. P. Johnson, and X. Lopez (1994). Protecting personal privacy in using geographic information systems. *Photogrammetric Engineering and Remote Sensing 60*(9), 1083–1095.

Paul, M. J. and M. Dredze (2011). You are what you tweet: Analyzing twitter for public health. In L. A. Adamic, R. A. Baeza-Yates, and S. Counts (Eds.), *ICWSM*, pp. 265–272. The AAAI Press.

Pike, W. and M. Gahegan (2007). Beyond ontologies: Toward situated representations of scientific knowledge. *International Journal of Human-Computer Studies 65*(7), 674–688.

Schade, S., C. Granell, and L. Diaz (2010). Augmenting SDI with linked data. In K. Janowicz, T. Pehle, G. Hart, and P. Maué (Eds.), *Proceedings of the Linked Spatiotemporal Data Workshop (LSTD 2010 GIScience 2010 Zürich)*, pp. 34–45.

Simmhan, Y. L., B. Plale, and D. Gannon (2005). A survey of data provenance in e-science. *ACM Sigmod Record 34*(3), 31–36.

Winter, S. and M. Truelove (2013). Talking about place where it matters. In *Cognitive and Linguistic Aspects of Geographic Space*, pp. 121–139. Springer.

Worboys, M. and M. Duckham (2004). *GIS: a computing perspective*. CRC press.

An image engineering approach to analysing mobile mapping data

Michael Borck
Dept. Spatial Sciences
Curtin University
m.borck@curtin.edu.au

Geoff West
Dept. Spatial Sciences
Curtin University
g.west@curtin.edu.au

Tele Tan
Dept. Mechanical Engineering
Curtin University
t.tan@curtin.edu.au

Abstract

Vehicle-based mobile mapping systems capture co-registered imagery and 3D point cloud information over hundreds of kilometres of transport corridor. Methods for extracting information from these large datasets are labour intensive. These need to be easily configured by non-expert users to process images and develop new workflows. Image Engineering provides a framework to combine known image processing, image analysis an image understanding methods into powerful applications. Such a system was built using Orange an open source toolkit for machine learning onto which image processing, visualisation and data acquisition methods were added. The system presented here enable users who are not programmers to manage image data and to customise their analyses by combining common data analysis tools to fit their needs. Case studies are provided to demonstrate the utility of the system. Co-registered imagery and depth data of urban transport corridors provided by the Earthmine dataset and laser ranging systems are used.

1 Introduction

Computer vision research is moving to the stage where quite complex systems can be used in real world applications, although in most cases the methods are turnkey. There are applications in which a non-expert user needs to configure a complex sequence of processes for their application. Such an application is the processing of co-registered imagery and 3D point cloud information acquired from a moving vehicle along transport corridors. GPS and inertial guidance allows the data to be registered to the world coordinate system enabling reasonably accurate location information to be acquired including the location of street side furniture, width of roads, power line to vegetation distances etc. Such systems can acquire enormous amounts of data quite quickly. For example the business district of Perth, Western Australia consists of 320kms of roads resulting in 120GBytes of imagery and 2GBytes of point cloud information (compressed).

In modern mobile mapping systems there are three common methods for data collection: image, point cloud, and a combination of the two through co-registered data. It is not clear which is the best as there are many algorithms for object detection that have varying measures of success. Imagery is cheap to acquire and can be interpreted by humans as well as by machine. It also contains colour information that is needed for some

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

applications. Point clouds are more expensive to acquire, have no colour information and often have gaps in the data. Most point clouds for mobile mapping have low density along the direction of travel unless multiple scans are performed e.g. six, three in each direction. Finally, imagery, stereo derived point clouds and laser scanning can be combined to exploit the strength of each method.

Data from two different systems was used in this paper. The Earthmine system captures panoramic imagery and uses stereo algorithms to generate co-registered 3D point clouds. Figure 1(a) shows the Earthmine system on a car. The other system, AAM, generates depth maps from laser ranging sensor. The AAM system provides generated co-registered imagery and 3D point clouds captured from mobile scanning systems such as shown in Figure 1(b).

(a) (b)

Figure 1: (a) Earthmine system showing two sets of panorama cameras (b) MDL mobile laser scanner system with camera

Earthmine has developed a server-based system that allows the querying via location to obtain data about a particular location. The data can be processed using a number of methods e.g. randomly, from a number of known locations, or by "driving" along the capture pathway.

In many applications of computer vision, workflow is an important consideration. A user would typically read in some acquired data, process it interactively and produce the desired result. In many recognition applications, much tuning is needed requiring a user skilled in such techniques. The challenge is to produce a workflow that an non-expert user can use to configure a complex process such as object detection. To do this, a user must be able to view selected parts of the data, identify objects of interest and train a system to use the best features for recognition through a feature selection process combined with some form of pattern recognition method such as a decision tree.

This paper presents a system that gives the non-expert the opportunity to develop powerful image processing applications. The system builds on existing open source frameworks and libraries. Case studies are provided to demonstrate the utility of the system.

2 Background

Zhang (2006) defines Image Engineering (IE) as the collection of three related computer vision processes, image processing (IP), image analysis (IA), and image understanding (IU). Figure 2 show how IP, IA, and IU build up three layers of IE. Each operate on different elements. IPs primary element is the image pixel, IAs processing unit is the object, and IUs operand is a symbol or label providing meaning to an image component or segment. Each layer works with different semantic levels. From low semantic level at IP to high semantic level at IU. The three layers follow a progression of increasing abstractness and of decreasing compactness from IP to IU.

IP primarily includes the acquisition, representation, compression, enhancement, restoration, transformation and reconstruction of images. IP manipulates an image to produce another (improved) image. IA is concerned with the extraction of information from an image. IA takes an image as input and outputs data. Here, the extracted data can be the measurement results associated with specific image properties or the representative symbols of certain object attributes. IU transforms the data into descriptions allowing for decisions and actions to be taken according to the interpretation of the images.

The development of software is a complicated and tedious process, requiring highly specialized skills in systems programming (Bentrad et al., 2011). A program is usually defined as a sequence of instructions executed by a computer, so any system that executes the users actions can be considered programmable. Figure 4(a) is a fragment of python code from Demšar et al. (2004) that performs 10-fold cross validation to test a naive Bayesian classifier and k-nearest neighbors algorithm on a voting data set. Using such scripts is only suitable for experts with enough programming skills and does not allow for visual exploration and manipulation of the data.

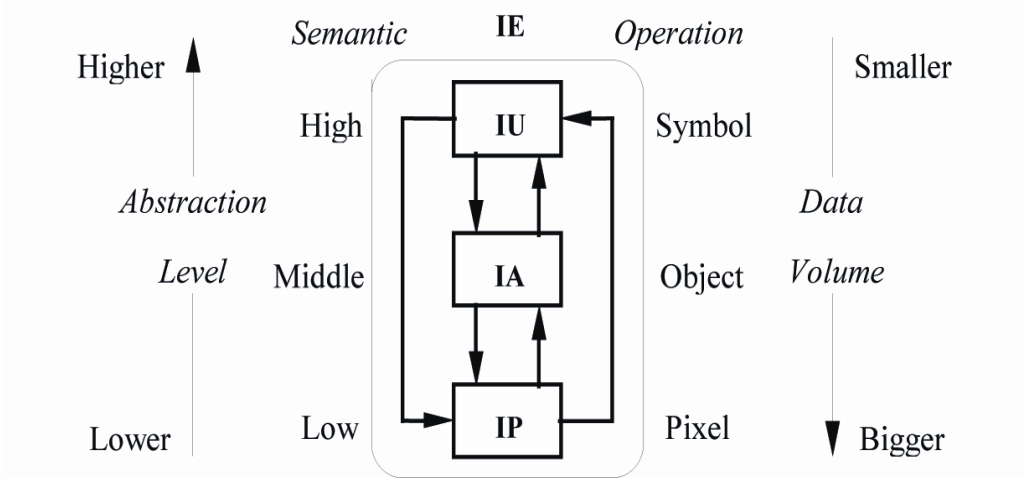


Figure 2: The three layers of image engineering (Zhang, 2006)

There are significant advantages to supplying programming capabilities in the user interfaces of a wide variety of programs (Shneiderman, 1983).

The user interface of an image processing environment is a key aspect of the proper functioning of an environment. A good user interface can significantly reduce the development effort of new image processing applications. The user interface also determines the usability of the environment to the various classes of users (Koelma and Smeulders, 1994).

One approach is the use of graphics as the programming language, or visual programming. By connecting the graphic components a user can visualise the data and the path of the data. This provides a clear expression of the flow of control in an application. The data flow metaphor is the best choice for the visual programming interface in image processing (Koelma and Smeulders, 1994). Using a visual environment provides a higher-level description of the desired actions. This makes the programming task easier even for professional programmers (Myers, 1992).

The Image Engineering system developed in this paper takes this approach. Using a visual environment allows for fast prototyping, interactive exploration of algorithms and workflow, and development of applications. The system presented exhibits characteristics and provides the benefits of visual programming, program visualisation (Myers, 1990), direct manipulation (Shneiderman, 1983), program-by-example, program-with-example (Lieberman, 2000), and demonstration interfaces (Myers, 1992).

Figure 3(a) is a fragment of python code that creates a Histogram of Gradients (HoG) (Dalal and Triggs, 2005) feature vector using the python image processing library scikit-image (van der Walt et al., 09). Figure 3(b) show the equivalent schema on the Orange canvas. Figure 3(c) opens the HoG widget showing visually options and response images. What is interesting, is the widget can process multiple images on the input stream without the user needing to understand loop constructs, unlike the code fragment that requires additional code to perform the same task. This is common for most widgets developed.

3 Underlying Technologies

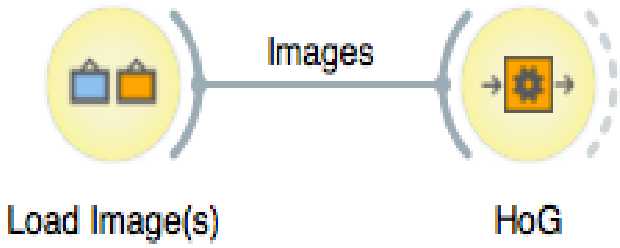
The system was built using Orange an open source toolkit for machine learning (Štajdohar and Demšar, 2013) onto which we have added image processing, visualisation and data acquisition methods recruiting functions from image processing libraries. In this paper functionality from scikit-image (van der Walt et al., 09) and OpenCV (Bradski, 2000), Scipy (Jones et al., 01) and Python (Van Rossum, 2003)have been used in the development of the system.

Orange is a general-purpose machine learning and data mining tool (Štajdohar and Demšar, 2013). It features a multi-layer architecture suitable for different kinds of users, from inexperienced data mining beginners to programmers who prefer to access the tool through its scripting interface.

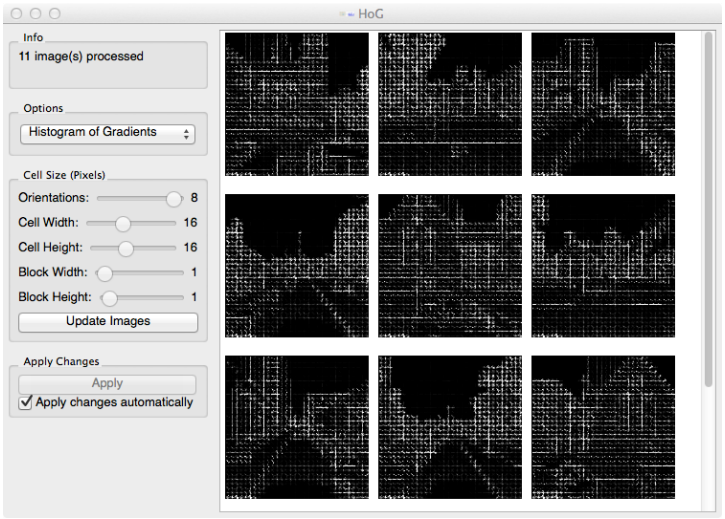
Orange Canvas provides a graphical interface for these functions. Its basic ingredients are widgets. Each widget performs a basic task, such as reading the data from a file in one of the supported formats or from a data base, showing the data in tabular form, plotting histograms and scatter plots, constructing various models and testing them, clustering the data, and so on.

```
from skimage.feature import hog
from skimage import data, color, exposure

image = color.rgb2gray(data.lena())

desc, hog_image = hog(image, orientations=8,
                      pixels_per_cell=(16, 16),
                      cells_per_block=(1, 1),
                      visualise=True)

# Process/save descriptor, perhaps display HoG image
# ....
```

(a)



(b)



(c)

Figure 3: (a) Fragment of python code calculating the HoG descriptor for an image (b) Orange schema calculating HoG descriptor for an image (c) Widget displaying HoG images a parameters of library API. Notice that the widget automatically process multiple images on the input stream.

The widgets expose the parameters of the underlying API. The advantage of widgets is in their modularity. Widgets can be connected through channels and communicate with each other by sending and receiving data. The output of one widget is used as an input for one or several other subsequent widgets. Communication channels are typed (i.e. the data type is determined to be integer, text, table, etc.) and the system establishes the proper type of data connections automatically. This property relieves the user from the need to design data structure, which is one of the greatest obstacles for lay users (Myers, 1990) (Olson et al., 1987).

A collection of widgets and their communication channels is called a schema, which is essentially a program designed by the user for a specific data analysis task. The programming process creates a schema with widgets and their connections is done visually through an easy-to-use graphic interface. Schemas can be saved and compiled into executable scripts for later reuse. The power of Orange Canvas is its interactivity. Any change in a single widget for example, loading another data set, changing the filter, modifying the logistic regression parameters can instantly propagates down the scheme.

4 Image Engineering Extension

Orange does not provide any image processing facilities but does allow you to develop your own widgets, extend scripting interface or even create your own self-contained add-ons, all seamlessly integrating with the rest of Orange, allowing components and code reuse. This allows anyone to leverage the power of the orange canvas to suit specific needs.

Figure 4(b) shows the Orange schema of the code fragment in Figure 4(a). Figure 4(c) shows the details of the *Test Learners* widget and the parameters that can be explored. Notice that unlike the code fragment details about the Area Under the Curve and Classification Accuracy and various other metrics available to the user.

The Image Engineering extension developed provides widgets designed to process image streams. The extension provides a complete workflow from image stream, to image processing, to feature vector, to machine learning, to classifier, to application. Some bespoke widgets were developed for functionality not provided by image processing libraries. For example in our systems, widgets were developed to calculate curvature, surface normals etc. from range images and statistical description of images.

Machine learning algorithms in Orange require a numerical representation of objects. If machine learning is to be performed then a widget that outputs numerical values need to be placed in the workflow as an interface between the machine learning and image processing data flow. When representing images as numerical values, the values might correspond to the pixels of an image, or the frequency of edge pixels, or perhaps a statistical description of an image region depending on the widget used. Widgets developed generally output a response image and numerical representations. Table 1 provides an overview of some of the widgets that have been developed.

5 Case Studies

Recognising objects in a scene is a primary goal of computer vision. The difficulty of this tasks depends on several factors such as: the number of objects, the complexity of the object etc. The appropriate techniques for object recognition depend on the difficulty of the task. One approach is to identify image patches, or regions of interest, that may contain objects of interest and then focus later processing, potentially computationally intensive, these regions of interest. Using a classifier, built prior, process the image patches to verify or confirm that the object of interest if present. To demonstrate the utility of the Image Engineering extension developed three case studies are provided demonstrating the above approach.

The case studies are based on the task of asset management of street furniture in urban transport corridors using the co-registered imagery and depth data. They have been simplified to focus on traffic lights but should be obvious to the reader how they could be extend to manage different assets.

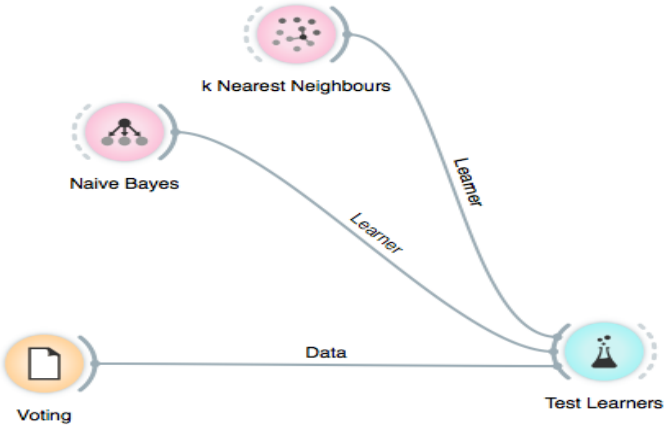
In each case study, no knowledge of how to program is required and a basic understanding of image processing is assumed. Workflows can be discovered through experimentation. Simple dragging, dropping and connecting widgets is all that is required. In one example the output from one case study is reused in another.

5.1 Locating Interesting Regions

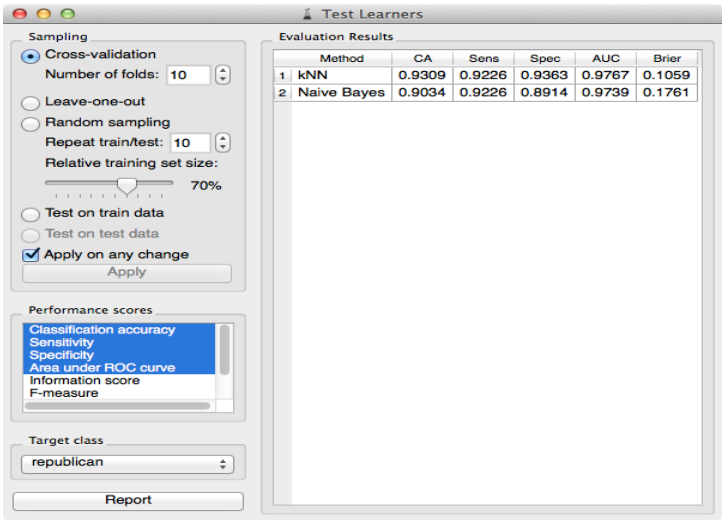
This case study uses AAM depth maps to identify interesting regions in co-registered imagery. In a scene objects exist at different depths. The intuition is that local peaks in a depth map indicate that an object is closer to the

```
from orange import BayesLearner, kNNLearner, ExampleTable
from orngTest import crossValidation
from orngStat import computrCDT

# set up the learners
bayes = BayesLearner(name='naive bayes')
knn = kNNLearner(name='knn')
learners = [bayes, knn]

# compute accuracies on data
data = ExampleTable("voting")
results = crossValidation(learners, data, folds=10)
cdt = computeCDT(results)

# output the results
# ...
```

(a)



(b)



(c)

Figure 4: (a) Fragment of python code performing 10-fold cross validation to test a nave Bayesian classifier and k-nearest neighbors algorithm on a voting data set. (B) Orange canvas schema performing 10-fold cross validation to test a nave Bayesian classifier and k-nearest neighbors algorithm on a voting data set. (c) Test Learners window exposing the parameters of the machine learning library

Table 1: A subset of widgets and associated functionality provided by image acquisition, processing and analysis extension.

Widget	Functionality
Load Image(s)	Reads image(s) from file system.
Test Image	Load one of many test images.
Earthmine	Connects to Earthmine server. Output imagery and depth maps
Channel	Select a image channel e.g. R,G,B,H,S,V. Output selected channel as an image.
Colour Space	Convert to a different colour space e.g. RGB to HSV. Output converted image.
Contours	Find the contours of an image
Depth	Find the surface normal, depth difference, curvature of the image. Output response image and histogram of response.
Edges	Compute edge map. Edge detectors available Canny (Canny, 1986), Sobel. Output edge map and edge density measure
Filter	Apply either Harr, Daubechilies or Gabor filter (Lee, 1996) to the input. Output response image and mean and variance of the filtered response.
Fourier	Convert to/from Fourier space. Output image or Fourier space image.
HoG	Calculate HoG Descriptor. Output HoG image and descriptor
Image Operations	Find Region Min/Max, Produce Histogram, Histogram Equalisation.
Interest Points	Find Harris (Harris and Stephens, 1988), SURF (Bay et al., 2008), SIFT (Lowe, 2004), or FASTER (Rosten et al., 2010) interest points. Outputs response image and the density of the interest points within the input region.
Local Binary Pattern	Calculate the LBP (Zhao and Pietikainen, 2006) of the input. Output image, mean and variance
Morphology	Apply, Dilate, Erosion, Opening, Closing, Skeleton operations on the input. Output response to morphological operation.
Salience	Calculate a saliency map using Frequency Tuned (Achanta et al., 2009), Edge based (Rosin, 2009) or Luminance and colour (Achanta et al., 2008) algorithms. Output saliency image.
Select	Select regions in an image or grid image into patches. Outputs set of regions
Segment	Segment the input using Quick Shift (Vedaldi and Soatto, 2008), SLIC (Achanta et al., 2010), and graph based (Felzenszwalb and Huttenlocher, 2004) algorithms. Output segmented image and segment labels.
Smooth	Apply either Gaussian or median filter to input image. Output smoothed image.
Statistics	Calculate and output the mean, standard deviation, skew, energy, entropy of the input image.
Threshold	Threshold image using either OTSU (Otsu, 1979), adaptive, median or user defined value. Output binary thresholded image.
Transform	Rotate, Resize or apply some other transformation. Output transformed image.

camera than its surroundings. The following method segments out things that are closer than their surroundings. A detailed description of the process can be found in Borck et al. (2014)

First a histogram equalization algorithm applied to the depth map. Then a maximal filter is used so local peaks stand out. The image is then dilated to expand the size of the local peaks. To remove the background, which essentially is unaltered in the process, the dilated image is subtracted from the output of the maximal filter. This creates an image that contains only local peaks. This local peak image is then thresholded and bounding boxes for each thresholded segment are determined. These bounding boxes are the hypothesised interesting regions and overlaid on the co-registered imagery to visually evaluate the process. For each step a widget is used to process the image stream. Figure 5 show the schema for detecting interesting regions as well as widgets for loading, dilation, background subtraction and the overlay of bounding boxes. Observe that the workflow is capable of processing multiple images without the user needing to understand and complex programming constructs such as looping or data structures.

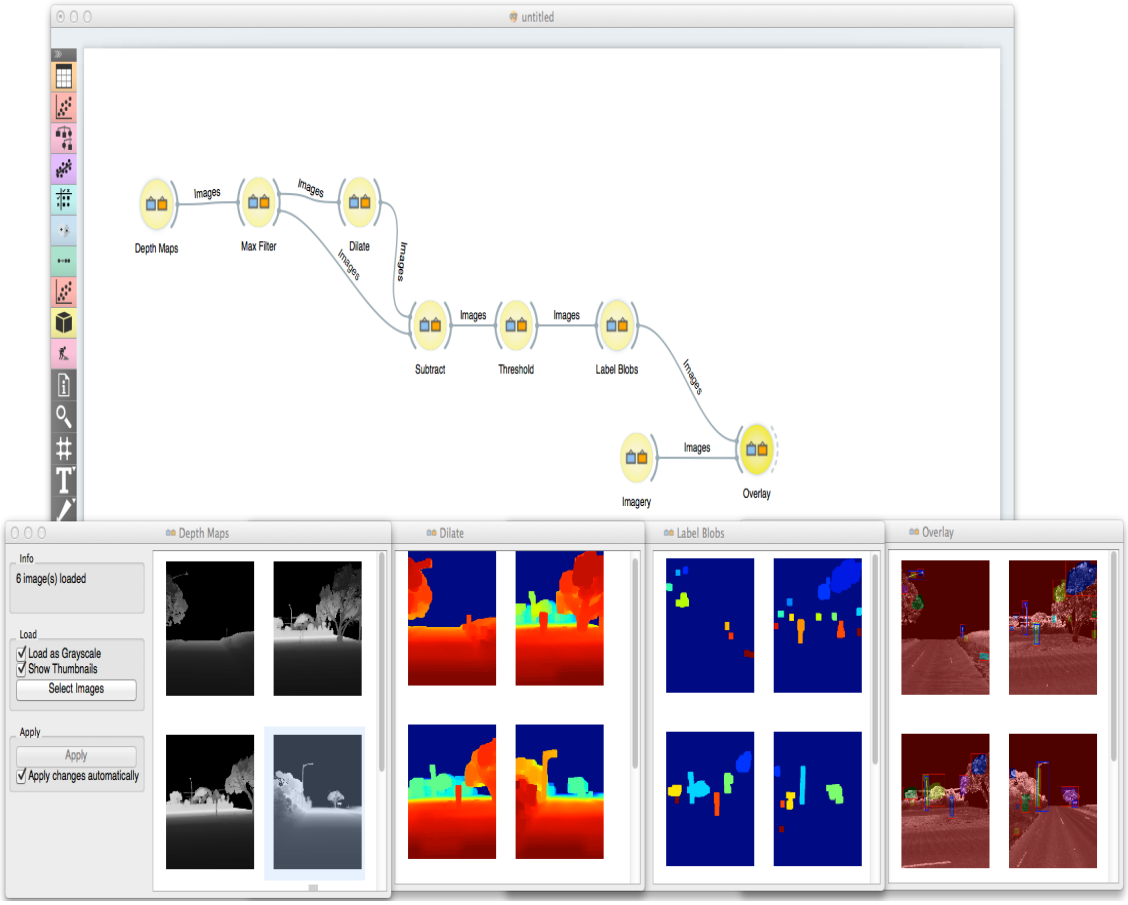


Figure 5: Schema for detecting interesting regions and widget output for loading, dilation, background subtraction and the overlay of bounding boxes.

5.2 Asset Detection

The process of using positive and negative examples with machine learning algorithms to build a object detection system is well defined in the computer vision literature. The real question is which combination of features and which machine learning algorithm is fit for the purpose. This case study describes a workflow to assess image and depth features, machine learning algorithms and infer an optimal set of features and a classifier to produce an traffic light detection system. Imagery and depth data were used from the Earthmine system. This case study implements a simplified workflow of a more complex multi class example as described in Borck et al. (2014).

The optimal classifier is determined by examining the graphs, confusion matrix, classifier metrics and other visualisations provide by the widgets use in the workflow. Once an optimal classifier has been identified it can be

saved and used in other workflows. In this case the Support Vector Machine (SVM) was the optimal classifier and is connected to the *Save Classifier* widget. Observe the two paths for the positive (Traffic Lights) and negative (Background) training examples. The feature vector is built using the HoG, Gabor and Statistics widgets. Also, since the output of the Gabor widget is an image, this need to be transformed into a descriptor. In this case a statistical summary, mean and variance, is used as the numerical representation of the Gabor feature. Notice in the schema where the output of the Gabor widget is connected to the input of the statistics widget. The HoG widget has two outputs, a descriptor and a HoG image. The HoG image allows you to visualise the feature and the descriptor is included in the feature vector. Figure 6(a) is the schema training a classifier to detect traffic lights from background. Canvas widgets can be opened by double clicking on the widget. Figure 6(b) shows open widgets for the Traffic light images, HoG feature, and the data table where each row is a feature vector. This is an example of visually exploring the workflow.

5.3 Asset Verification

This case study demonstrates a simple asset verification system and is a simplified version of a prototyped developed for a government department. Using the traffic light detection system developed earlier it is possible to label an image patch as either a traffic light or background. The Earthmine widget allows access you to specify a location. By using the Earthmine widget known locations were visited and traffic lights were extracted by manually by drawing bounding boxes around the traffic lights. A feature vector of each extraction region was built using the Hog, Gabor and Statistics widgets. The classifier developed in Section 5.2 was loaded and the newly created feature vector of unseen traffic lights were connected to the prediction widget. The prediction widget labels each feature vector as being either a traffic light or background. Figure 7 show the schema of the asset verification system.

6 Future Work

The extension developed is immediately useful for processing imagery and co-registered depth maps. Work is under way to include functionality provided by the Point Cloud Library (Rusu and Cousins, 2011). This would then provide a visual environment for the direct processing of point clouds.

There is an obvious overhead of using the visual environment, from maintaining internal data structures to updating visualisations. Many of the tasks in a schema could be executed in parallel. The functions could be distributed over several processors. The scheduling overhead would be small compared to the computational complexity of of typical image processing functions.

Widget functionality is influenced by the conceptual grouping of functions from the underlying library. This may not be the most suitable mapping onto higher level tasks performed by the widgets. Icons are used to express meaning graphically. The design of good icons is essential. Research into appropriate conceptual mapping of task and suitable expressive icons needs to be conducted. The cognitive dimensions framework of Green and Petre (1996) provides an evaluation technique that is easy to understand and quick to use and specific to visual environments and may be suitable for this purpose.

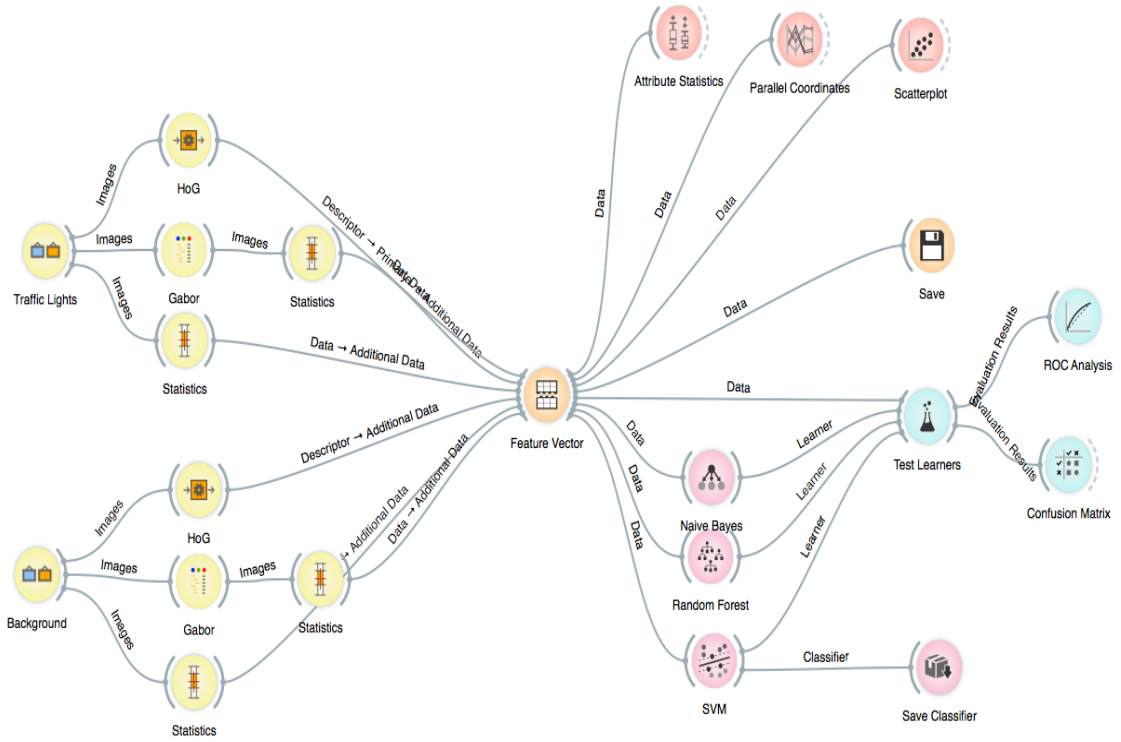
Initial feedback on prototypes developed has been positive and the perceived benefit from the user is high but a more detailed investigation into tangible benefits need to be conducted.

7 Conclusions

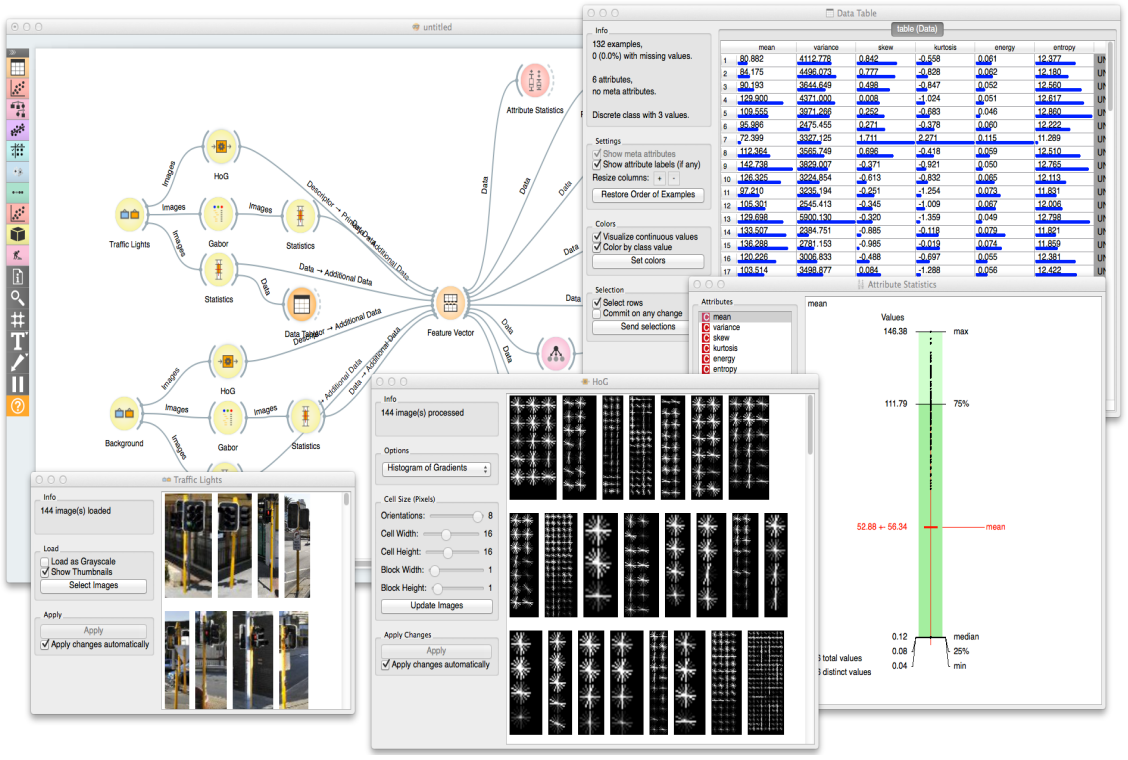
An Image Engineering extension to the Orange data analysis framework was developed. This extensions includes widgets for image acquisition, image processing and image analysis. The extension is modular and gives a non programmer user the opportunity to develop powerful image processing applications in a two-dimensional, data flow metaphor visual environment. This allows for fast prototyping and interactive exploration of algorithms and workflows without the need to learn a textual programming language. To demonstrate the utility of the system three case studies were presented. Imagery and depth data of urban transport corridors were used from the Earthmine dataset and laser ranging system.

Acknowledgment

This work is supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealths Cooperative Research Centres Programme. It provides PhD scholarship for Michael Borck and partially funds Professor Geoff West's position. The authors would like to thank John



(a)



(b)

Figure 6: (a) Example of schema training a classifier to detect traffic lights from background. (b) Example of exploring a workflow with widgets *Load Image(s)*, *HoG*, *Attribute Statistics* and *Feature Vector* opened to investigate and experiment with the workflow.

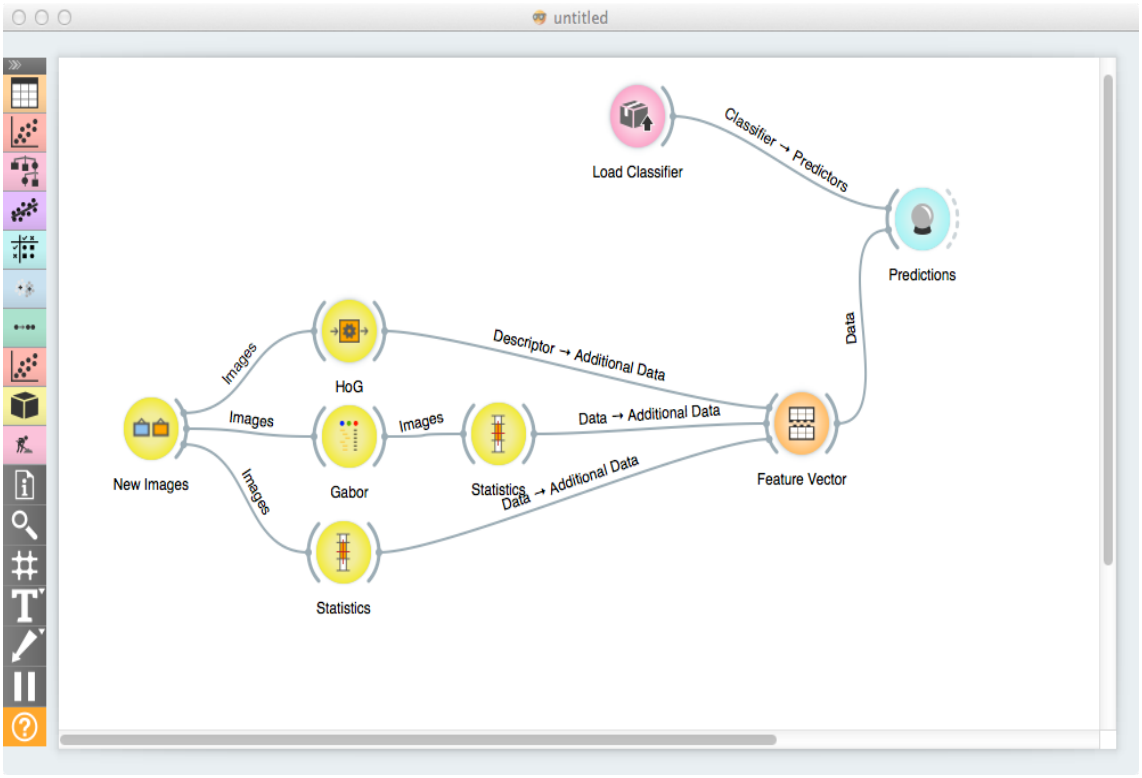


Figure 7: Schema for verifying images. First the feature vector is built then using a trained classifier predict if image regions is a traffic light.

Ristevski and Anthony Fassero from Earthmine and Landgate, WA for making available the dataset used in this work.

References

Achanta, R., F. Estrada, P. Wils, and S. Süsstrunk (2008). Salient region detection and segmentation. *Computer Vision Systems 5008*, 66–75.

Achanta, R., S. Hemami, F. Estrada, and S. Susstrunk (2009, june). Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1597 –1604.

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk (2010). Slic superpixels. *École Polytechnique Fédéral de Lausssanne (EPFL), Tech. Rep 149300*.

Bay, H., A. Ess, T. Tuytelaars, and L. V. Gool (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding 110*(3), 346 – 359.

Bentrad, S., D. Meslati, et al. (2011). Visual programming and program visualization-towards an ideal visual software engineering system. *ACEEE International Journal on Information Technology 1*(3).

Borck, M., R. Palmer, G. West, and T. Tan (2014). Using depth maps to find interesting regions. *to appear in proceedings of 2014 IEE Region 10 Technical Symposium*.

Borck, M., G. West, and T. Tan (2014). Use of multiple low level features to find interesting regions. *to appear in proceedings of 3rd Internation Conference on Pattern Recognition Aplications and Methods 2014*.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Canny, J. (1986, November). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 8*(6), 679–698.

Dalal, N. and B. Triggs (2005, June). Histograms of oriented gradients for human detection. In C. Schmid, S. Soatto, and C. Tomasi (Eds.), *International Conference on Computer Vision & Pattern Recognition*, Volume 2, pp. 886–893.

Demšar, J., B. Zupan, G. Leban, and T. Curk (2004). Orange: From experimental machine learning to interactive data mining. In J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi (Eds.), *Knowledge Discovery in Databases: PKDD 2004*, pp. 537–539. Springer.

Felzenszwalb, P. F. and D. P. Huttenlocher (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision 59*(2), 167–181.

Green, T. R. G. and M. Petre (1996). Usability analysis of visual programming environments: a 'cognitive dimensions' framework. *Journal of Visual Languages & Computing 7*(2), 131–174.

Harris, C. and M. Stephens (1988). A combined corner and edge detector. In *Alvey vision conference*, Volume 15, pp. 50. Manchester, UK.

Jones, E., T. Oliphant, P. Peterson, et al. (2001–). SciPy: Open source scientific tools for Python.

Koelma, D. and A. Smeulders (1994). A visual programming interface for an image processing environment. *Pattern Recognition Letters 15*(11), 1099–1109.

Lee, T. S. (1996). Image representation using 2d gabor wavelets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 18*(10), 959–971.

Lieberman, H. (2000). *Your wish is my command: Giving users the power to instruct their software.* Morgen Kaufmann.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*(2), 91–110.

Myers, B. (1992). Demonstrational interfaces: A step beyond direct manipulation. *Computer 25*(8), 61–73.

Myers, B. A. (1990). Taxonomies of visual programming and program visualization. *Journal of Visual Languages & Computing 1*(1), 97–123.

Olson, G. M., S. B. Sheppard, and E. Soloway (1987). *Empirical studies of programmers: second workshop*, Volume 2. Intellect Books.

Otsu, N. (1979, January). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics 9*(1), 62–66.

Rosin, P. L. (2009). A simple method for detecting salient regions. *Pattern Recognition 42*(11), 2363 – 2371.

Rosten, E., R. Porter, and T. Drummond (2010, Jan). Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 32*(1), 105 –119.

Rusu, R. B. and S. Cousins (2011). 3d is here: Point cloud library (pcl). *Library*, unknown.

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *Computer 16*(8), 57–69.

Štajdohar, M. and J. Demšar (2013, 4). Interactive network exploration with orange. *Journal of Statistical Software 53*(6), 1–24.

van der Walt, S. et al. (2009–). scikit-image: Image processing in python.

Van Rossum, G. (2003, September). *The Python Language Reference Manual.* Network Theory Ltd.

Vedaldi, A. and S. Soatto (2008). Quick shift and kernel methods for mode seeking. In *Computer Vision–ECCV 2008*, pp. 705–718. Springer.

Zhang, Y.-J. (2006). An overview of image and video segmentation in the last 40 years. *Advances in Image and Video Segmentation*, 1–15.

Zhao, G. and M. Pietikainen (2006). Local binary pattern descriptors for dynamic texture recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Volume 2, pp. 211–214. IEEE.

A Vector Agent Approach to Extract the Boundaries of Real-World Phenomena from Satellite Images

Kambiz Borna
School of Surveying
University of Otago
PO Box 56, Dunedin, NZ
kambiz.borna@ otago.ac.nz

Antoni Moore
School of Surveying
University of Otago
PO Box 56, Dunedin, NZ
tony.moore@otago.ac.nz

Pascal Sirguey
School of Surveying
University of Otago
PO Box 56, Dunedin, NZ
pascal.sirguey@otago.ac.nz

Abstract

This paper explores the application of vector agents (VA), a geometry-led type of computational agent, to extract the boundaries and classify real-world objects from satellite images. This method has been successfully implemented and tested on a multi-spectral satellite image to extract a set of objects with real-world counterparts. Comparison to the outcome of an Object-Based Image Analysis at a single scale of segmentation has also been made. The results of the presented approach enables real-world phenomena to be modelled with geographical objects derived intelligently from images. These objects have geometric, state and neighbourhood behaviours allowing them to iterate towards realistic output boundaries and robust classification in the simulation process.

1 Introduction

In the process of determining real world phenomena in remotely sensed images (e.g. an agricultural parcel or a building), the object boundaries are often interpreted as a set of irregular polygons. These boundaries enclose homogeneous areas and are collectively able to convert a spatial phenomenon, defined in a continuous environment (e.g. the remotely sensed image), into a set of objects in a discrete space (e.g. the classified image). However, from a geographic object-based image analysis (GOBIA) perspective, it may not be easy to determine these regions precisely. For one, they are extracted without a direct relationship with real-world objects (Benz et al., 2004) in a sequential process of segmentation and subsequent classification. Hence, segmented objects cannot adjust their geometry once they are classified. They are also extracted based on a set of crisp experimental parameters such as scale, colour, and shape (Hay et al., 2005). In turn, existing approaches are not capable of modeling vague/fuzzy/diffuse boundaries and/or complex geometric behaviours of such boundaries in a dynamic manner on the basis of the real-world phenomena that such boundaries are meant to represent.

To tackle these limitations, a vector agent (VA) approach that encapsulates spatial reasoning capabilities (e.g. use of orientation and size knowledge), will be investigated in order to extract the boundary of real-world phenomena from a satellite image. Each image object in the model can be interpreted as a level of abstraction of a real-world object in subsequent iterations of an evolving process from pixel to real-world object. In this regard, this VA implementation moves toward a classification solution. Hence, the modelled objects are a set of vector agents that evolve dynamically in a spatio-temporal environment (Figure 1), and once evolved are subject to a single scale. These VAs provide a dynamic geometry and impose the properties of real-world and simulation environment for the modelled objects. This is in contrast to previous VA implementations (Hammam et al., 2007; Moore, 2011) in which VAs manipulated their own geometry according to fractals. All VA realisations can be placed within the Geographic Automata (GA) framework of Torrens and Benenson (2005).

2 Implementation and Outcome

The initial scenario is formed based on three different classes: agriculture, bare soil and water, found in a subset of an IKONOS image (Figure 2a). First, each desired object is automatically initialized in space. This is done by using

the feature space defined by the reflectance information from each of the four spectral bands of the sample image. The feature space exhibits sets of pixels as contiguous clusters with similar spectral reflectance properties. A candidate pixel, which has a minimum spectral distance from the mean of each cluster, is extracted for each class. In addition, a spectral threshold is defined for each class as the maximum Euclidean distance for candidate pixel to belong to the modelled class object. This is equivalent to a minimum Euclidean distance classifier. After initializing, the desired object repeatedly seeks to find such candidate pixels in image space, in an effort to define the full extent of its VA boundary. Also, the other agents are automatically initialized in space, in the same way. Once an agent finds a neighbor pixel that meets the class criteria, this triggers the evolving process of the VA controlled by a set of rules. These rules are defined by structure and geometry primitives to adjust the boundaries of the modelled objects.

An example of the result of this process is shown in Figure 2(b), while Figure 2(c) illustrates the outcome of a single-scale segmentation of the same subset completed with the Trimble eCognition OBIA classification software. The results are promising and indicate that the VA object classification method is capable of evolving towards a modelling of the image in readily classified objects, while traditional OBIA requires the successive and somewhat independents process of segmentation then classification.

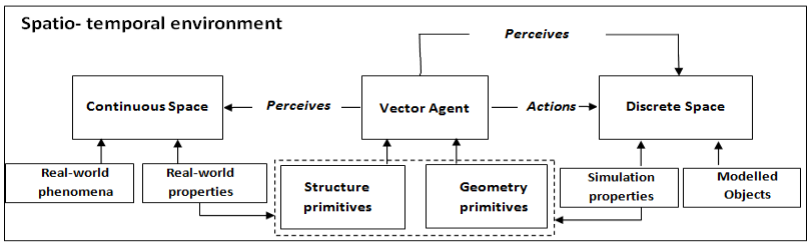


Figure 1: Schematic of the processes involved by the proposed approach.

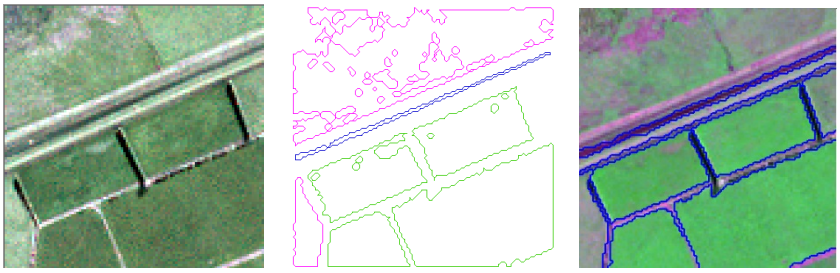


Figure 2: (a). Multispectral IKONOS image (Dunedin, 2005), (b). The boundaries of the modelled objects based on VA classification, (c). OBIA segmentation (scale = 45, colour = 0.4) for agriculture, bare soil and water classes.

3 Conclusions

This study proposes a new intelligent vector agent approach to extract the boundaries of real-world objects in the image space through spatial reasoning. This approach enables the desired objects to extract their own boundaries based on the existing information for a great variety and complexity of real-world objects. In this study, the different behaviours of desired objects, such as initialising, expanding and shrinking, have been individually investigated. For future development, the full scope of the geometry, state and neighbourhood interactions rules will be explored.

References

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M.2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J.*

Hammam, Y., Moore, A., and Whigham, P.2007. The dynamic geometry of Geographical Vector Agents,*Computers, Environment and Urban Systems*, vol.31, no.5, pp. 502-519.

Hay, G.J., Castilla, G., Wulder, M.A. and Ruiz, J.R. 2005.An automated object-based approach for the multiscale image segmentation of forest scenes. *Int. J. of Applied Earth Observation and Geoinformation*, 7, 339–359.

Moore, A. 2011. Geographical Vector Agent Based Simulation for Agricultural Land Use Modelling, in Marceau,D. and Benenson, I. (Eds) *Advanced GeoSimulation Models*.

Torrens, P., and Benenson, I. 2005. Geographic Automata Systems, *International Journal of Geographic Information Science*, vol. 10, no.4, pp.385-412.

Estimating Urban Ultrafine Particle Distributions with Gaussian Process Models

Jason Jingshi Li
College of Engineering and Computer Science.
The Australian National University, Canberra, ACT 0200, Australia
jason.li@anu.edu.au

Arnaud Jutzeler
Artificial Intelligence Laboratory
EPFL, Lausanne, 1025, Switzerland
arnaud.jutzeler@epfl.ch

Boi Faltings
Artificial Intelligence Laboratory
EPFL, Lausanne, 1025, Switzerland
boi.faltings@epfl.ch

Abstract

Urban air pollution have a direct impact on public health. Ultrafine particles (UFPs) are ubiquitous in urban environments, but their distribution are highly variable. In this paper, we take data from mobile deployments in Zürich collected over one year with over 25 million measurements to build a high-resolution map estimating the UFP distribution. More specifically, we propose a new approach using a Gaussian Process (GP) to estimate the distribution of UFPs in the city of Zürich. We evaluate the prediction estimations against results derived from standard General Additive Models in Land Use Regression, and show that our method produces a good estimation for mapping the spatial distribution of UFPs in many timescales.

1 Introduction

Air pollution in urban environments have a direct impact on the health of the people. The World-Health-Organization (2011) estimated that over 1.3 million deaths per year world-wide are attributed to urban outdoor air pollution. Currently in most developed countries, a network of government-funded and operated static measurement stations continuously make highly reliable and accurate measurements on important air pollutants. However, the high cost of installation and maintenance of these stations limits the number of stations deployed in a given city. Consequently, only very limited information can be collected about the spatial distribution of air pollutants in the urban setting.

The OpenSense project, described in Aberer *et al.* (2010), is a multi-disciplinary project funded by the Swiss National Science Foundation to study mobile air quality monitoring and modelling in urban environments. It is deploying multiple mobile air quality monitoring stations on top of trams in the Swiss city of Zürich (Fig. 1), collecting measurements of ozone concentrations (O_3) and the counting of ultrafine particle (UFPs). To this date, it has publicly released over 25 million measurements over an urban area of 100 km^2. The data and their sensing methodology can be found in Li *et al.* (2012b) and Hasenfratz *et al.* (2014). These data form a sufficient basis to study the spatial variability of the pollutants in the urban environment.

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

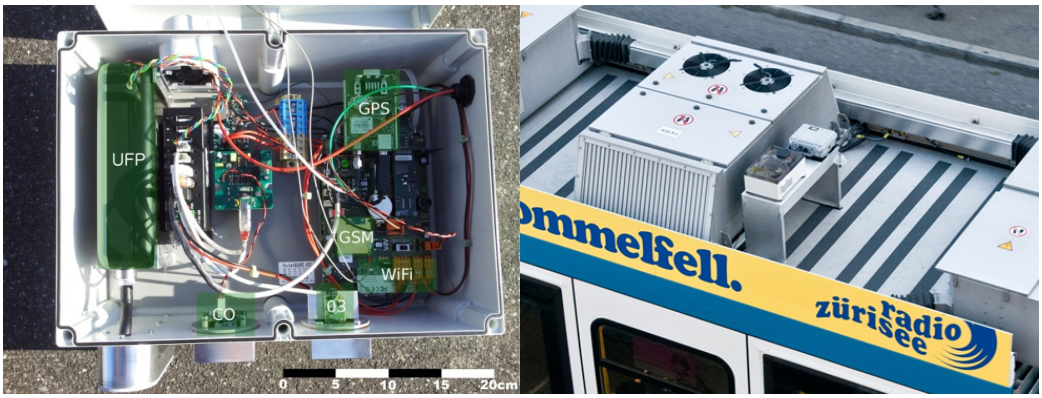


Figure 1: Left: Interior of an OpenSense sensing node; Right: deployment on a tram in Zürich

Traffic junctions, industrial installations and urban canyons all contribute to the high spatial and temporal variability of air pollution in urban areas. Small-scale spatial distribution of ambient air pollution have traditionally been studied with Land-Use-Regression (LUR) summarised in Hoek *et al.* (2008). It uses land-use and traffic characteristics of a particular grid region as explanatory variables to learn to estimate pollution concentrations under a Generalized Additive Model (GAM). The learnt model is then used to predict pollution levels for all locations with the available land-use information.

In this paper, we propose a novel approach of estimating urban ultrafine particle levels across different temporal aggregates from measurements collected from the trams. Similar to standard models in land use regression, it estimates the pollution levels within different grid-cells in the urban environment from a set of land-use features. Our model is based on constructing a Gaussian Process described in Rasmussen and Williams (2006), with additional consideration to spatial features in the covariance matrix. Following the practice in previous work of Hasenfratz *et al.* (2014), we evaluate the models (GAM, pure land use and mixed spatial-land-use) using standard random 10-fold cross validation.

The outline of this paper is as follows: we begin with a summary of the background to the paper: the data, the traditional models used in land-use regression, and introducing Gaussian Process Regression. We then introduce a new approach for estimating UFP levels, and evaluate it against the previous approach over a benchmark dataset.

2 Background

2.1 The Aggregate Datasets

The data were selected from UFP measurements collected on Zürich trams between April 2012 and March 2013 as part of the OpenSense project and the sensing methodology is described in Li *et al.* (2012b) and Hasenfratz *et al.* (2014). The data were partitioned into 13,200 grid cells of size 100m $\times$ 100m. The profile of a typical grid cell, such as the one containing Centralplatz in Zürich, is shown in Fig 2. We can see that instead of being fitted to a normal distribution (solid line shows the best-fit), the measurements fits much better as log-normal distribution (dotted line). This is consistent with literature on particle count concentrations in urban environments described in Mølgaard *et al.* (2012). The data were captured and transmitted in real time to a back-end server running Global Sensor Network (GSN) by Aberer *et al.* (2006), and removed to a local database to be preprocessed and aggregated before entered into the model.

Several preprocessing steps were used before the data were prepared for the model, including removal of measurements within the indoor tram depot, measurements with bad GPS data, and measurements with extraordinary high levels >100'000 particles per cm^3. These steps were described in detail in Hasenfratz *et al.* (2014), with the purpose of avoiding bias due to erroneous measurements.

We then aggregate the data within the different grid cells according to the different time windows, such as yearly, seasonal, monthly, biweekly, weekly, daily and half-daily. This is done to understand the trade-off between long and short term aggregate data. In order to evaluate and compare our results to previous work, we followed the convention of selecting only the 200 grid cells with the highest measurements count for the purpose of modelling and validation, as Hasenfratz *et al.* (2014) showed that the state-of-the-art models produced the

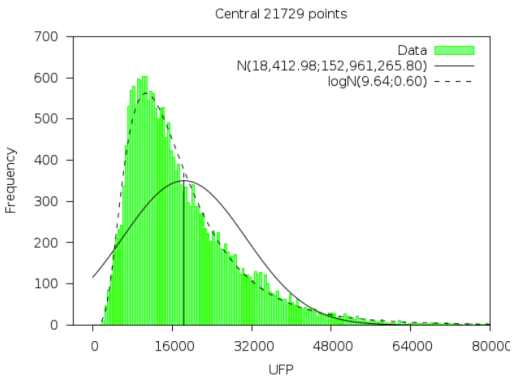


Figure 2: Distribution of UFP measurements collected in the grid cell near Centralplatz in Zürich during winter 2013. The black line shows the maximum likelihood estimated normal distribution whereas the dashed line shows the maximum likelihood estimated log-normal distribution.

most reliable predictions when only the top 200 grid cells with the highest measurement count are considered.

2.2 Land Use Regression

In literature, land-use regression models are used to assess intra-urban air pollution distributions, and a comprehensive review of these techniques can be found in Hoek *et al.* (2008). They typically combine monitoring of air pollution at 20-100 locations spread over the study area, and develop a model using predictor variables obtained through geographic information systems (GIS). The predictor variable generally include some traffic information, population density, designated land use and features of the landscape such as attitude and slope. Due to the cost of deployments, studies usually last 1-2 weeks in duration.

For particulate matter such as $PM_{2.5}$ and PM_{10} and UFPs, Generalized Additive Models (GAMs) have been used in land use regression to study their spatial and temporal variability. It typically use the following equation to model the relationship between the pollution level p and a set of explanatory variables $A_1, \dots A_n$.

$$ln(p) = a + s_1(A_1) + s_2(A_2) + \cdots + s_n(A_n) + \epsilon \tag{1}$$

where a is known as the intercept, ϵ the error term, and $s_1 \dots s_n$ are typically smooth regression splines with an upper limit of 3 on the degree of freedom. In this paper, we use the GAM data from Hasenfratz *et al.* (2014) as a benchmark to compare our model predictions.

2.3 Gaussian Process Regression

Also known as Kriging, Gaussian process regression (GPR) has been extensively used for decades in Geostatistics to model various spatial phenomena such as soil concentrations, weather-related events, etc., and in-depth overviews can be found in Cressie and Cassie (1993) and Rasmussen and Williams (2006). Similar to other non-parametric approaches, GPR does not require prior structural knowledge about the phenomenon. Indeed, the idea is precisely that structure is directly inferred from the data. Furthermore, GPR outputs statistical predictions and thus represents an adequate candidate to model phenomena that are inherently noisy and which one can only observe through noisy instruments. Recently it has been successfully applied in many machine learning tasks such as bioinformatics in Chu *et al.* (2005), sensor calibration in Monroy *et al.* (2012) and crowdsourcing Venanzi *et al.* (2013) It still represents a very active ongoing research area as seen in e.g. Bonilla *et al.* (2010), Cao *et al.* (2013) and Nguyen and Bonilla (2014). To allow the reader to have a better understanding of our models, in the following we will provide a very brief technical overview of Gaussian Process Regression.

A Gaussian Process (GP) is used to model a phenomenon that takes place in a certain input space $\mathcal{X} \subseteq \mathbb{R}^d$. We formally write $f(\mathbf{x})$ where $\mathbf{x} \in \mathcal{X}$ the function that models the phenomenon. The general idea is to assume that the function $f(\mathbf{x})$ is a specific realization of a prior Gaussian Process $\mathcal{GP}$, which is the generalization of a multivariate normal distribution to an infinity of random variables, that is to say a distribution over whole functions. A GP is fully defined by its mean function $m(\mathbf{x})$ and its covariance function $k(\mathbf{x}, \mathbf{x}')$ (also called

kernel) that are the generalization of the mean vector, respectively the covariance matrix of a multivariate normal distribution.

Regression with a GP is typically performed as follows. In general, we can only make from the phenomenon noisy observations $y_i = f(\mathbf{x_i}) + \epsilon_i$ where the additive noise ϵ is also assumed to be Gaussian $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. By using the marginalization property of GPs and the additive nature of the noise ϵ we know the joint distribution of the observations $\mathbf{y}$ at locations X and the values $\mathbf{f}_*$ at test points X_* to be:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} k(X, X + \sigma_n^2 I) & k(X, X_*) \\ k(X_*, X) & k(X_*, X_*) \end{bmatrix} \right) \tag{2}$$

Then for those test points X_* the regression consists in computing the predictive distribution $p(\mathbf{f}_*|\mathbf{y})$. Fortunately by the conditioning property of a joint multivariate Gaussian distribution this expression is tractable and even admit a closed formula. It results in another multivariate Gaussian distribution. For any single test points $\mathbf{x}_* \in X_*$ the predictive mean and variance are given by:

$$\bar{f}(\mathbf{x}_*) = m(\mathbf{x}_*) + k(\mathbf{x}_*, X)(k(X, X) + \sigma_n^2 I)^{-1}(\mathbf{y} - m(\mathbf{x})) \tag{3a}$$

$$\mathbb{V}[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, X)(k(X, X) + \sigma_n^2 I)^{-1} k(X, \mathbf{x}_*) \tag{3b}$$

The main challenge is to create and choose prior mean and covariance functions that carry adequate assumptions about the phenomenon. We describe in detail how we derived such functions in the following section.

3 Our Model

3.1 The Land-Use Model

Our first GP model uses only land-use variables as features to generate predictions on the mean UFP concentration measured by the sensors within the respective grid cells in the timeframe of the specified dataset. They follow from the features used in Hasenfratz *et al.* (2014). The model takes a vector $\mathbf{x_{LU}}$ containing the land-use variables values of a certain 100m × 100m grid cell as input. These land-use features were taken from the following sources:

- Swiss Federal Statistical Office
  - Population density, industry density, building heights, heating type, terrain elevation, terrain slope
- Canton of Zürich government
  - Average daily traffic volume
- *OpenStreetMaps.org*
  - Main road type, distance to next major road, distance to major traffic signal

As we wanted to start with no particular *a priori* structural knowledge, only very simple mean functions were tried such as the trivial fixed 0 function and a constant c. Deriving a suitable covariance function was, however, a bit more complex. Indeed, to be valid a covariance function must be positive definite. It is common practice to start from well-known parametrized families of positive definite functions and fit the parameters (that in the scope of GPR are called hyperparameters) using the data. All the covariance functions that were tried are stationary that is to say every points of the space shows the exact same covariance structure with its own surroundings or more formally we have $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$. Stationary covariance functions such as squared exponential, and various flavours of the Matérn class were tried, each one carrying different assumption about the smoothness of the process. Finally from preliminary tests results, a constant mean function and a squared exponential covariance function were selected. We note that the chosen covariance function was the one that carried the strongest smoothness assumptions. The prior GP is thus defined by:

$$m(\mathbf{x_{LU}}) = c \tag{4a}$$

$$k(\mathbf{x_{LU}}, \mathbf{x'_{LU}}) = \sigma_f^2 \exp\Big(-\frac{1}{2}(\mathbf{x_{LU}} - \mathbf{x'_{LU}})^\top M(\mathbf{x_{LU}} - \mathbf{x'_{LU}})\Big) \text{ where } M = \operatorname{diag}(\boldsymbol{\ell_{LU}})^{-2} \tag{4b}$$

The σ_f^2 is the magnitude hyperparameter, and the $\boldsymbol{\ell_{LU}}$ are the length-scale hyperparameters that determine the relevance of some or other land-use variables. To learn the values of all the hyperparameters $\theta = (c, \sigma_f^2, \boldsymbol{\ell_{LU}}, \sigma_n^2)$ one can either use optimization or sampling techniques. In our case, we used the standard approach that consists in optimizing the log marginal likelihood:

$$\log p(\mathbf{y}|X, \theta) = -\frac{1}{2}\mathbf{y}^\top (K + \sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2}\log 2\pi$$

Every evaluation of this expression takes $O(n^3)$ with n being the number of training points X. From then the evaluation of its derivatives with regards to hyperparameters takes $O(n^2)$ per hyperparameter.

3.2 The Mixed Spatial Land-Use Model

Even though the explanation of the phenomenon given by the land-use variables may already be quite good, it is very likely that part of it still elude us because of some contributions to the phenomenon that are badly or not at all reflected in the variables. To address this matter, we tried to incorporate geographical informations into the model with the hope that such missed contribution will at least be partly explained locally.

The problem with parametric models such as GAM is that we cannot easily add geographic informations into the model in a sensible way. For example if we naively add the longitude and latitude as covariates, we would be making very strong assumptions rather unrealistic.

However, with GPR (and this is why it has been extensively used in Geostatistics) it is natural to include such informations in the reasoning. This is done by including a consideration for geographical distance in the covariance function. We call our second model a mixed spatial-land-use model, which is a variant of the first one in which we added a term in the covariance structure. We also tried different isotropic kernels to be this additional term. From the preliminary experiments the following covariance function was selected:

$$k(\begin{bmatrix}\mathbf{x_{LU}}\\ \mathbf{x_S}\end{bmatrix}, \begin{bmatrix}\mathbf{x'_{LU}}\\ \mathbf{x'_S}\end{bmatrix}) = \sigma_{f_{LU}}^2 \exp\Big(-\frac{1}{2}(\mathbf{x_{LU}} - \mathbf{x'_{LU}})^\top M(\mathbf{x_{LU}} - \mathbf{x'_{LU}})\Big) + \sigma_{f_S}^2 \exp\Big(-\frac{\|\mathbf{x_S} - \mathbf{x'_S}\|}{\ell_S}\Big) \tag{5}$$

It is worth noting that it is the exponential function, the less smooth of the considered covariance functions, that was chosen to be the additional term in function of the geographical distance. The values of the hyperparameters $\theta = (c, \sigma_{f_{LU}}^2, \sigma_{f_S}^2, \ell_S, \boldsymbol{\ell_{LU}}, \sigma_n^2)$ were once again fixed using marginal likelihood maximization.

4 Evaluations

We implemented our own Java framework to perform GPR. However, the conjugate gradient optimizer, used to maximize the log marginal likelihood, was taken from the Matlab toolbox GPML v.2 (see Rasmussen and Nickisch (2010)) and translated in Java. Most of linear algebra operations were carried out using EJML[1] library. The experiments were conducted on a server with 64 AMD Opteron processing cores and 96 GB of RAM. In the experiments, we compared the following three different type of models on the UFP datasets described earlier.

1. **GAM** A General Additive Model from Hasenfratz *et al.* (2014);
2. **GP_LU** Our land-use only GP model;
3. **GP_LUXY** Our mixed spatial-land-use GP model.

From the benchmarking data supplied by Hasenfratz *et al.* (2014), we get 989 datasets comprise of 597 half-daily, 309 daily, 44 weekly, 23 biweekly, 11 monthly, 4 seasonally and a single yearly aggregated dataset from measurements taken from Zürich trams between April 2012 and March 2013. For each aforementioned type of model, we trained yearly to half-daily models to predict mean pollution level within grid cells (in particle count per cm^3). We evaluated the quality of the models predictions using standard randomised 10-fold cross validation. That is, for each dataset, we randomly partitioned the data into 10 equal parts, and iteratively we used 9 parts as training set of the model to generate predictions to be compared against the 1 remaining part.

Fig. 3 shows the satellite image of the urban area covered in the deployment, the output of the pollution map for the season of summer in 2012, and the comparison of the prediction against ground truth of the same season under random 10-fold cross validation. Fig.4 shows the scatter plots of model predictions against ground truth

[1] http://code.google.com/efficient-java-matric-library/

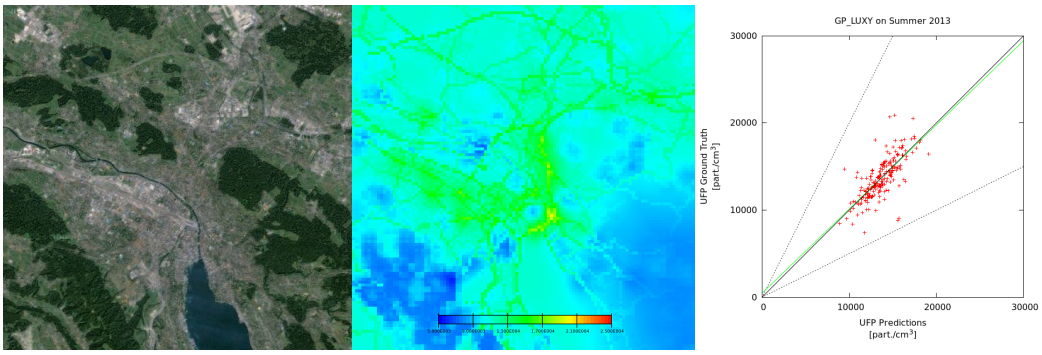


Figure 3: left: a satellite image of Zurich; centre: the predicted summer mean UFP level from mixed spatial-land-use GP model; and right: the scatter plot of the predictions from the same model against ground truth under random 10-fold cross validation.

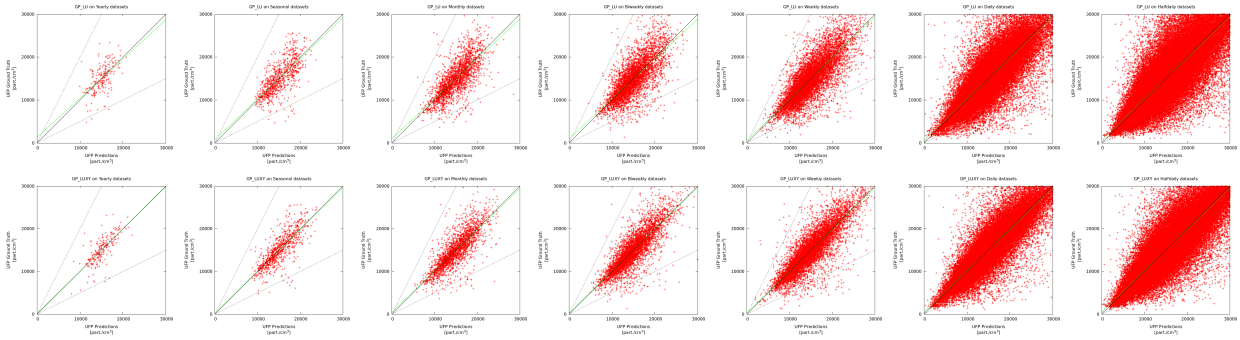


Figure 4: Scatter plot of prediction vs. ground truth after random 10-fold cross validation on yearly, seasonal, monthly, biweekly, weekly, daily and half-daily data. Top row: GP_LU model, bottom row: GP_LUXY model.

across all time scales, where all predictions of the same time scale are located on the same plot. It is worthy to note that similar to the previous model presented in Hasenfratz *et al.* (2014), our models also show little to no bias, as evident from the fact that across all cases the linear regression lines (in green) are very close to the optimal 1-to-1 lines. It indicates the absence of systematic model errors.

4.1 RMSE

First we compare the Root Mean Square Error (RMSE) of the predictions derived from the models under random 10 fold cross validation (Fig. 5). It is a standard metric of predictive power for measuring the accuracy of prediction models. It is obtained by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(p_i - g_i)^2}{N}} \tag{6}$$

where p_i denotes the i^{th} prediction, g_i the ground truth of the i^{th} prediction, and N the total number of predictions. In Fig. 5, the plot on the left displays the overall mean of the RMSE, while the box-plot on the right displays the minimum, lower quartile, median, upper quartile and maximum of the average RMSE of the whole 10-fold validation tests on all the datasets of the same time scale. The yearly data came from a single dataset, thus it is presented as a single value. It shows that as expected, the higher temporal resolution leads to higher uncertainty in the prediction, the GP models outperforms GAM across all temporal resolutions, and the mixed spatial-land-use model produced less error than the land-use only model.

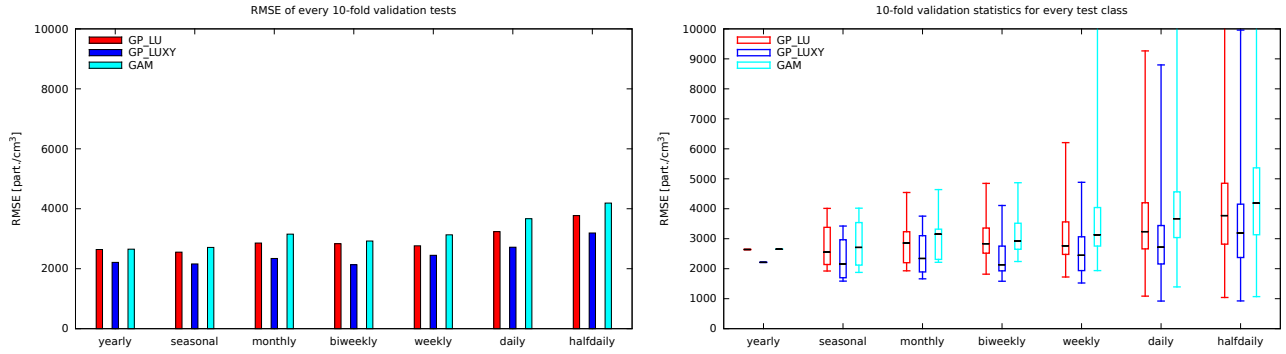


Figure 5: Mean (left) and distribution (right) RMSE of model predictions across all datasets in random 10 fold cross validation (the lower the better)

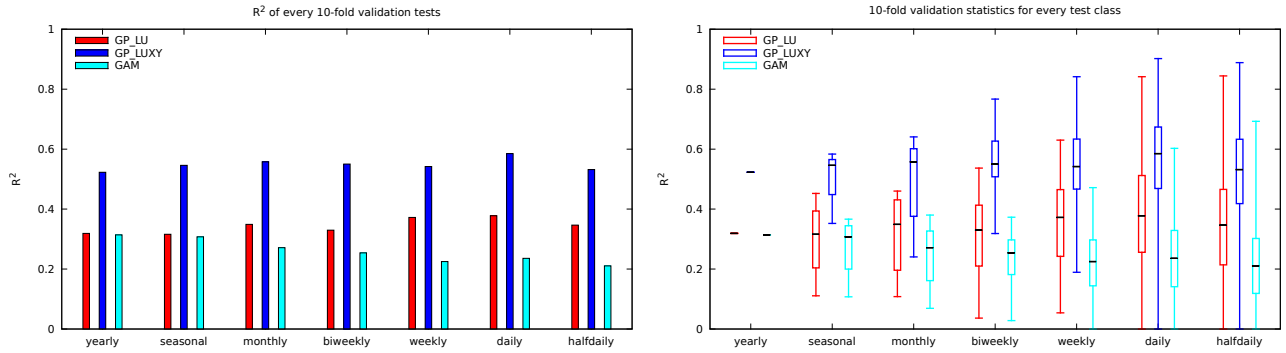


Figure 6: Mean (left) and distribution (right) R^2 score of model predictions across all datasets in random 10 fold cross validation (the higher the better)

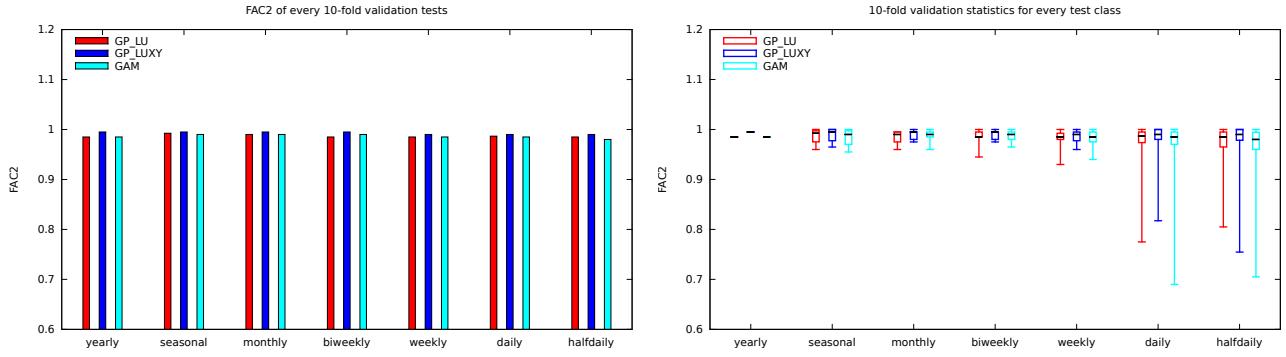


Figure 7: Mean (left) and distribution (right) of FAC2 score of model predictions across all datasets in random 10 fold cross validation (the higher the better)

4.2 R^2 Score

We then compare the R^2 coefficient, also known as the coefficient of determination of the model predictions (Fig. 6). It indicates how well the observed outcomes are replicated by the model predictions as the proportional variation of outcomes explained by the model. Its formula is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(p_i - g_i)^2}{\sum_{i=1}^{N}(g_i - \bar{g})} \tag{7}$$

where p_i denotes the i^{th} prediction, g_i the ground truth of the i^{th} prediction, and $\bar{g}$ is the mean of the ground truth. In Fig. 6 observe that the variance of the R^2 score also increases as the the time scale shrinks across all models. We see that the results of GP models in general have a higher R^2 than GAMs, and introducing the spatial covariance in the GP model also improves the R^2 score across all time scales.

4.3 FAC2 Score

Finally, we compare the $FAC2$ score of the model predictions (Fig. 7). It measures the fraction of data points that lie inside the factor of two area. It is a robust measure of prediction as it is not overly influenced by high and low outliers. It is derived by:

$$FAC2 : 0.5 \leq \frac{p_i}{g_i} \leq 2 \tag{8}$$

The box plots in Fig. 7 show the FAC2 distributions for all models across all temporal scales. We can see that they all have very high FAC2 values for yearly, seasonal, monthly, biweekly and weekly data. Daily and half daily predictions have lower FAC2 values, with GP models perform slightly better than GAM.

5 Conclusion and Future Work

We implemented two schemes based on Gaussian Process for estimating mean UFP concentrations in urban areas of Zürich, Switzerland. We show that they provide an alternative to GAM approaches in land-use regression, and there is a general trade off between the length of the time scale and the quality of the model predictions. We also show that across the timescales the proposed GP models presents an improvement on the current state of the art. The resulting maps may be useful for application such as assessing population exposure to air pollutants similar to that of Carroll *et al.* (1997), uncover areas of high air pollution for persons with allergies, or evaluate the trustworthiness of measurements contributed by a community of sensors as described in Li *et al.* (2012a) and Faltings *et al.* (2014).

Possible future work includes moving away from a grid-based model to make use of urban spatial features described in Li *et al.* (2012b), developing models that handles different aspects of sensor reliability and measurement bias; detecting and filtering spurious measurements, and combining meteorological information and real time data to produce the best real-time estimations for individual exposure analysis and route planning. Our approach based on Gaussian Process Regression is very general, and it is interesting to see if it can be generalised to particulate dispersion outside urban environments to applications such as bush-fire detection; and whether it can be applied to estimating other air-borne or water-borne pollutant dispersions.

Acknowledgements

We thank our collaborators at ETHZ David Hasenfratz and Olga Saukh for supplying the benchmarking data and model from their previous work. This work is supported by OpenSense project funded by NanoTera.ch, and the ARC Discovery Project (DP120103758) "Artificial Intelligence Meets Sensor Networks".

References

K. Aberer, M. Hauswirth, and A. Salehi. A middleware for fast and flexible sensor network deployment. In *VLDB*, 2006.

K. Aberer, S. Sathe, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings, and L. Thiele. OpenSense: Open community driven sensing of environment. In *ACM IWGS*, 2010.

Edwin V Bonilla, Shengbo Guo, and Scott Sanner. Gaussian process preference elicitation. In *NIPS*, pages 262–270, 2010.

Yanshuai Cao, Marcus A Brubaker, David Fleet, and Aaron Hertzmann. Efficient optimization for sparse gaussian process regression. In *NIPS*, pages 1097–1105, 2013. URL http://papers.nips.cc/paper/5087-efficient-optimization-for-sparse-gaussian-process-regression.pdf.

RJ Carroll, R Chen, EI George, TH Li, HJ Newton, H Schmiediche, and N Wang. Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association*, 92(438):392–404, 1997.

Wei Chu, Zoubin Ghahramani, Francesco Falciani, and David L Wild. Biomarker discovery in microarray gene expression data with gaussian processes. *Bioinformatics*, 21(16):3385–3393, 2005.

Noel AC Cressie and Noel A Cassie. *Statistics for spatial data*, volume 900. Wiley New York, 1993.

Boi Faltings, Jason Jingshi Li, and Radu Jurca. Incentive mechanisms for community sensing. *IEEE Transactions on Computers*, 63(1):115–128, 2014.

D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, and L. Thiele. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *Proceedings of the 12th International Conference on Pervasive Computing and Communications (PerCom'14)*, 2014.

Gerard Hoek, Rob Beelen, Kees de Hoogh, Danielle Vienneau, John Gulliver, Paul Fischer, and David Briggs. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33):7561 – 7578, 2008. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2008.05.057. URL http://www.sciencedirect.com/science/article/pii/S1352231008005748.

Jason Jingshi Li, Boi Faltings, and Radu Jurca. Incentive schemes for community sensing. In *The 3rd International Conference in Computational Sustainability*, 2012a.

Jason Jingshi Li, Boi Faltings, Olga Saukh, David Hasenfratz, and Jan Beutel. Sensing the air we breathe - the opensense zurich dataset. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI12), Toronto, Canada*, July 2012b.

Bjarke Mølgaard, Tareq Hussein, Jukka Corander, and Kaarle Hmeri. Forecasting size-fractionated particle number concentrations in the urban atmosphere. *Atmospheric Environment*, 46(0):155 – 163, 2012. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2011.10.004. URL http://www.sciencedirect.com/science/article/pii/S1352231011010491.

J Monroy, Achim Lilienthal, J Blanco, Javier González-Jimenez, and Marco Trincavelli. Calibration of mox gas sensors in open sampling systems based on gaussian processes. In *IEEE Sensors'12*, pages 1743–1746, 2012.

Trung V Nguyen and Edwin V Bonilla. Fast allocation of gaussian process experts. In *International Conference on Machine Learning*, 2014.

Carl Edward Rasmussen and Hannes Nickisch. Gaussian processes for machine learning (gpml) toolbox. *J. Mach. Learn. Res.*, 11:3011–3015, December 2010. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1756006.1953029.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. ISBN 0-262-18253-X.

Matteo Venanzi, Alex Rogers, and Nicholas R Jennings. Crowdsourcing spatial phenomena using trust-based heteroskedastic gaussian processes. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

World-Health-Organization. Air quality and health. In *Fact Sheet No. 313*, 2011.

A Novel Algorithm for Road Extraction from Airborne Lidar Data

Li Liu
School of Civil & Environmental Engineering
The University of New South Wales
Sydney, NSW 2032 Australia
l.liu@unsw.edu.au

Samsung Lim
School of Civil & Environmental Engineering
The University of New South Wales
Sydney, NSW 2032 Australia
s.lim@unsw.edu.au

Abstract

Road data in 3-dimensional forms is required for a variety of geospatial applications e.g. road maintenance, transport planning and location-based services. Although airborne lidar can produce dense point clouds from which 3-dimensional road information can be retrieved in detail, lidar data is often incomplete due to the line-of-sight requirement, and therefore a better result of information extraction from lidar data can be achieved if supplement data is used. This paper presents a novel algorithm for road extraction from lidar point clouds and associative vector data. A moving window-based classification technique using various cues determined from the lidar data is applied in a hierarchical way to separate road from other objects. In order to fill the gaps caused by the line-of-sight problem e.g. shadows of trees and buildings, the vector data is used in the refinement process to improve the results by including a buffer, deleting false positives, interpolating and fitting the road-representing points into polylines. To validate the proposed algorithm, four samples of lidar data sets are tested. The test result shows that it is a practical method for road extraction from airborne lidar data for both structured and unstructured lanes.

1 Introduction

Feature extraction of urban environments from geospatial data acquired by airborne mapping has been a significant research topic in photogrammetry and remote sensing for many decades (Mayer, 2008). For example, many research efforts have been attempted to extract roads from aerial images. However, roads have different appearances and characteristics in different scenes, and therefore road extraction from aerial images has to be adapted to particular datasets or scenes (Grote, *et al*, 2012).

Urban object extraction is still a sought-after topic, with a focus shifting to the detailed representation of objects, to using data from new sensors, and to advanced data processing techniques (Rottensteiner, *et al*., 2013). Compared with aerial photogrammetry, airborne lidar is a relatively new technique for geospatial data capture, providing much denser points. Triggered by the advent of lidar, experiments have been conducted to extract geo-features (e.g. trees, buildings, roads etc.) from lidar points. While a significant progress has been made in the extraction of buildings and trees from lidar data, little research has been performed on the extraction of roads (Vosselman and Zhou, 2009). Due to the thriving 3-dimensional (3D) mapping and geo-database updating, road extraction from lidar data is now feasible but is still in its infancy.

Most algorithms have exploited height differences, normal vectors, and contextual information to extract roads from 'raw' lidar data. In general, common strategies consist of segmentation and clustering (Choi *et al*. 2007; Zhu and Mordohar, 2009; Zhao and You, 2012), region growing (Akel *et al*., 2005; Clode *et al*. 2004a, 2004b, 2007) and curb detection (Vosselman and Zhou, 2009). Choi *et al*.(2007) extracted roads by utilizing a series of circled buffers. Also, a maximum possible slope of roads is applied to eliminate the erroneous object clusters. Zhu and Mordohar (2009) generated a road likelihood map with lidar points and extract dominant road regions with a minimum cover-set algorithm. Zhao and You (2012) developed the elongated structure templates that detect candidate road regions and a voting scheme is introduced to refine the road parameters.

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

However, it is not feasible to detect highly occluded roads. Akel *et al*. (2005) applied a region growing segmentation method based on elevation and normal vectors to detect road areas. Clode *et al*. (2007) introduced a hierarchical system to extract road points from airborne lidar data, and the results are convoluted with Phase Coded Disk (PCD) to generate the vectorised results. Vosselman and Zhou (2009) detected small height jumps between curbstones and road surfaces according to height differences and an elevation threshold. Zhou and Vosselman (2012) applied a method to mobile laser points and refined the detection process by a sigmoidal function. However, the results show that it is not well-suited to mobile laser points because of the occlusion by large vehicles and trees.

Because of the shadow in some highly-densed vegetation areas, the results in these areas are not promising. For the purpose of better accuracy and completeness, some researchers utilized aerial images as well as lidar data. Hu *et al*. (2004) exploited imagery and lidar data together in combination to extract road points. Ground points are generated by lidar data and imagery is used to distinguish roads from open areas. The results show that the accuracy and completeness were improved. Zhu *et al*. (2004) introduced the associated road image (ARI) extracted from lidar points to enhance the results from real road images (RRIs) from aerial imagery. However, road extraction from images and lidar points is still difficult since the registration of images and lidar points remains to be a problem.

Existing geospatial data is a complementary source to serve as a priori geometric knowledge about roads (Vosselman, 2003; Hatger and Brenner, 2003; Oude Elberink and Vosselman, 2006). Vosselman (2003) used lidar points and cadastral maps to reconstruct roads. Hatger and Brenner (2003) applied a fast region-growing algorithm to extract road geometry parameters from lidar data and existing road databases. Similarly, Oude Elberink and Vosselman (2006) fused lidar data and 2-dimensional topographic maps to extract road points. However, these methods suffer from the map scale and the generalization process.

In this paper, we propose a novel algorithm for road extraction from airborne lidar point clouds and vector data. The proposed method uses a hierarchical moving window that detects road points based on slope, elevation, and intensity values. Because of the highly-densed vegetation areas, the completeness of the results can be challenging to achieve. To refine the initial extraction results, vector data are utilised to improve the accuracy and completeness of the results, including creating a buffer, deleting false positives, filling gaps and fitting into lines. The main difference between existing algorithms and ours is that the vector data is used only in the refinement process to minimise the effect caused by the scale problem and the generalization problem in the vector data. If the vector data is used to partition the neighborhood, there are chances that road points may be left out when a buffer is created to retain lidar points within the buffer since there may be errors in the vector data. Our goal of this work is to develop an algorithm suitable to both structured and unstructured roads.

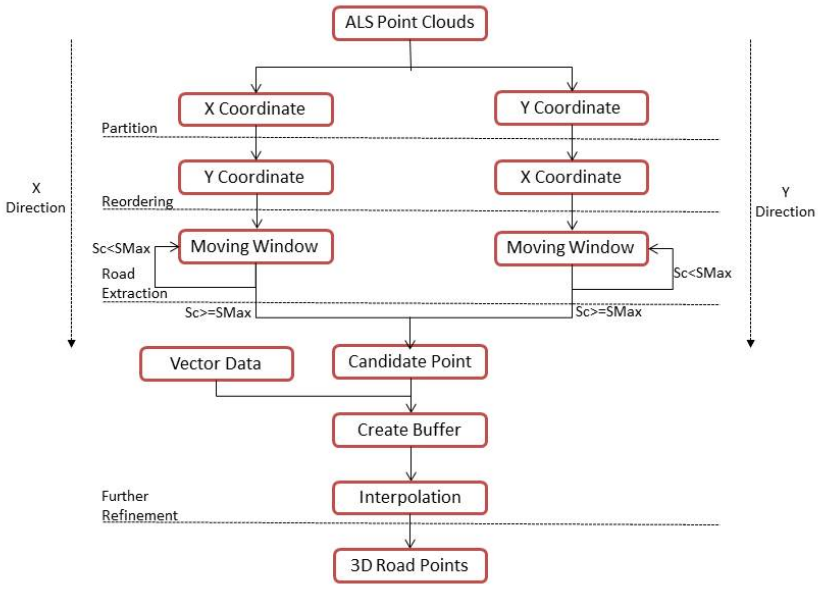


Figure 1 A framework of road extraction from airborne lidar data (Sc represents the current size of a moving window; SMax is the maximum size of the moving window)

2 Road Extraction

The proposed method has the following core steps as shown in Fig.1. Firstly, partitioning road points according to x/y coordinates and reordering the points according to y/x coordinates. Secondly, setting the

minimum and maximum sizes of moving windows and detecting road points within the given window. Thirdly, enlarging the window size and repeating the detection process until the window size meets the maximum threshold. Lastly, refining the results by vector data, filling the gaps and extracting the centerlines.

2.1 Overview

Fig.2 shows a top view of airborne lidar dataset. Transected by a line in Fig.2, a part of the vertical profile of this data (the red rectangle in Fig.2) is illustrated in Fig.3 (a), including mostly three types of points: building points, road points, and vegetation points. As Cheng *et al*. (2013) did, a moving window operator is proposed to detect road points. The basic idea is that different objects have different local geometric features: a road is a relatively flat and continuous surface; a building's rooftop may be also flat but is the highest part whiles the elevation of vegetation varies. As seen in Fig. 3 (b, c, d), when the size of the moving window gradually increases, road points can be distinguished from other points. Details will be illustrated in Sec.2.3.

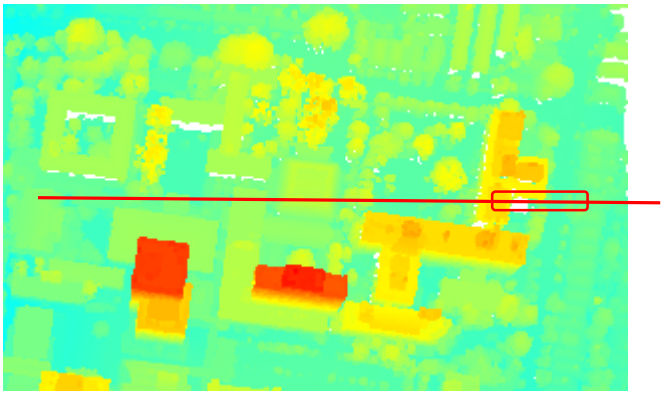


Figure 2 Top view of a sample airborne lidar dataset

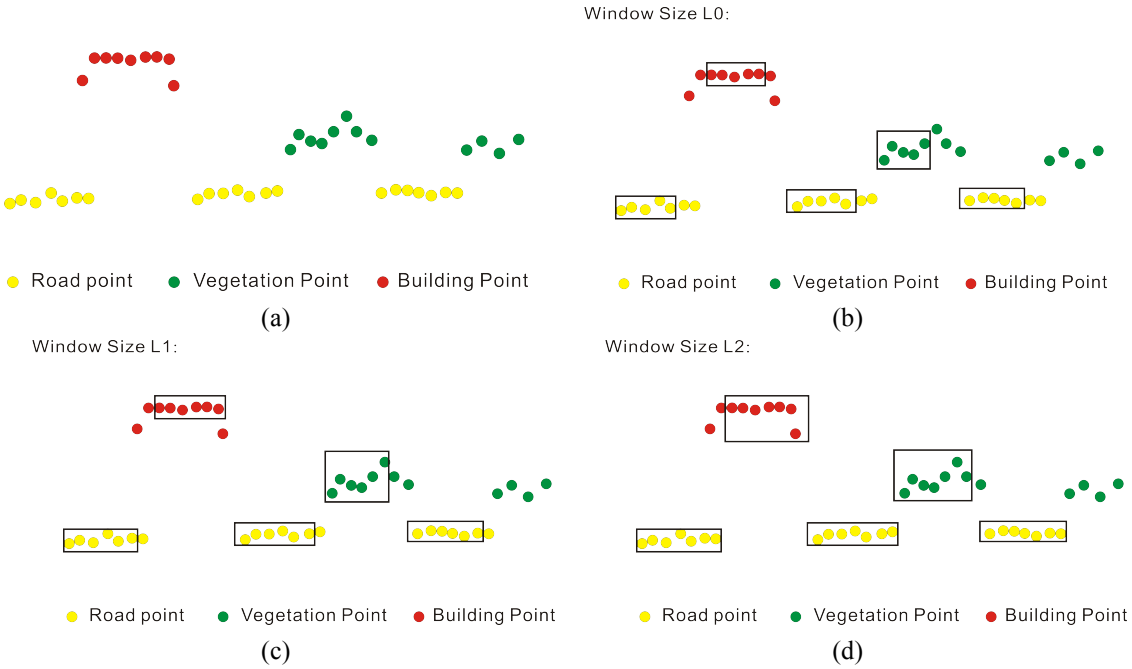


Figure 3 An initial process of roads extraction: (a) Part of the profile of lidar points from the transect line in Fig.2, (b) initialise the window size with L0, (c) increase the window size to L1 and repeat the process,(d) increase the window size to L2 and repeat the process

2.2 Partitioning lidar points

Airborne lidar data contains typically a huge amount of points. Therefore, lidar data possessing is overly time-consuming and complicated. To reduce the computation time, Vosselman and Zhou (2012) and Yang *et al.* (2013) divided lidar points according to scanlines and then extracted road points scanline by scanline; Boyko and Funkhouser (2011) used road maps to define the road surroundings and extract features within the surroundings. Pu *et al*. (2011) partitioned lidar data using buffers along the vehicle trajectories. Since road maps and trajectory data may suffer from a low accuracy, we partitioned lidar points according to the coordinates. At first, all points are partitioned by x/y coordinates, and then they are reordered by the y/x coordinates. In order to be suitable for road extraction in different directions, the dataset is also partitioned by x and y coordinates respectively.

2.3 Road Extraction

After partitioning the lidar data with respect to x and y axes, a moving window is initialized. This process requires the following steps:

Step 1. Set the minimum and maximum window sizes.

Step 2. Set the current window size as the minimum.

Step 3. Detect road points along x and y axes . If the points meet the criteria, increase the classification index of the points by one;

Step 4. Increase the window size by the pre-defined increasing step and repeat Step 3. If the window size is larger than the maximum threshold, do Step 5.

Step 5. Repeat Steps 2-4 until all x and y axes are processed.

Any point whose classification index is above the threshold is marked as the candidate for road points. To distinguish road points from other points, we defined three rules for slope, elevation and height differences, and intensity, as follows.

2.3.1 Rule 1: Slope

Since a road is a smooth and continuous surface, the slope of two adjacent road points should be low. Therefore the slope of two adjacent road points should meet the following condition:

$$\Delta S \le \Delta S_{\max} \tag{1}$$

where ΔS is the slope of two adjacent points, and $\Delta S_{\max}$ is the maximum slope of the surrounding points.

2.3.2 Rule 2: Elevation and Height Differences

To some extent, a road is the lowest part of the surrounding points (Haugerud and Harding, 2001). Elevation of the road is usually lower than the surrounding and the height difference of the road is small. Therefore the road points should meet these conditions:

$$\begin{cases} E_P \le E_{\max} \\ E_{P\max} - E_{P\min} \le \Delta H_{\max} \end{cases} \tag{2}$$

where E_P is the height of a lidar point of interest, $E_{\max}$ is the height threshold for surrounding lidar points calculated by the accumulative histogram algorithm, $E_{P\max}$ is the maximum height of the selected road points within the moving window, $E_{P\min}$ is the minimum height of the selected road points within the moving window, and $\Delta H_{\max}$ is the self-defined threshold of the height difference.

2.3.3 Rule 3: Intensity

The distribution of intensity values of ground points should meet the Gaussian Normal Distribution (Crosilla *et al*., 2013). In addition to that, intensity values of a road are obviously different from the surroundings, either higher or lower. Therefore the road points should meet the condition:

$$I_P \le I_{\max} \tag{3}$$

where: I_P is the intensity of the lidar point of interest, $I_{\max}$ is the intensity threshold for the surrounding lidar points, calculated by the accumulative histogram algorithm.

As seen in Fig. 3(b), if the window size is not properly chosen, some building rooftops can be misclassified as road points. In order to avoid the misclassification, we apply statistical methods to constrain the classification: when the point meets all these criteria with one window size, we add one to its classification index. And the classification index of road points should be larger than the preset threshold; otherwise the point would be marked as an object point.

2.4 Refinement of Candidate Road Points

After the initial roads extraction, some open areas may be misclassified as roads since they present similar characteristics as roads. Further refinement is needed to obtain more accurate results. In this paper, we utilise vector data to refine the initial results. The refinement process includes creating a buffer, deleting false positives, interpolating and fitting to a line, as shown in Fig.4. At first, a buffer is created along each vector segment to constrict the surrounding area and filter out the false positives. Points that fall into the buffer of one road segment are labelled as part of the road segment and will not be taken into account afterwards. If points are outside the buffers of all road segments, they will be labelled as noise and be removed. And then a linear interpolation will be applied along the main direction of each road segment to fill the gaps. To get a smooth surface, fitting is adapted to refine the results. And the centerline is calculated according to the refined results.

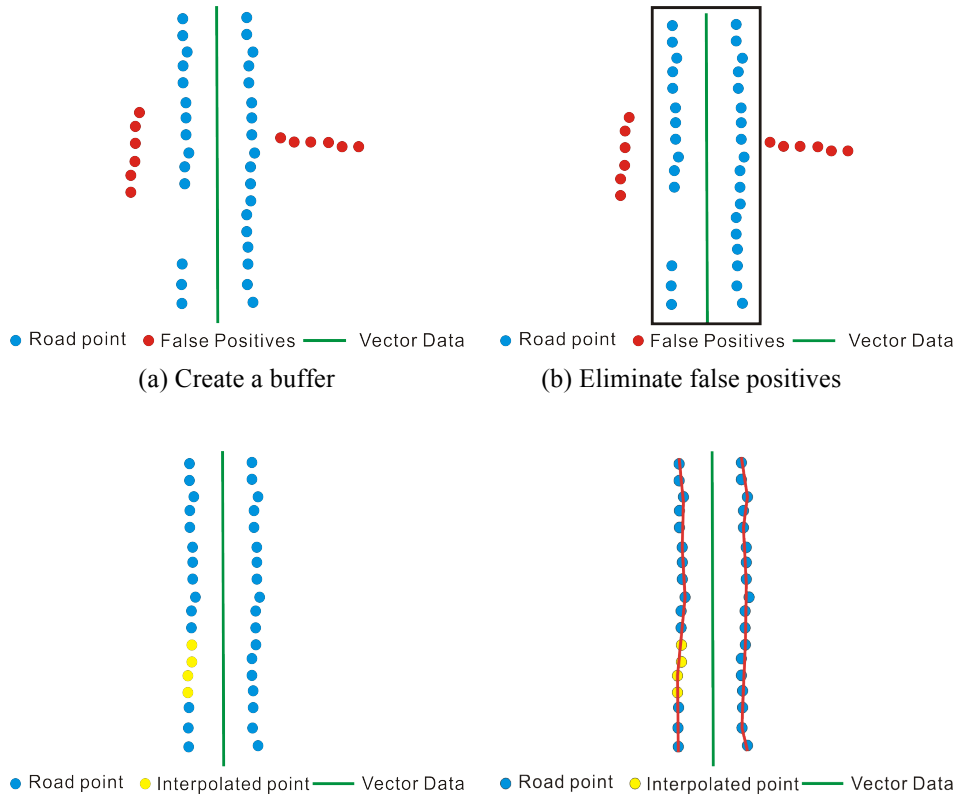


Figure 4. Refinement process: (a) buffering, (b) eliminating false positives, (c) interpolation, (d) curve fitting.

3 Results

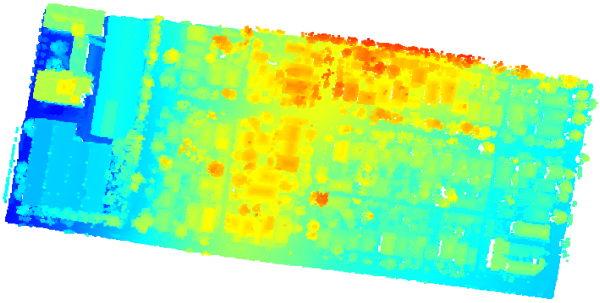
3.1 Test Data

Four samples of different areas and scenes are chosen for our study as shown in Fig.5 and summarized in Table 1. The first sample is a residential area near the University of New South Wales (UNSW), including 144,255 points. The maximum elevation difference within the dataset is 50.73 m. There are high slopes along the horizontal streets, and in some areas roads are heavily blocked by trees. The second sample is a part of Anzac Parade (an arterial road) with its surrounding residential areas, with 128,847 points. It consists of high-rise buildings, small residential houses, vegetation, structured roads and unstructured roads. In some part of Anzac Parade, it is highly occluded by cars. The third sample is a part of Barker Street (a main road) with

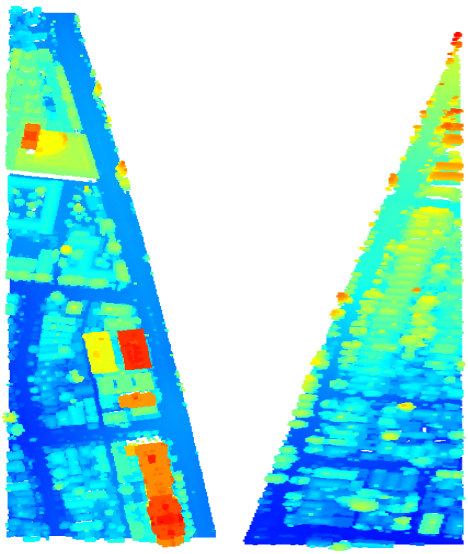
its surrounding areas with 124,690 points. Residential houses, trees and roads make up the entire dataset and there is a high slope along the street leading to the fact that in some areas the heights of houses are even lower than that of adjacent roads. The last sample is a part of Botany Street (a main road) with its surrounding areas with 69,064 points. It mainly consists of houses, vegetation and roads. The vector data available is the road network whose resolution is 2 m. Because of the fast development of cities, road geometry can change quickly, which causes a problem to the vector data updating. In this sense, the vector data should comply with the lidar dataset. In this experiment, only the coordinates of starting points and ending points for road segments are available. And the assumption of the vector data is that the road geometry is accurate enough to be used for refinement of the lidar data processing results.

Table 1 Characteristics of lidar samples

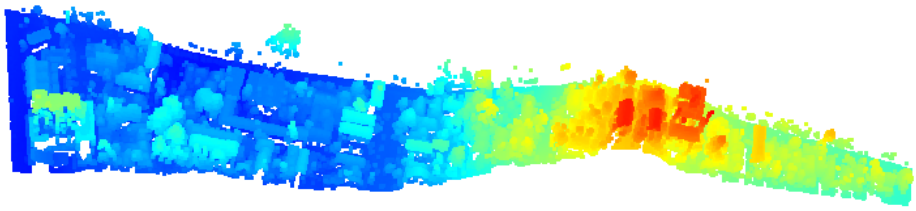
	Length (m)	Width (m)	Maximum Elevation (m)	Minimum Elevation (m)	Point density (pt/m^2)
Sample 1: Residential area	507.32	270.23	81.52	30.79	1.05
Sample 2: Anzac Parade	713.10	191.98	57.14	19.93	0.94
Sample 3: Barker Street	155.59	839.68	74.85	24.64	0.95
Sample 4: Botany Street	708.25	132.66	73.60	43.75	0.74



(a)



(b) (c)



(d)

Figure 5 Four airborne lidar datasets: (a) a residential area, (b) Anzac Parade, (c) Botany Street, (d) Barker Street.

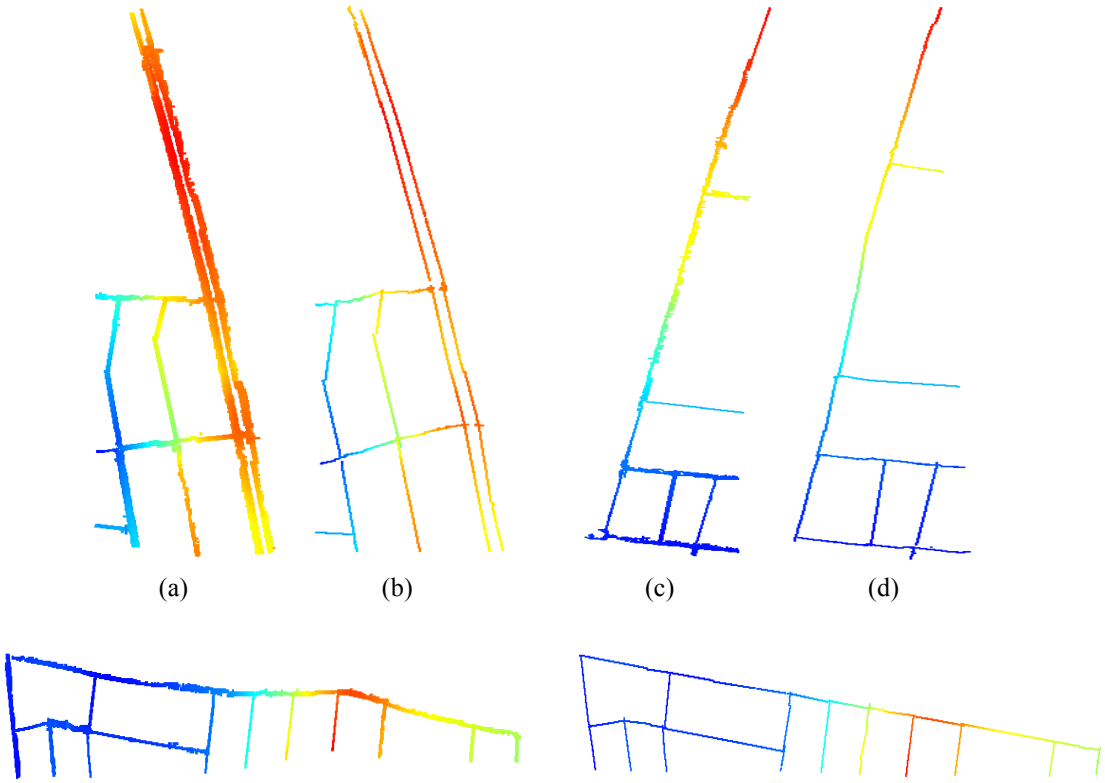
3.2 Extraction Results

Based on the proposed method, the minimum window size is set to 15; the maximum window size is set to 35; and the increase step is defined as 5. Then the slope, the height difference, the classification index threshold and the gap between two candidate road segments are also specified as listed in Table 2. The size of window changes gradually from 15 points to 35 points. If a point meets the criterion in all neighborhoods from a window size of 15 to 35, it is regarded as a road point.

Table 2 List of values of parameters

Symbol	Description	Value
S_{Max}	The maximum of moving window size	35
S_{Min}	The minimum of moving window size	15
S_i	Increase step of moving window size	5
ΔS_{max}	The maximum slope of two adjacent road points	20°
ΔH_{max}	The threshold of road height difference	0.8 m
N_i	The value of classification index	4

Fig.6 shows the road extraction results and respective centerlines of four data samples, which indicates the feasibility of the proposed method.



(a) (b) (c) (d)

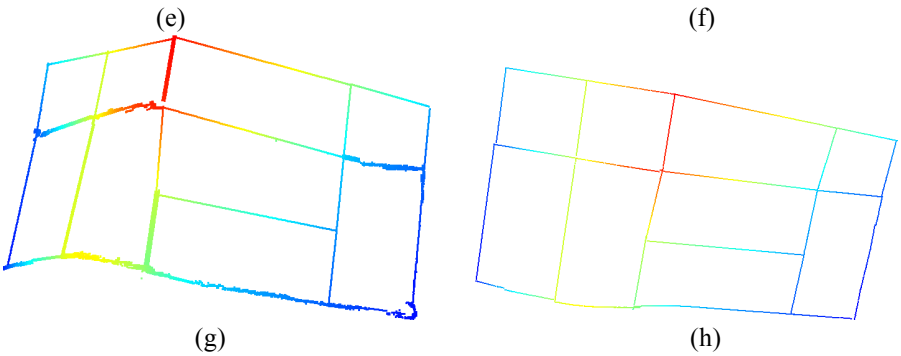


Figure. 6 Results of the extracted road network and respective centerlines: (a) extraction results of Anzac Parade, (b) estimated centerlines of Anzac Parade, (c) extraction results of Botany Street, (d) estimated centerlines of Botany Street, (e) extraction results of Barker Street, (f) estimated centerlines of Barker Street, (g) extraction results of a residential area, (h) estimated centerlines of the residential area.

3.3 Quantitative evaluation of 3D road extraction

Heipke *et al*., (1997) stated that the quantitative evaluation of road extraction can be addressed in three ways: completeness, correctness and quality, which are also used by Clode *et al*.(2007) and Yang *et al*.(2013). Completeness is the ratio of the total length of the extracted road that matches the reference roads to the total length of the reference data. Correctness is the ratio of the total length of the extracted road that matches the reference roads to the total length of the extracted data. Quality is the ratio of the total length of the extracted road that matches the reference roads to the total length of the sum of extracted data and undetected roads. Firstly, reference or ground truth data is extracted from images by digitizing the road centerlines and the length of reference roads is measured several times to get the average value. The length of extracted roads is calculated by summing up the lengths of each segment which can be achieved by a coordinate-based computation. Then results of the proposed method are compared with the reference data by calculating the maximum bias between the extracted road segment and the reference segment. If the maximum bias is above the threshold, the whole road segment is regarded as an unmatched road. The maximum tolerance of bias is set to 1.5 m. The statistics for our quantitative analysis are given in Table 3. Accuracy metrics are calculated by the following equations (4-6):

$$Completeness = TP/Lr \quad (4)$$

$$Correctnes = TP/Le \quad (5)$$

$$Quality = TP/(Le + FN) \quad (6)$$

where:

Lr is the total length of the reference roads;
Le is the total length of the extracted roads;
TP is true positives, namely, the total length of the extracted roads that matches the reference roads;
FN is false negatives, namely, the total length of the undetected roads.

The statistics for each dataset are computed as follows:

Table 3 Statistics for quantitative analysis

	Length of reference roads (m)	Length of extracted roads (m)	Length of undetected roads (m)	Total length of Matched roads (m)
Sample 1: Residential area	2227.14	2218.51	32.31	1809.58
Sample 2: Anzac Parade	2493.20	2445.45	61.98	1740.50
Sample 3: Barker Street	1958.33	1949.94	0	1643.72
Sample 4: Botany Street	1268.89	1215.81	58.20	1167.55

The quantitative analysis results are summarized in Table 4, which indicates that Sample 2 shows the worst case. A further inspection indicates that the majority of unmatched roads occurred in two polylines. The main reason is that only starting nodes and ending nodes are available in the vector data and the way points are not present, which obviously contributes to the biases. The majority of undetected roads

are the traffic islands and crossings of the two roads. As for traffic islands, although most of the edges are extracted in the initial extraction, they are merged to other road segments when centrelines are extracted. As for the crossings, while they are listed as separated road segments in the reference data, they will be ignored and regarded as one segment in the initial extraction.

Table 4 The results of quantitative analysis

	Completeness (%)	Correctness (%)	Quality (%)
Sample 1: Residential area	81.25	81.57	80.40
Sample 2: Anzac Parade	**69.81**	**71.17**	**69.41**
Sample 3: Barker Street	84.30	83.93	83.93
Sample 4: Botany Street	92.01	96.03	91.64

4 Concluding Remarks

Roads are important infrastructure and therefore road extraction from lidar data is of great importance in terms of civil engineering and urban planning. In this paper, we presented a novel method for road extraction from airborne lidar data and vector data. The proposed method utilises the local geometry and the radiometric feature of roads to classify road points from the lidar point clouds. To effectively process the data, all points are partitioned with respect to x and y axes and reordered by coordinates. A moving window is initialized to detect the road points. After the initial process, those points whose classification indices are above the threshold are marked as road points..In order to refine the results, vector data is used to eliminate false positives and, in the end, gaps between road segments are filled by a linear interpolation method. The proposed method successfully extracts the roads from airborne lidar data. The quantitative results show that it is a promising method for road extraction, not only suitable for structured roads, but also feasible to unstructured lanes.

Although the presented method is able to extract roads with reasonable completeness, quality and correctness, it is vulnerable to the problem of traffic islands extraction. If not only the starting and ending nodes but also way points are available in the vector data, our a priori geometry knowledge of road segments will help reduce biases in the refinement process. Also, the choice of a window size is critical to the extraction results. If the maximum window size is too small, some building rooftops may fall into the whole window, thus being misclassified as roads. On the other hand, if the minimum window size is too small, it would take too much time for iteration. The aforementioned vulnerability of the proposed algorithm will be the focus of our further research.

References:

Akel, N.A., Kremeike, K., Filin, S., Sester, M., Doytsher, Y., (2005). Dense DTM generalization aided by roads extracted from LIDAR data. In: IAPRS 36 (Part 3/W19), 54–59.

Boyko, A., Funkhouser, T. (2011). Extracting roads from dense point clouds in large scale urban environment. ISPRS J. Photogrammetry and Remote Sens. 66 (6), S2–S12.

Cheng, L., Zhao, W., Han, P., Shan, J., Liu, Y.X., Li, M.C., (2013). Building region derivation from Lidar data using a reversed iterative mathematic morphological algorithm. Optics Communications (286): 244-250.

Choi, Y.W., Jang, Y.W., Lee, H.J., Cho, G.S., (2007). Heuristic road extraction. In: International Symposium on Information Technology Convergence, IEEE Computer Society.

Clode, S., Kootsookos, P. Rottensteiner, F., (2004(a)). The Automatic Extraction of Roads from Lidar Data. In: IAPRSIS, *Vol*. XXXV-B3, pp. 231 – 236.

Clode, S., Zelniker, E., Kootsookos, P., Clarkson, V., (2004(b)). A Phase Coded Disk Approach to Thick Curvilinear Line Detection. In: 17th European Signal Processing Conference, 6-10 September, 2004,Vienna, Austria, pp.1147-1150.

Clode, S., Rottensteiner, F., Kootsooks, P., Zelniker, E., (2007). Detection and Vectorisation of Roads from Lidar Data. PE&RS, *Vol*. 73(5), 517–536.

Crosilla, F., Macorig, D., Scaioni, M., Sebastianutti, I., Visintini, D., (2013). Lidar data filtering and classification by skewness and kurtosis iterative analysis of multiple point cloud data categories. Applied Geomatics. *Vol*. 5, Issue 3, pp 225-240

Elberink, S.J.O., Vosselman, G., (2006). 3D modelling of topographic objects by fusing 2D maps and lidar data. In: IAPRS, *Vol*. 36, part 4, Goa, India, September 27-30.

Gröte A., Heipke, C., Rottensteiner, F., (2012). Road Network Extraction in Suburban Areas. The Photogrammetric Record. Vol. 27,No. 137, pp: 8–28.

Hatger, C., Brenner, C., (2003). Extraction of road geometry parameters from laser scanning and existing databases. IAPRS, *Vol*. 34, 2003, 225-230.

Haugerud, R.A., and Harding, D.J., (2001), Some algorithms for virtual deforestation (VDF) of lidar topographic survey data: IAPRS , XXXIV-3/W4, p. 211–217.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., (1997). Evaluation of automatic road extraction. IAPRS XXXII (1997), pp. 47–56.

Hu, X., Tao, C.V., Hu, Y., (2004). Automatic road extraction from dense urban area by integrated processing of high-resolution imagery and LIDAR data. IAPRS 35 (Part B3), 288–292.

Mayer, H., (2008). Object extraction in photogrammetric computer vision. ISPRS J. Photogrammetry & Remote Sens. 63(2):213-222.

Pu, S., Rutzinger, M., Vosselman, G., Oude Elberink, S. (2011) Recognizing basic structures from mobile laser scanning data for road inventory studies. ISPRS J. Photogrammetry and Remote Sens. 66: S28-S39.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J., D., Breitkopf, U., Jung, J., (2013), Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS J. Photogrammetry & Remote Sens. In Press.

Vosselman, G., Zhou, L., (2009). Detection of curbstones in airborne laser scanning data. In: IAPRS XXXVIII – 3/W8, pp. 111-117, Paris, France, 2009.

Vosselman, G., (2003). 3-D Reconstruction of Roads and Trees for City Modelling. In: IAPRS, *Vol*. 34, part 3/W13, Dresden, Germany, pp. 231-236.

Yang, B. S., Fang, L.N., Li, J., (2013). Semi-automated extraction and delineation of 3D roads of street scene from mobile laser scanning point clouds. ISPRS, *Vol*.79: 80-93.

Zhao, J.P., You, S.Y., (2012). Road Network Extraction from Airborne Lidar Data Using Scene Context. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Zhou, L. and Vosselman, G., (2012). Mapping curbstones in airborne and mobile laser scanning data. International Journal of Applied Earth Observation and Geoinformation 18(2012), 293-304.

Zhu, P., Lu, Z., Chen, X., Honda, K., Eiumnoh, A., (2004). Extraction of city roads through shadow path reconstruction using laser scanning. PE&RS 70(12), 1433–1440.

Zhu. Q.H., Mordohai, P., (2009). A Miumum Cover Approach for Extracting the Road Network from Airborne Lidar Data. In: IEEE 12th Conference on Computer Vision Workshop

Intersection Delay Estimation from Floating Car Data via Stacked Generalization: A Case Study on Beijing's Road Networks

Xiliang Liu, Feng Lu, Hengcai Zhang

State Key Lab of Resources and Environmental Information System,

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences,

11A, Datun Road, Chaoyang District, Beijing 100101, P. R. China

{liuxl, luf, zhanghc}@lreis.ac.cn

Abstract

Modeling the time cost on driving through the intersections among road networks can significantly help improve the reliability of car navigation systems or web map systems, as well as other real traffic related applications (Liu et al., 2013). Some researchers rely on parametric approaches including the historical mean (Sun 2007; Zhao et al. 2013), linear interpolation (Ban et al. 2009), auto regression moving average (ARMA) (Zhang and Liu 2009a), etc. These methods perform well in stationary traffic flow scenario, but may fail facing with the dynamics of the urban networks. Some researchers notice the instability of traffic flow and employ sophisticated non-parametric models such as artificial neural network (ANN) (Zhang et al. 2011), support vector machine (SVM) (Zhang and Liu 2009b) and so on. These methods show advantages in over saturated conditions, but require prior knowledge about the models and extra specific parameter tuning, which restrict the generalization of these models. To date, it seems still hard to get a satisfying result just from one model facing with the complexity of traffic phenomena due to the complexity of transport system. In view of the limitations of traditional approaches, in this paper, we propose a pervasive framework to estimate the intersection delays from floating car data (FCD) based on stacked generalization ensemble learning theory (Wolpert, 2002), trying to integrated the advantages of different models and to get rid of extra artificial interference. Firstly we analyze the optimal linear combination based on error-ambiguity decomposition and redesign the learning strategy in the level-1 phrase of stacked generalization; Secondly we integrate six classical and frequently used approaches into this framework, including linear least squares regression (LLSR), autoregressive moving average (ARMA), historical mean (HM), Artificial Neural Network (ANN), Radical Basis Function Neural Network (RBF-NN), Support Vector Machine (SVM), and

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

conduct experiments with a real floating car dataset collected with more 20,000 taxies in a four-month period of 2011 in Beijing city. In order to testify the effectiveness of the proposed model, we further compare our estimation result with other four linear combination methods, namely the equal weights method (EW), optimal weights method (OW), minimum error method (ME) and minimum variance method (MV). Final results on 400 main intersections among Beijing's road networks confirm that the proposed stacked generalization approach behaves more robust and accurate than any single approach, and outperforms those classical linear combination strategies both in variance and bias. We give possible explanations based on mathematical analysis and spatial-temporal distribution of floating car data. In addition, the proposed stacked generalization approach provides a scalable, open-ending framework for the intersection delay estimation, which means any new promising models can be easily incorporated in our future work.

References

Ban, X. J., Herring, R., Hao, P., Bayen, A. M. (2009). Delay pattern estimation for signalized intersections using sampled travel times. *Transportation Research Record*, (2130), 109-119.

Liu, X., Lu, F., Zhang, H. and Qiu P. (2013). Intersection delay estimation from floating car data via principal curves: a case study on Beijing's road network. *Frontiers of Earth Science*, 7(2), 206-216.

Sun, L. (2007). An approach for intersection delay estimate based on floating vehicles. *Ph.D thesis*. Beijing: Beijing University of Technology (in Chinese).

Wolpert, D.H. (2002). Stacked generalization, *Neural Networks*, 5(2), 241-260.

Zhang, H., Lu, F., Zhou, L., Duan, Y. (2011) Computing turn delay in city road network with GPS collected trajectories. In: *Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis*, September 17-21, Beijing, China, ACM, 45-52.

Zhang, Y., and Liu, Y. (2009a) Comparison of Parametric and Nonparametric Techniques for Non-peak Traffic Forecasting. In: *Proceedings of World Academy of Science, Engineering and Technology*, 8-14.

Zhang, Y., and Liu, Y. (2009b) Traffic forecasting using least squares support vector machines. *Transportmetrica*, 5(3):193-213.

Zhao, M., Li, X. (2013). Deriving Average Delay of Traffic Flow around Intersections from Vehicle Trajectory Data, *Frontiers of Earth Science*, 7(1), 28-33.

Integrated land evaluation: Story of a track not taken[1]

Nicholas Chrisman
Geospatial Sciences, RMIT University
Melbourne VIC Australia
nicholas.chrisman@rmit.edu.au

Abstract

Many cite CGIS and Tomlinson as the origin of GIS, yet it is instructive to examine the community of practice into which this idea was presented in 1967. A group of practitioners in CSIRO Lands Directorate had evolved a scheme for integrated land evaluation. This paper considers the track not taken, and the delicate connection between technology and conceptual frameworks. This story has particular relevance to spatial data integration. The prior technology dealt with a deeper kind of integration than we manage with current approaches.

1 Preamble

The first round of GIS History was written by the winners, often as a tale of great feats of dare-doing (Foresman, 1998, for example). Recently much more nuanced accounts focus on the interaction of society, organizations, and GIS technology (Harvey and Chrisman, 2004; Chrisman, 2006a, 2006b). History requires careful reading of the contemporaneous documents, not what the actor comes to think later on. But that remains the history of the victors, in the sense that we focus on the developments leading to the current day. What happens to the tendencies that disappear? How do we remember the tracks not taken, the technological developments that do not lead to vibrant industries with gigantic conferences? This paper will consider this more speculative form of history.

But this is not merely a return to a parallel world that might have been. The lost approaches have much to offer in the realm of spatial data integration. It is crucial to look back to history to enrich the current strategies.

1.1 Before the beginning

The game of origins can get quite convoluted. After all, every event occurs in some circumstances, and thus there is literally nothing new under the sun. Rather than a simple branching tree, leading back to a unitary root, the history of technology may reach a level of "rhizomes" where lateral connections link lillies in a complex maze (Deleuze and Guatari, 1976; Harvey and Chrisman, 2004).

Roger Tomlinson (1998) claims to have built the "first GIS", a claim that many accept at face value (despite multiple origins such as Garrison and others, 1965). Certainly Tomlinson and his colleagues at the Canada Land Inventory in Ottawa did build a functioning system that included many innovations. I do not seek to minimize these advances. But Tomlinson's first published paper indicates the community into which this GIS development contributed. It brings us to this city, Canberra 46 years ago. Tomlinson brought his first paper to a conference (26-31 August 1968) hosted by CSIRO, which assembled a collection of professionals engaged in "terrain evaluation," an integrative practice of natural resource survey (Stewart, 1968). This group (Stewart, 1952; Mabutt and Stewart 1963; Mabutt 1968) understood what Tomlinson wanted to achieve, since that was the kind of work they already performed—with different technologies.

This article will consider the quite Australian paradigm of terrain evaluation as it was practiced by the CSIRO Lands Directorate for a few decades, and what that practice has to offer in a world of linked data, spatial information integration, cloud computing and volunteered geographic information.

2 Integrated terrain evaluation

In the review article that led off the collected proceedings of the conference in 1968, Mabutt (1968) divided land inventory into three general approaches: genetic, parametric and landscape. According to this classification, geological

In: S. Winter and C. Rizos (ed.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

maps are the classic example of ‘genetic’ maps, being classified according to their origins. Any attributes are derived from the historical sequence. Beyond geology, and certain soils surveys, the genetic approach is not the primary technique for land evaluation and measurement. These sciences played a key role in early land surveys, but track of process is harder to trace for many other concerns. Most of our current GIS is filled with ‘parametric’ information – where the maps are structured to present a particular variable, with spatial units to depict various attribute values. This is a measurement framework termed categorical coverage, and in wide use. The lines on the map are determined by a particular level of the attribute. This parametric view of a map is so dominant that genetic maps are often presented in these terms.

By contrast, Mabutt’s landscape approach considered many attributes at once. "The land complex as a whole is the object of study, even where a particular attribute may be of prime interest to a land classifier." (Mabutt, 1968, p. 16) In certain academic literature, this approach is presented as ‘gestalt’ (Hopkins, 1977), implying a connection to a branch of cognitive psychology from the previous century. Certainly the landscape approach is interested in the whole, not each item. Yet, the motivation is not based on psychology. Mabutt argues for a synthesis of various viewpoints on the landscape, thus the integration is very much the issue. It was accomplished by various forms of fieldwork and collaboration by a group of specialists representing the parametric disciplines. My anonymous sources (uncited to protect the informants) who worked in these teams tell of the force of personality playing perhaps more of a factor than it should in the purity of the method. This was a method profoundly connected to human elements, and professional training. In a modern era when we think of technical factors, we often undervalue the influence of these human factors.

The landscape approach, as developed by CSIRO in Australia, involved sending an interdisciplinary team into the field together. They would not produce distinct layers, but one single final interpretation. Figure 1 and Table 1 present one of these "land systems" defined on the basis of geology, vegetation, and soils (Story and others, 1963).

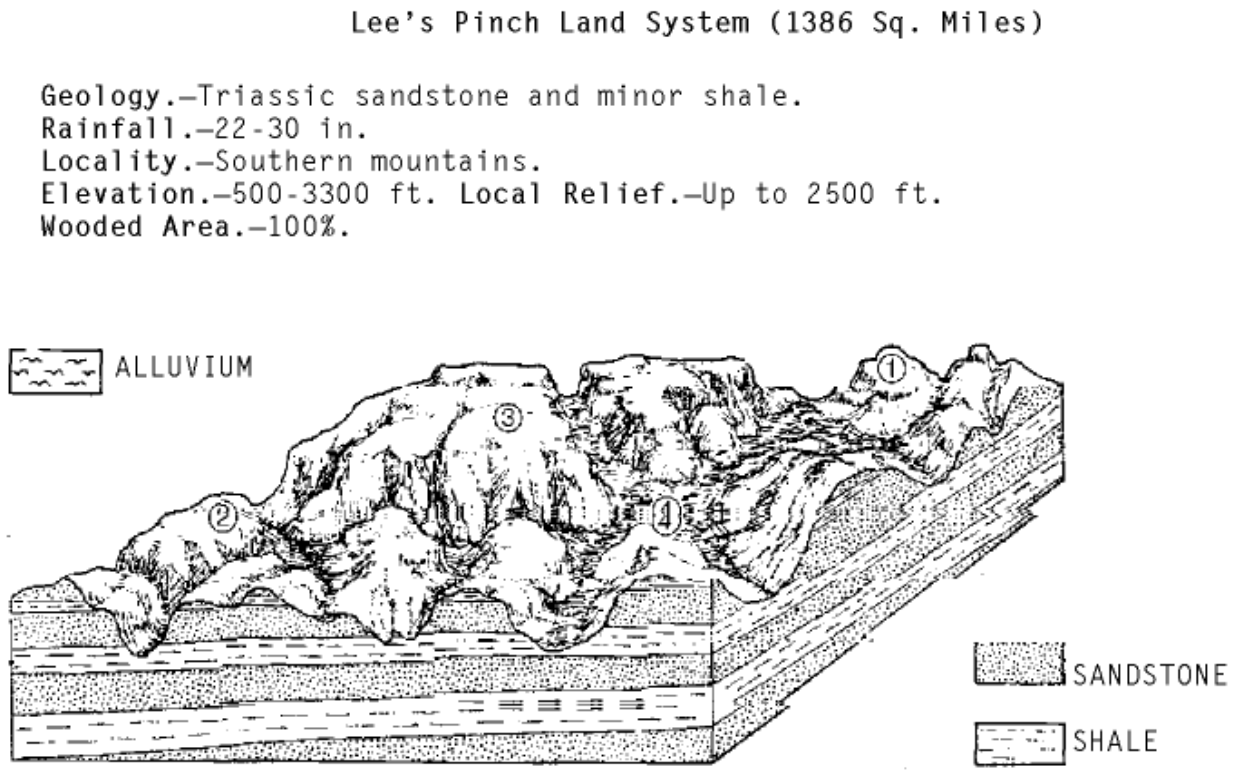


Figure 1: Lee’s Pinch Land System *Source: (Story and others, 1963).*

Table 1: Explanation the four land units of Lee's Pinch Land System *(Note 1 foot = 0.3 m)*

Unit	Area	Land Forms	Soils	Vegetation
1	30%	Rugged hills with rounded summits; irregularly benched slopes often littered with boulders and with very frequent sandstone outcrops including low cliffs up to 30 ft. high; fairly narrow flat-floored valleys 400 - 1000 ft deep	Mainly shallow coarse-textured skeletal soils and bare rock; in moist cool sites humic surface-soils; infrequently on interbedded shales or arkosic sandstones shallow podzolic soils (Binnie, Pokolbin); in stable sites coarse-textured earths	Shrub woodland of ironbark and gum 4080 ft high, iron-barks common, with *E. punctata, E aggiomerata,*and *E. oblonga,* and with scattered or dense *Callitris endlicheri, Casuarina torulosa,* and *Persoonia spp*. below; shrubs usually abundant and mixed,*Leguminosae* common; ground cover poor, of grasses and herbs
2	30%	Rugged hills margined by sandstone cliffs 50 - 500 ft high usually overlooking steep shaly slopes littered with boulders; cavernous weathering of the cliffs; narrow inaccessible valleys 500 - 2500 ft deep	Similar to unit 1; predominantly coarse-textured non-humic skeletal soils; probably more bare rock	As for unit 1, but with more herbs, shrubs, and non-eucalypt trees in ravines and at bases of cliffs
3	35%	Stony, hilly plateaux with ridges and escarpments up to 200 ft high; very steep margins including cliffs up to 100 ft high; narrow gorges along the major rivers	Restricted observations; similar to units 1 and 2; deep yellow earth (Mulbring) in level, stable site on plateau	Shrub woodland of ironbark and gum 30 ft high, including*E. punctata*, *E. trachyphola*,and stringybarks; ground cover poor; many non-eucalypts in ravines and at bases of cliffs
4	<5%	Sandy alluvium occupying valley floors in unit 1; liable to frequent flooding and deposition of sand in middle and upper reaches	Restricted observations; deep sandy stratified alluvial regosols (Rouchel); sedimentation in valley bottoms frequent and calamitous owing to low soil stability on sandstone hills	Shrub woodland or ironbark and gum with an admixture of non-eucalypt trees, sometimes cleared and under pioneer grasses

Source: *General report on the lands of the Hunter Valley.* (Story and others, 1963)

This "Lee's Pinch" land 'system' in the Hunter Valley had four 'units', each in relationship to each other and to the distribution of soils, vegetation and topography. The underlying geology had shale and sandstone in interleaved beds, nearly horizontal. This geology formed the basis for the soils, and thence the vegetation. This multi-faceted landscape system was portrayed on a single map, produced out of negotiation among the various specialists sent into the field, and then delineated on airphoto surveys (Christian, 1952; Mabutt and Stewart, 1963; Christian and Stewart, 1968). The various elements of the landscape were to be handled together as an integrated whole, not through separate maps and separate layers. Attributes would be assigned to the land system and the units, not to specific polygons.

The techniques developed at CSIRO led to a major effort of the Food and Agriculture Organization of the United Nations (FAO 1976), and a whole community of practice with developments in Netherlands and elsewhere. Practitioners in this form of land inventory were trained around the World, including a substantial number at ITC in the Netherlands (Zonneveld, 1972). The popularity of this approach did not diminish in the 1970s as GIS and remote sensing began to develop. For example, one researcher from USGS travelled to Australia to work with the landscape approach, and attempted to derive landscape (integrated) units from the parametric measurements of the early Landsat imagery (Robinove, 1979). The experiment ran into some difficulties matching what the satellite can detect to the land units scheme. By this time, the technique was termed "integrated terrain unit mapping" (ITUM) at least in North America. One of the greatest proponents of this technique in the United States was Jack Dangermond (1979). ESRI's President championed ITUM (presenting a case study in the Zulia region of Venezuela at the first Harvard Computer Graphics Week in 1979) at the very time his company was developing software packages based on a different (parametric) model. At that point, Aerial Information Systems (a company related to ESRI), specializing in air photo interpretation, had more employees than ESRI. This company continues to work with ITUM concepts, for example in a long-term contract with the State of New Jersey (http://www.aisgis.com/projects/NewJersey.html).

The ITUM method, as implemented by ESRI and its sister company in the Zulia case, Alaska and New Jersey at about the same time, involved making one coverage of polygons that could be reclassified without additional linework to portray a number of variables. So, it could present something that looked like each parametric map, but from a common database. The attraction of this approach was quite apparent at that time, for technical reasons. Polygon overlay procedures were inefficient and unable to deal with imprecisions in borders (Goodchild, 1978; Dougenik, 1980; Chrisman and others, 1992). The experience of CGIS had established the terminology of 'slivers' to refer to small thin areas created by the overlay of lines intended to be the same but that differed by small amounts. These slivers enlarged the database, creating a flood of spurious entities that could not be treated with the algorithms of the day.

3 Competitors and fellow-travellers

In 1968, when CSIRO ran its seminar, there were other efforts going in with more or less communication. In 1967 (essentially simultaneously), the Landscape Architecture Research Office (1967) at the Graduate School of Design, Harvard University hosted a series of speakers on the subject of 'environmental resource analysis' – much the same topic as landscape evaluation. Harvard did this under contract with the Washington-based Conservation Foundation, but with a clear goal to develop strategic directions in research (Chrisman, 2006, p. 42-43).

This event occurred before Tomlinson's announcement to the World, but it concentrated on the content and analytical procedures not the technology. The three experts invited were Ian McHarg (University of Pennsylvania), Philip Lewis (University of Wisconsin) and Angus Hills (Ontario).

Angus Hills was a mapping expert, quite connected to the CSIRO group from the Ontario forestry agency. Hills (1966) presented his version of terrain assessment, which had been adopted as the basis for the Canada Land Inventory (CLI), the project that had created the requirement for Tomlinson's computer system CGIS. In a variant of the Australian land systems approach, Hills set up CLI with the concept of mixtures in each polygon. Figure 2 presents the descriptive legend to decipher the CLI maps. Each polygon could be coded with one (or many) capability classes, calibrated in tenths. The example shows a polygon that is 60% class 1 and 40% class 3. In addition it is coded with additional subclasses that may occur in some part of the polygon.

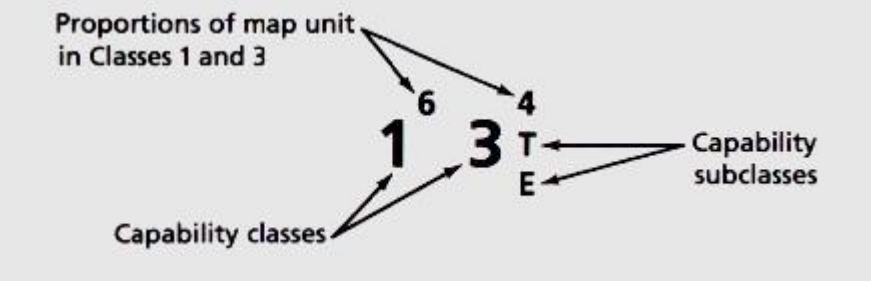


Figure 2: Legend from Canada Land Inventory *Source: CLI map sheet Ville Marie.*

Clearly any software written to handle CLI would have to be rather delicate to ensure that it could handle these mixture codes and the logical issues raised by a variable number of attributes attached to each polygon. These complications have not been included in the story about CGIS as the first GIS. While the group from Harvard listened to Hills, they were more attracted to the presentations of McHarg and Lewis.

Phil Lewis presented his analytical framework for environmental analysis, largely based on connectivity and 'corridors' (Lewis, 1963). Lewis's vision of the environment was interconnected and systematic, much as the views of Mabutt. Most of Lewis's work started with points connected by lines, though some masses of polygons were used to define the background and the connectivity. This was tricky stuff along the line of pattern recognition that would be hard to automate in the first instance. Lewis presented a way to integrate based on a human interpreter. While perhaps more scientifically valid, simpler technical fixes could be more attractive.

Ian McHarg presented the work that he was doing that later appeared in his massively popular book (McHarg, 1969). McHarg used simple map layers, mostly black and white, to exclude areas from the analysis. He stacked them up, looking for the least threatened areas, or the most suitable sites. A confirmed showman, McHarg made it all look easy and direct. No need to integrate the maps, on the light table they were all seen together. McHarg was presenting the basic overlay approach, where each map presented on parameter.

4 Overlay-based organization

The parametric mapping approach has venerable origins, and equally contested history (Manning, 1913, Steinitz and others, 1976; Cloud, 2005). Polygon overlay played a key role in making GIS software viable (Burrough 1986). Most academic research on GIS in the early phase placed the overlay function at the core (Tomlinson 1974; Chrisman 1982; Tomlin 1990). This software capability became a kind of litmus test to separate mere mapping from GIS (Boyle and Tomlinson, 1981). Certainly, this centrality has diminished as the capabilities have had to expand to respond to many different marketplaces and user communities. Yet, the metaphor of map layers (Figure 3) remains a central element of the graphic interface, even using software that no longer is as strongly tied to the topological coverage model. The logic for placing overlay at the core deserves some reexamination.

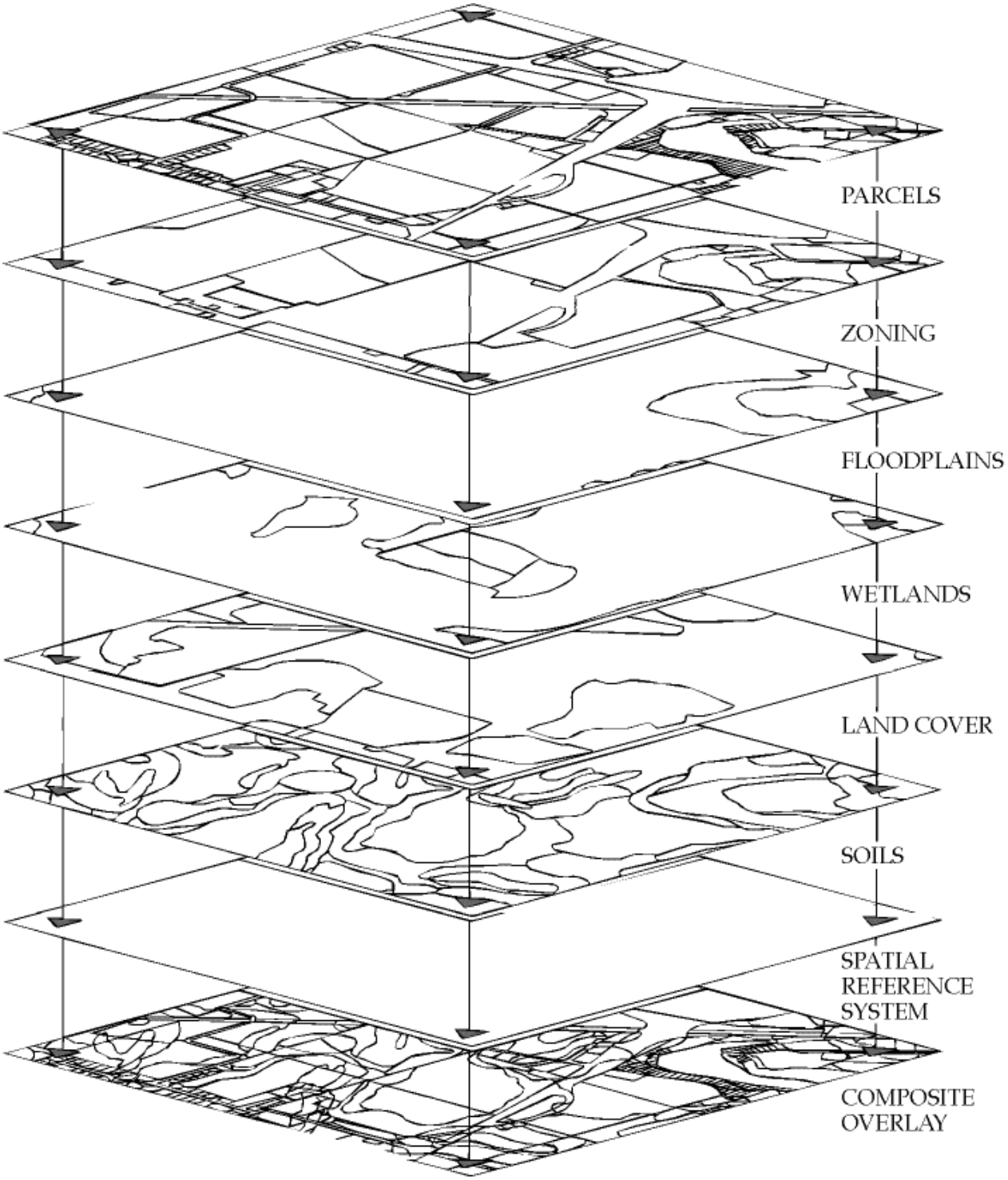


Figure 3: Concept for a Multi-Purpose Land Information System (1984)
Source: Sullivan, Chrisman and Niemann (1984)

The layer cake diagram presented in Figure 3 is one of many produced in approximately the same period (1984), and copied again and again ever since. This particular one was perhaps the first produced with actual data layers, in this case, Section 12 (one square mile) in Westport Wisconsin. It was originally produced in a press run of 20,000 on card stock to be handed out at the Wisconsin State Fair to explain multipurpose land information systems to the farming community and the general public. The text on the reverse was titled "Conceptual Model of a Multipurpose Land Information System", perhaps a bit highbrow for the average State Fair booth, but received with substantial interest when I served my stint at the booth handing them out. The text did make some mention of computing technology, after all, the state motto is "Progress". However, the primary emphasis was on the independence of the sources of various layers of information. This layer-cake diagram was about the institutional relationships as much as it was about the overlay procedure.

At that time, a certain element in the community was proposing a multipurpose cadastre as the primary direction for implementation (National Research Council, 1980). That concept emphasized the property parcel as the base unit, much as the ITUM also relied on a single base polygon- but one that was derived to fit the circumstances. The NRC panel proposed to make all other variables simply an attribute of parcels. Thus, land use: attribute of parcel; floodplains:

attribute of parcel; wetlands: attribute of parcel; agricultural productivity: attribute of parcel and so on. They demonstrated this with a layer cake diagram showing the parcel layer as the base on which everything was built. The alternative layer cake presented in Figure 3 was part of an attempt to demonstrate that one particular layer did not need to be the base. The geodetic framework (blank!) was provided as the means to integrate. The diagram was constructed in the first years of the GPS constellation, then just a few satellites with a few hours of observation a day. The potential was already clear.. Viewed from 2014, the geodetic underpinning is even more important.

Thus, this diagram may look obvious now, but it was contentious when constructed. The land systems concept from CSIRO had much more to offer than simply selecting one layer like the parcels. Land ownership is important, but it does not define all map boundaries so that everything is an attribute of parcels. The land units, however, were refined and tuned to suit the situation by the domain experts. Originally developed for rapid appraisal in an overview of huge areas, it developed into a niche tool for expert photointerpreters to develop many internally consistent land inventories without needing the polygon overlay tool. It was actually technically superior to use ITUM technology up to 1979.

5 Discussion

Playing the origins game is hardly very fruitful in this case. Both overlay and integrated terrain units (ITU) have deep historical roots. Both contributed to the emergence of the current digital technology. Yet, it is clear that the integrated terrain unit mapping technique did not become dominant. There are complex reasons why layer-based logic and polygon overlay procedures took the lead. It is too common to ascribe the reasons to a simple technological march of progress where the past is simply a prologue to the age of enlightenment.

It would be hard to argue that polygon overlay procedures are more "accurate". Since the earliest days of CGIS, it became apparent that overlay produces a flood of slivers [small objects induced by slight differences between two boundaries] (Goodchild 1978; Chrisman 1987a). The ITU requires all such disparities to be resolved in the compilation phase, a phase which engages experts and human interpretation. This procedure introduced a division of labor, and a division of knowledge. While ITU kept the compilation phase under disciplinary control, the polygon overlay software displaced this effort to the user and the uncontrolled vagaries of software packages. This user was meant to resolve the disagreements between the various source layers as a part of their analysis. Of course, the user typically has little knowledge that such slivers are even there, and does not understand the procedures that would have to be applied to resolve each kind of sliver appropriately. Software packages took charge and resolved complex issues with simplistic heuristics. The multiple tolerance overlay has been discussed in the research literature for a number of years (Dougenik 1980; Pullar 1991; 1993; Harvey 1994), without breaking into the mainstream of user software.

The strongest reason for the overlay approach is that it matched administrative hierarchy, with its implicit divisions of labor and responsibility, and the divisions of knowledge between disciplines and communities of practice. The Dane County layer cake (Figure 3) represents the organizations currently making the maps, and accepted their several responsibility. The concept of ‘mandates’ and ‘custodians’ derived from this project (Chrisman, 1987b). The overlay procedure is presented as a final step simply to produce the analytical product. Data quality issues would not be apparent until you ask certain questions.

This contrasts with ITU were the expert compilation will alter all sources to bring them into coherence with each other in the preparation phase, upfront. Agencies are much more likely to associate with a federation in which they retain autonomy and control over the parts they consider to be theirs. The concept of "custodians" of data layers came from this administrative logic, not any particular technical merit. There are lots of strong reasons to support ITU, but they are likely to lose to the impressive solidity of the administrative reasoning behind custodians managing their individual layers. The division of labor and the division of knowledge is exactly the center of the design process.

In a study of an early computer company building a new model, Woolgar (1991) saw a particular set of social relationships where the developers did not particularly design the computer, rather they attempted to configure the user. The layer-cake model takes the overlay algorithm as the base, and then configures the administrative and professional world around that capability. In much the same way, the integrated survey set the goal as a single map, and made all the activity fall around that choice.

6 Conclusion

This paper therefore concludes by evoking social factors that often determine the nature of the GIS and GIS products. In particular, a layer-based design has strong support from the administrative divisions of labor and knowledge. This may not be particularly surprising, but it continues the demonstration that the divisions between GIS and society are perhaps not drawn in the right places, and might be impossible to draw at all.

Drawing the lines between what is the "technical" part and what is the responsibility of the less- sophisticated user is a frontier of substantial interest for future research. In study of another kind of software, Rachel and Woolgar (1995) noted that the key element in locating what was considered "technical" was who got to make their decisions first. In their business software organization, the programmers decided things then told the documentation team that the decisions were "technical", meaning mostly that they were already made. In the GIS situation, the roles may be somewhat more subtle, but the effect of time and priority of decision-making still will be important.

Acknowledgements

This research is supported by RMIT University, School of Mathematical and Geospatial Sciences. Elements derive from prior work supported by US National Science Foundation, Canada Social Sciences Research Council and Networks of Centres of Excellence.

References

Burrough, P. A. (1986). *Principles of Geographical Information Systems for Land Resource Assessment.* Oxford: Clarendon Press

Chrisman, N. R. (1982). Methods of spatial analysis based on error in categorical maps. unpublished Ph.D. thesis, University of Bristol.

Chrisman, N. R. (1987a). Accuracy of map overlays: a reassessment. *Landscape and Urban Planning 14:* 427-439.

Chrisman, N. R. (1987b). Design of information systems based on social and cultural goals. *Photogrammetric Engineering and Remote Sensing 53:* 1367-1370.

Chrisman, N. R., J. A. Dougenik, and D. White. (1992). Lessons for the design of polygon overlay processing from the ODYSSEY WHIRLPOOL algorithm. *Proceedings, 5th International Symposium on Spatial Data Handling, 2,* 401-410.

Chrisman, N.R. (2006). *Charting the Unknown.* Redlands: ESRI Press.

Christian, C. S. (1952). Regional land surveys. *Journal of the Australian Institute of Agricultural Sciences, 18: 140-146.*

Christian, C. S. and Stewart, G. A. (1968). Methodology of integrated surveys. *Aerial Surveys and Integrated Studies 233-280.*

Cloud, J. (2002). American cartographic transformations during the cold war. *Cartography and Geographic Information Science 29:* 261-282.

Dangermond, J. (1979). A case study of the Zulia Regional Planning Study, describing work completed. In Moore, P.J. (ed.) *Harvard Library of Computer Graphics, 3,* p. 35-62. Cambridge MA: Harvard Laboratory for Computer Graphics.

Davies, T., Edwards, D. (2012). Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *IDS Bulletin, 43*(5), 117-127, doi:10.1111/j.1759-5436.2012.00372.x

Deleuze, G., and Guattari, F. (1976). *Rhizome introduction*. Paris: Editions de Minuit.

Dougenik, James A. 1980: WHIRLPOOL: a geometric processor for polygon coverage data. *Proceedings, AUTO-CARTO IV* pp. 304-311.

FAO (1976). A *Framework for Land Evaluation.* Rome: Food and Agriculture Organization.

Foresman, T. (ed.). (1998). *History of geographic information systems: Perspectives from the pioneers.* London, U.K.: Taylor & Francis.

Garrison, W. L., R. Alexander, W. Bailey, M.F. Dacey, and D. F. Marble. (1965). Data system requirements for geographic research. In: *Proceedings, Third Goddard Memorial Symposium.* pp. 139-51.

Goodchild, M. F. (1978). Statistical aspects of the polygon overlay problem. In: G. Dutton (ed.), *Harvard Papers on Geographic Information Systems 6,* pp. 1-22. Reading, Massachusetts: Addison Wesley.

Harvey, F. (1994). Defining unmovable nodes/segments as a part of vector overlay: the alignment overlay. *Spatial Data Handling 94 1:* 159-176.

Harvey, F., and Chrisman, N.R. (2004). The imbrications of geography and technology: The social construction of geographic information systems. In: S. Brunn, S. Cutter, and J. W. Harrington Jr. (eds), *Geography and Technology.* p. 65-80, Amsterdam: Kluwer Academic Press.

Hills, G.A. (1966). The classification and evaluation of land for multiple uses. *Forestry Chronicle 42:* 21-25.

Hopkins, L.D. (1977). Methods of generating land suitability maps: A comparative evaluation. *American Institute of Planners Journal, 43:* 386-400.

Landscape Architecture Research Office, Graduate School of Design Harvard University. (1967). *Three approaches to environmental resource analysis*. The Conservation Foundation, Washington D.C.

Lewis, P. (1963). *Recreation in Wisconsin.* State of Wisconsin, Department of Resource Development, Madison, Wisconsin.

Mabbutt, J. A. (1968). Review of concepts of land classification. In Stewart, G.A. (ed.) *Land Evaluation,* p. 11-28. Melbourne: Macmillan.

Mabbutt, J. A., and G. A. Stewart. (1963). The application of geomorphology in resources surveys in Australia and New Guinea. *Revue de Géomorphologie dynamique* 14(7-8-9): 97-109.

Manning, W. (1913). The Billerica town plan. *Landscape Architecture* 3: 108-118.

McHarg, I. L. (1969). *Design with nature*. Garden City, New York: Natural History Press.

National Research Council, (1980). *Needs for a multipurpose cadastre*. National Academies Press, Washington DC.

Pullar, D. V. (1991). Spatial overlay with inexact numerical data. *Proceedings, AUTO- CARTO 10* 313-329.

Pullar, D. V. (1993). Consequences of using a tolerance paradigm in spatial overlay. Proceedings, AUTO-CARTO 11 *288-296.*

Rachel, J. and Woolgar, S. (1995). The discursive structure of the socio-technical divide: the example of information systems development. *The Sociological Review* **43:** 251-273.

Robinove, C.J. (1979). Integrated terrain unit mapping with digital Landsat images of Queensland, Australia. *USGS Professional Paper 1102.* Reston: USGS.

Steinitz, C., Parker, P. and Jordan, L. (1976). Hand-drawn overlays: Their history and prospective uses. *Landscape Architecture* 66: 444-55.

Stewart, G. A. (ed.). (1968). *Land evaluation.* Melbourne, Australia: Macmillan.

Story, R., Galloway, R. W., van de Graaf, R. H. M. and Tweedie, A. D. (1963). General report on the lands of the Hunter Valley. *Australian Land Research series 8.* Canberra: CSIRO.

Sullivan, J.G., Chrisman, N.R., and Niemann, B.J. (1984). Concept of a Multipurpose Land Information System. Madison: University of Wisconsin.

Tomlin, C. D. (1990). *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs NJ: Prentice Hall.

Tomlinson, R. F. (1968). A geographic information system for regional planning. In: G. A. Stewart (ed.), *Land Evaluation*. Melbourne, Australia: Macmillan. pp. 200-210.

Tomlinson, R. F. (1974). The application of electronic computing methods to the storage, compilation and assessment of mapped data. unpublished Unpublished Ph.D., University of London.

Tomlinson, R. F. (1984). Geographic information systems - a new frontier. *Proceedings of the International Symposium on Spatial Data Handling* **2:** 1-14.

Tomlinson, R. F. (1989). Presidential Address: Geographic information systems and geographers in the 1990s. *Canadian Geographer 33:* 290-298.

Tomlinson, Roger F., and A. Raymond Boyle. (1981). The state of development of systems for handling natural resources inventory data. *Cartographica 18:* 65-95.

Tyrwhitt, J. (1950). Surveys for planning. In: Association for Planning and Regional Reconstruction (ed.), *Town and Country Planning Textbook*. London: Architectural Press.

Woolgar, S. (1991). Configuring the user: the case of usability trials. In Law, J. (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination,* p. 58-97. London: Routledge.

Zonneveld, I. S. (1972). Land evaluation and landscape science. Chapter 7 in *ITC textbook of photo-interpretation*, ITC: Enschede, The Netherlands.

The Geography of World War I Cartoons: Gallipoli

Antoni Moore
School of Surveying
University of Otago, PO Box 56, Dunedin, New Zealand
tony.moore@otago.ac.nz

William Cartwright
School of Mathematical and Geospatial Sciences
RMIT University, PO Box 2476, Melbourne, Australia
william.cartwright@rmit.edu.au

Abstract

We have traditionally used maps to provide information about space. We have fashioned a design, development, fabrication and consumption process (and associated procedures) that have enabled essential artifacts to be made available and for them to be used effectively and efficiently. However, with the use of non-traditional representational artefacts , whereby for example users can disassociate the source of information from the actual display of that information, the consideration that these artefacts may be required to provide information not just about 'SPACE', but also information about 'PLACE'. This paper reports on from the initial stages of research that is investigating at cartoons from the First World War, and particularly those cartoons that relate to the Gallipoli campaign of April 1915 – January 1916. Cartoons from this period are being investigated to ascertain their potential value as alternative, more personal, sources of information about the perceptions of the geography of the Gallipoli Peninsula of soldiers in the field, their commanders, politicians and the media.

Keywords: geography, cartoons, World War I, Gallipoli

1.Introduction

Events in history, each having their own relative importance and impact on the world of today, are recorded objectively as fact as far as being possible (notwithstanding the adage that history is written by the victor). However, for many of these events, particularly the most significant ones, there are artistic reactions to what is being played out. The cartoon, which could be regarded as an extension of fine art (the term used to be reserved for full scale preparatory sketches for paintings), is a medium that can render a complex and subtle historical event in easily understandable terms (in fact it is this accessibility that has contributed to its lowbrow status – McCloud, 1993) as well as hitting home with a sharp perspective on the situation being depicted.

As a collated group, these artworks can reveal further information about the event and times they originated from through thematic and other patterns elicited from the collective. Furthermore, all of these historical events took place at a location and so geography is a dominant theme to be elicited either explicitly or implicitly from art.

Here, our research is being directed to ascertain the geography depicted in War cartoons related to the Gallipoli campaign of April 1915 – January 1916. Many reports of this campaign deliver their content through words,

In: S. Winter and C. Rizos (Eds.): Research@Locate'14, Canberra, Australia, 07-09 April 2014, published at http://ceur-ws.org

photographs, and maps. These provide, generally, succinct, but impersonal narratives about what happened during the campaign in the Dardenelles. If one wants to better understand personal impressions of the campaign these formal reports will not provide insight into what it was really like to be there and the personal impact of the war on those directly involved and those indirectly affected by the conflict and those that support soldiers on the front line – in the theatre of war, at home and in political and social mechanisms that decide the fate of soldiers from afar. Therefore, by investigating the cartoons produced during the Gallipoli campaign we seek to ascertain how representations of geography (in their various forms) in these cartoons were used to support particular messages and how understanding the geography that were adjuncts to these messages provides insight into the personal, national and international perspectives about that campaign.

We are concerned with underlying geographies explicitly and implicitly contained within war cartoons associated with the World War I battle of Gallipoli. We seek to ascertain the differences between perspectives and the differences between contexts.

2.Gallipoli 1915

The Gallipoli campaign of 1915-16 came about because of the deadlock on the Western Front, which turned British eyes towards other possible theatres, plus appeals for assistance from Russia early in January 1915 (Travers, 2001, Velsley, 1997). A plan preferred by Winston Churchill, then the First Lord of the Admiralty was to be a naval operation (Heffernan 1996). The original plan was for a combined Anglo-Franco naval fleet, using mainly outdated battleships; to force the Narrows, sail into the Sea of Marmara and then on to Istanbul. Once this was done, three Divisions of the Greek Army would advance on Istanbul.

The entrance to the Dardanelles and the Narrows is shown in the map in figure 1.

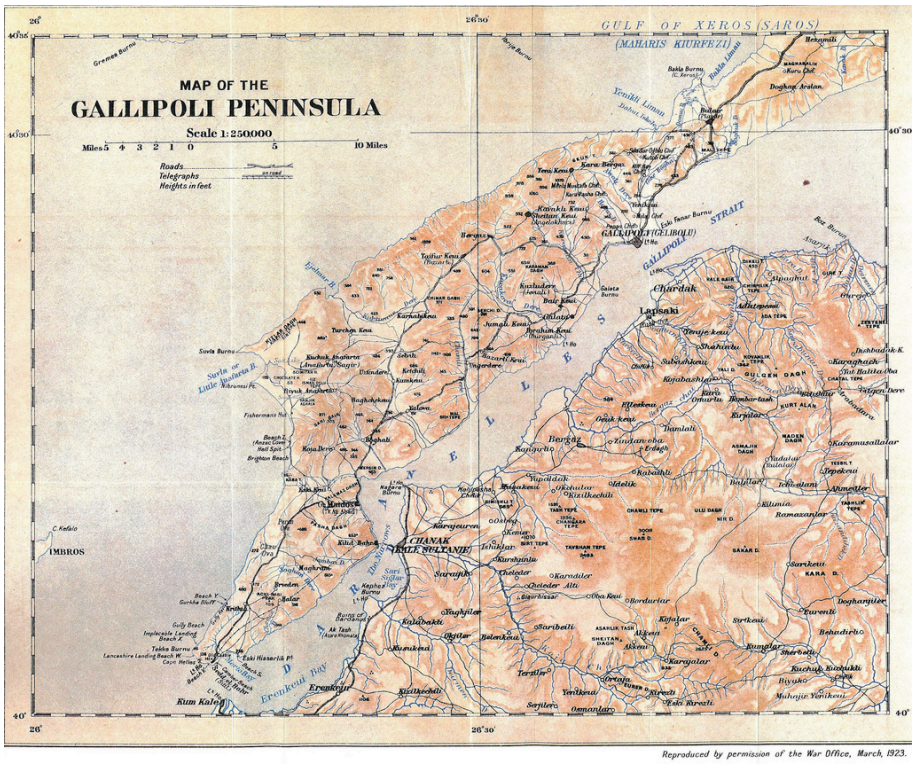


Figure 1: Map of the Gallipoli Peninsula. Source: http://farm8.staticflickr.com/7187/7090070097_7b1f3e5be1_b.jpg

However, this was later amended to be a naval engagement, after Russian opposition to the use of Greek troops. The revised naval plan was to force the Narrows, penetrate the Sea of Marmara and bombard Istanbul, compelling Turkey to surrender (Sea Power Centre, 2005). The opening attack began on February 18, 1915m (Corbett, 1921). The map in figure 2 shows the bombardment plan.

Figure 2: Map showing the bombardments of Turkish forts, 19 February 1915.
Source: Corbet (1921) http://www.naval-history.net/WW1Book-RN2-143.JPG

Forcing of The Narrows was attempted on March 18, 1915, when seventeen allied warships, supported by an assortment of other craft, like mine sweepers (Millett, 2002). Mine fields and hidden guns prevented the success of this plan. Six battleships were sunk or severely damaged (Millett 2002). Some military analysts considered that this plan could not have eventually worked anyway. *"I am still of the opinion however, that the Royal Navy could not have "rushed" the Narrows and go through in sufficient numbers to tackle the hostile fleet it would have met in the Sea of Marmara"* (Aiguillette, 1962, p. 63).

Then Britain prepared another plan, for a larger military operation that would capture the Gallipoli Peninsula, allow the waters to be cleared of mines and opening it for the fleet to sail to Istanbul (Sea Power Centre, 2005). To support the military operation, France provided a Division (the First Division of the *Corps Expédition d'Orient* made up of North African (Arab and European), Foreign Legion and Senegalese troops (Hughes 2005), Britain its 29th Division, and Australian and New Zealand troops were moved from Egypt (Travers, 2001). The stage was set for the invasion and subsequent landings on the beaches of the Turkish Gallipoli (Chanakale) Peninsula by British, ANZAC and Indian troops and at Kum Kale (on the Asiatic shore) by French troops (who acted as a diversionary force by capturing a Turkish fort on the Eastern shores of the Dardanelles (Millett 2002)) (who were moved to Cape Helles on 26 April, where they held the eastern part of the Allied line) (Hughes, 2005) on April 25, and the Allied attacks of 28 April at Helles ,1915 (Travers, 2001b).

The Australasian landings took place at "Beach Z" and the Anglo French landings were at Cape Hellas, to the south. The Allied forces fought ashore, but were unable to seize the strategic heights that dominate the lower third of the Peninsula (Millett 2002). The campaign extended over a 10-month period (Millett 2002), until the final evacuation in January 1916 (Mason 1936; Millett 2002).

3.Gaining information about personal geography through alternative representations

The geographies depicted in cartoons should be considered to be 'naive' geographies, where a simplistic interpretation (and subsequent representation) of geography is offered. Naive geography was defined by Egenhofer and Mark (1995) as "the body of knowledge that people have about the surrounding geographic world" – the primary theories of space, entities and processes (Mark and Egenhofer, 1996). This was also described as being "…

captures and reflects the way humans think and reason about geographic space and time. Naive stands for instinctive or spontaneous" (Egenhofer and Mark, 1995, p. 4). Representations of naïve geographies offer the prospect for better learning about, and therefore understanding different geographies. In the context of this research we seek to ascertain how the geography of the Gallipoli campaign (physical, political and personal) might be better understood through the interpretation of cartoons from that period.

In this context, we take 'cartoon' to encompass all drawn graphics that represent situations in a simplified style. The characteristics of this style are relatively sparse lines (compared with a drawing that aims to represent something with realism), strong outlines and a simple palette of colours (if used), applied straightforwardly. Most will be drawn to depict a humorous situation, the humour derived from a real event, and if so, is probably satirical or ironic in nature (this represents the definition of cartoons, as described by Kleeman, 2006). However, we argue that also permissible are some of the graphics drawn in this style that are used for propaganda or straight diary-like purposes by the individual. Most cartoons will not be drawn directly from observation, but will have differing amounts of "true" features present. There is license to exaggerate and use caricature in cartoons. Some cartoons are effective with just graphics but most will have some text that either is used in tandem with the graphic to form the message of the cartoon, or even convey most of the message, with the graphic effectively just a supporting sketch (McCloud, 1993).

4.Classification of relevant cartoons

As a foundation for classifying cartoons related to the war, and Gallipoli in particular, a survey was undertaken to uncover the extent and the type of cartoons drawn that were related to this topic. In selecting cartoons useful to this research, cartoons were only selected if they had a 'geographic' element. As well, we sought to find differences between the European view of the campaign and the colonial (here, the Australasian) viewpoint. All of the cartoon examples described in this section come from the Allied (UK / Australia / New Zealand) point-of-view. With one notable exception, they do not explicitly reflect on the colonialism context. Further exploratory research will endeavor to unearth cartoons that can be used to study the representations produced (in cartoons) of both views of the campaign.

The cartoons found have been initially classified as:

- Propaganda Cartoons
    - with Geography contained in text;
    - with Literal Geography depicted in Graphics;
    - with Geography derived from Symbolic Graphic Element;
    - with Geography derived from Visual Metaphor; and
    - Cartoon with Map or Map-Related Object.
- Satirical Cartoon
    - with Geography contained in text;
    - with Literal Geography depicted in Graphics;
    - with Geography derived from Symbolic Graphic Element;
    - with Map or Map-Related Object;
    - with Geography contained in text; and
    - with Literal Geography depicted in Graphics.
- Personal Cartoon
    - with Geography derived from Symbolic Graphic Element;
    - with Geography derived from Visual Metaphor;
    - with Map or Map-Related Object.

The following sections of the paper provide some examples of some of these classifications and, where appropriate, provide examples – photograph or map - of the geography represented in the cartoon. This was done to provide some 'ground truthing' (Cartwright et al., 2001), to our collection of cartoons.

The first example (figure 3a) is a Propaganda Cartoon with Geography contained in the text. The example does not graphically depict a 'mappable place, though it implicitly signifies any place that the Kaiser and the Sultan ever

met, specifically just after Gallipoli. It is possible that such information was recorded. However, the cartoon's placement in this category derives from a second geographical clue in the caption, the Sultan's assertion that the Turkish Army has driven the English Army (and by extension the Anzacs) "into the sea". This can be implicitly linked to the known (and geographically explicit) embarkation point of the retreating army, at Anzac Cove. Its inclusion in the propaganda perspective derives from it perversely trying to find a bright side of what was a disastrous campaign for the Allies, even from a satirical publication such as *Punch*. The picture in figure 4b ground truths the cartoon into the Place of Gallipoli, showing the reality of the evacuation. (The Allied forces were evacuated from the Gallipoli Peninsula by January 1916.)

Figure 3a: Propaganda Cartoon with Geography contained in text From Punch magazine. Bernard Partridge, January 1916. Source: http://img69.imageshack.us/img69/2162/cctlwwkkgrhqfiee0eubdhb.jpg

Figure 3b: W Beach (Lancashire Landing) at Cape Helles, Gallipoli, 7 January 1916, just prior to the final evacuation of British forces. Source: http://upload.wikimedia.org/wikipedia/commons/9/92/W_Beach_Helles_Gallipoli.jpg

There may be an implicit signifier of the relationship with colonial armies, in that the Anzacs are not mentioned. Then again, in the vast majority of the Australia and New Zealand drawn cartoons, the English are not mentioned either. This may be for the sake of brevity, enhancing the impact of the cartoon by not getting enmeshed with the details of all contributing armies.

The cartoon in figure 4a moves us to the actual geography of the campaign. We classify this as a "Propaganda Cartoon with Literal Geography depicted in Graphics". It is notable for graphical clues of geography, depicting a part of the battleground. Though the painted drawing of a hill with the sea behind is probably generic, an abstracted geography of the peninsula with some implicit truth is depicted. It is not just "a" hill but "the Hill", with this geographic content in the text emphasising the graphical content through repetition. The "Hill" ties the cartoon to a specific location.

Figure 6a also works graphically on a similar abstracted level. There is a depiction of a generic trench battle between an Anzac soldier and two Turkish soldiers – this is enough to geographically fix this scene in the trench network of the Gallipoli campaign. However, we get extra information as to the identity of the soldier through the symbolic depiction of his Maori ancestor. This not only singles him out as a New Zealander, a geographical refinement, but also states an implicit linkage with the battles that the Maori took part in on New Zealand soil. [There is an irony in this as the Maori were defending their homeland in those 19th Century wars, much like the Turkish are in this cartoon.]

The use of drawing to convey the geography in Figure 6a has much the same function as the propaganda equivalent in Figure 4a. We have a scene on the coast conveying the Anzacs' retreat from Gallipoli, affording the definition of a specific location for this cartoon. There is an expression of colonialism in the restraint of the Anzac soldier by the British Army officer, and the implicit geographies that suggests

Figure 4a: Propaganda Cartoon with Literal Geography depicted in Graphics Source: http://img69.imageshack.us/img69/2162/cctlwwkkgrhqfiee0eubdhb.jpg

Figure 4b: Landing on the beach at Kapa Tepe, Gallipoli Peninsula, Turkey. McKenzie, Fiona, fl 2004: Photographs relating to Charles and Christina Andrews. Ref: PAColl-8147-1-08. Alexander Turnbull Library, Wellington, New Zealand.http://natlib.govt.nz/records/22453227

Figure 5a: Propaganda Cartoon with Geography derived from Symbolic Graphic Element. Source: http://img69.imageshack.us/img69/2162/cctlwwkkgrhqfiee0eubdhb.jpg

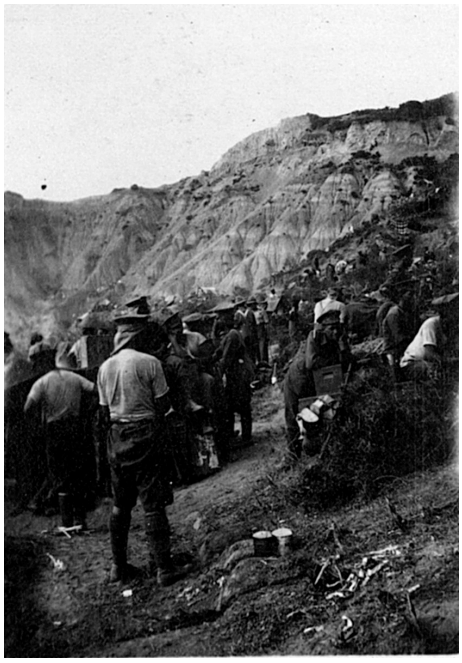


Figure 5b: Maori soldiers at Gallipoli. Source: http://ww100.govt.nz/sites/default/files/images/005%20Maoris-crop.jpg

Figure 6a: Satirical Cartoon with Literal Geography depicted in Graphics Source: http://img69.imageshack.us/img69/2162/cctlwwkkgrhqfiee0eubdhb.jpg

Figure 6b: Panoramic photograph Suvla Gay, Gallipoli. Source: German Military Archive. http://upload.wikimedia.org/wikipedia/commons/0/0b/Bundesarchiv_Bild_183-R36255%2C_T%C3%BCrkei%2C_Dardanellen%2C_Suvla-Bucht.jpg

Next is a propaganda "cartoon" that explicitly features maps or map-related objects. While there are a large number of recruitment posters that depict the Dardanelles as a map that we could have chosen from for this example, in this case we have an early film animation, a frame of which is shown in Figure 7a. The animation starts with a map of Australia and New Zealand, the island objects of which are moved to create the anthropomorphic representation in the frame capture. On its side, Australia itself serves as the head and neck of the 'person', with Cape York making an effective nose. The North Island of New Zealand moves to suggest a hat, with the South Island and Tasmania placed to represent hair (in fact, all three smaller islands are placed to imply a larger hat, which can be perceived with closure). The 'To Turkey' sign is reminiscent of the many signs erected at Gallipoli – officially or self-made by the troops. The sign illustrated in figure 7b is a typical example.

Figure 7a: Propaganda Cartoon with Map or Map-Related Object: "Miss Australasia.
Source: National Film and Sound Archive, Australia
http://aso.gov.au/titles/newsreels/miss-australasia/clip2/

Figure 7b: Trench sign to Stinking Farm. Source: Australian War Memorial. RELAWM06263. http://www.awm.gov.au/sites/default/files/IMG_0285-550x151.jpg

The cartoon in figure 8a is an example of a Satirical Cartoon with Geography derived from Symbolic Graphic Element. It uses a symbolic burrow complete with rabbit to depict the harsh conditions of trench life in the battle. As such, the geography on the ground represented is that of the trenches, whose location and geography are known. When comparing the cartoon to the photograph in figure 9b, showing soldiers from the 14th Battalion 'dug-in' in the hillside at Gallipoli, it can be seen that the cartoon accurately depicts life for the infantry soldier during the campaign.

The final satirical cartoon (figure 9a) uses a map rendering of Gallipoli (complete with labelling) as a backdrop to a war of words between Turkey and New Zealand. A turkey and a kiwi represent the countries metaphorically and there is metaphor in the kiwi's use of the word "yard" to represent country, evoking a sense of place. There is an implicit criticism of New Zealand being aggressors, though the cartoon could also be interpreted as defiant propaganda. Here the geography of the Peninsula in the cartoon does faithfully include many geographical features from the Peninsula and shows 'The Narrows', in the photograph populated by Allied warships.

Cartoon by David Low. First published in *The Bulletin*, 23 January, 1919.

Figure 8a: Satirical Cartoon with Geography derived from Symbolic Graphic Element. Source: Source: Australian Government. Department of Veterans Affairs. http://www.dva.gov.au/aboutDVA/publications/commemorative/awf/Pages/topic10.aspx

Figure 8b: Members of the 14th Battalion at Gallipoli, 1915. Source: http://upload.wikimedia.org/wikipedia/commons/3/31/A03803_14th_Battalion_AIF_Gallipoli.jpg

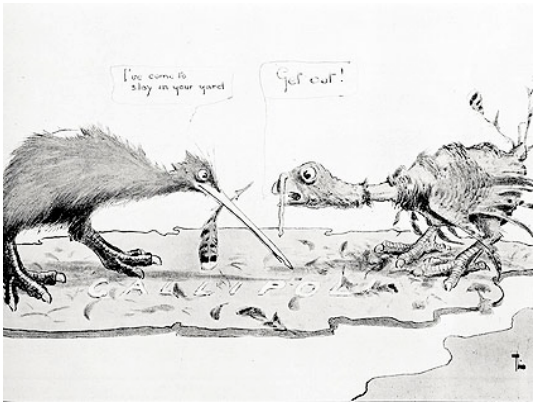


Figure 9a: Satirical Cartoon with Map or Map-Related Object. Source: Source: http://www.aucklandlibraries.govt.nz/EN/Events/ExhibitionsSlideshow/slideshow/images/gallipoli.jpg

Figure 9b: French troops on the heights at Gallipoli Source: http://www.warhistoryonline.com/wp-content/uploads/2013/03/cc33bc0896c6808ab4b36cec7dbd0b47_1M.png

Figure 10a: Personal Cartoon with Map or Map-Related Object. Source: http://img69.imageshack.us/img69/2162/cctlwwkkgrhqfiee0eubdhb.jpg

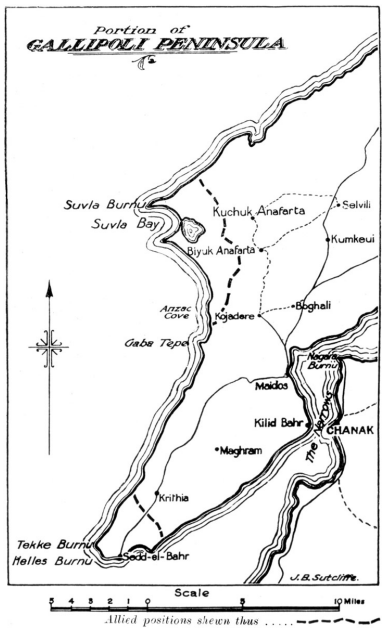


Figure 10b: Map by Australian War Museum showing allied lines at the time of the evacuation. Source: http://www.gutenberg.org/files/25341/25341-h/images/imagep059.jpg

Cartoons also sometimes included representations of maps. The cartoon in figure 10a includes a map showing Cape Hellas and the extent of British advances at the end of the campaign. When compared to the map in figure 10b, it is obvious that maps were used as a reference by the cartoonist. The extent of the British advances at Cape Hellas, shown on the map in figure 10b by the dashed line at the southwestern corner of the Peninsular is shown on the cartoon by drawings of Turkish emplacements and a camp. Regarding the actual evacuation itself, when

comparing the geography in the cartoon to the photograph in figure 4b of the evacuation from W Beach, it can be seen that the actual curve of the bay at Cape Hellas is faithfully depicted in the cartoon.

5.Developing a structure for extracting geography from cartoons

In order to structure the methodology for extracting the geography from cartoons a conceptual structure was developed by the authors. Figure 11 shows the conceptual structure that will guide the deconstruction of the cartoons collated. There are three axes to the structure (i.e. it is 3D); each will be dealt with in turn.

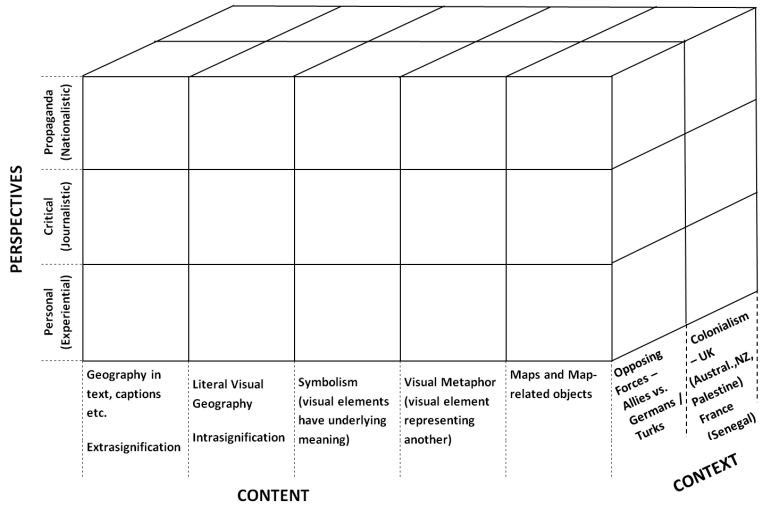


Figure 11: Conceptual structure for cartoon deconstruction

Firstly we have a categorisation of geographic content, which can be divided into elements of intrasignification and extrasignification. Starting with the latter, extrasignification covers the non-graphical cartoon elements that contribute to the main group of graphical elements, mainly text such as captions and titles. Intrasignification is achieved through the drawn element of the cartoon and that graphical component is here divided into four subcategories, drawing in part (along with the text category) from Kleeman's (2009) structure for extracting the geographic content of cartoons.

Moving from the least abstracted to the most abstracted, there is the literal depiction of natural and manmade geographic features in cartoons. These features may be unique (a hill of distinguishing topography and morphology; a striking building), enabling a more precise geography to be extracted, or generic, which leads to the production of a more uncertain geography (e.g. a non-descript building or unremarkable hill would be enough to link the cartoon to built-up areas or hilly topography, but there is no further refinement to specific regions or locations than this).

The next graphical content category is visually identical to the first, but is reserved for features in the cartoon that are symbolic of some larger entity or issue as well as being literal representations of geography. In applying this visual symbolism, features may be modified in some way, for example, exaggeration to make a feature larger than it normally is, or the use of caricature (if, for example, a soldier of a particular army is used to represent that entire army in the cartoon, caricaturing may be applied, perhaps pandering to commonly-held stereotypical appearances of the time).

This notion of a visual element representing another element (or a group) is abstracted further with the use of visual metaphor. Metaphors are used to facilitate the communication of difficult or complex concepts or ideas in a way that is easy to understand. War cartoons may use visual metaphor to great effect as shorthand for ideas, ideologies that are otherwise invisible and intangible.

The final content category is for cartoons with explicit map content present to differing degrees. Maps represent visual elements subject to the greatest amount of abstraction and are the richest in geographic content, having the

potential to pinpoint a location of cartoon focus with accuracy. They may be present as explicit maps or 'cartoonised' map forms that constitute a backdrop to cartoon protagonists.

6.Further research

In the deconstruction of cartoons it is anticipated that there will be explicit (e.g. the content of maps, explicitly named locations in text) and implicit (e.g. distinctive uniforms link soldiers to a specific army from a specific country) signifiers of geography, a consideration made by other efforts to extract the geographic component of art and literature (e.g. the Literary Atlas of Europe, Reuschel et al, 2009). This will require further explorations into cartoon collections and publications (currently underway) and further refining the research model. As well, automated routines will be developed to uncover and deconstruct cartoons related to this campaign.

7.Conclusion

This paper has provided an overview of the research currently being undertaken by the authors. It gave an overview of the Gallipoli campaign of April 1915 – January 1916 and how the initial sea warfare strategy was changed into a combined sea/land operation. It was found that the cartoons sourced thus far could be classified into three general categories, viz:

- Propaganda Cartoons;
- Satirical Cartoon; and
- Personal Cartoon.

This general classification was then further subdivided into sub-classifications, namely:

- with Geography contained in text;
- with Literal Geography depicted in Graphics;
- with Geography derived from Symbolic Graphic Element;
- with Map or Map-Related Object;
- with Geography contained in text;
- with Literal Geography depicted in Graphics; and
- with Geography derived from Visual Metaphor.

It then provided information about some of the cartoons sourced as part of this initial stage of research. It also elaborated on how we view the cartoons samples selected, as illustrations of this category of cartoon.

The research has also looked at the depiction of geography in the sample cartoons included in this paper. When comparing the geography included in the cartoons with the actual geography of Gallipoli (here shown in archival photographs) it was seen that the cartoonists when developing their drawings considered the actual geography.

As well, navigational artefacts like signposts, placed throughout the Peninsula in trenches and on pathways – official or hand-made by troops – were used in some cartoons to give a sense of Place to what was illustrated.

This classification of cartoons will be used as a foundation structure for extracting geography from cartoons in a formal manner, as described in the penultimate section of the paper.

8. **References**

Aiguillette. (1962) "Now. If We Had Air Drops: "Aiguillette" Applies Hindsight", *Marine Corps Gazette (pre-1994);* May 1962; 46, 5, p. 63; ProQuest Military Collection

Cartwright, W., Crampton,, J., Gartner,, G., Miller,, S., Mitchell,, K., Siekierska,, E. and Wood,, J. (2001) "User Interface Issues for Spatial Information Visualization", CaGIS, vol. 28, no. 1, pp. 45 – 60.

Corbett,J. S. (1921) History of the Great War, Naval Operations, Vol. II. http://www.naval-history.net/WW1Book-RN2a.htm#IX. Web page accessed December 2013.

Egenhofer, M. J. and Mark, D. M. (1995) "Niave Geography". In Frank, A. U. and Kuhn, W., eds, Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 988, pp. 1 -15.

Heffernan, M. (1996) Geography, Cartography and Military Intelligence: The Royal Geographical Society and The First World War *Transactions of the Institute of British Geographers NS 21.*

Hughes, M. (2005) The French Army at Gallipoli. *The RUSI Journal,* 150, 64-67.

Kleeman, G. (2006) Not just for fun: Using cartoons to investigate geographical issues. New Zealand Geographer, 62(2): 144-152

Kleeman, G. (2009). Through the eyes of others: The role of curriculum perspectives in Australian school Geography. Geographical Education. 22: 18-27.

Mason, A. T. (1936) An Introduction to the Gallipoli Campaign. *Marine Corps Gazette (pre-1994),* 20.

Mark, D. M. and Egenhofer, M. J. (1996) "Common-sense geography: foundations for intuitive Geographic Information Systems", paper presented at GIS/LIS '96, 6 pp.

McLeod, S. (1993) Understanding Comics. William Morrow Paperbacks

Millett, A. R. (2000) Most Significant Amphibious Operation: Invasion of Gallipoli. *MHQ : The Quarterly Journal of Military History,* 12, 2.

Reuschel, Piatti, B., and A-K, Hurni, L. , W. (2009) " Modelling Uncertain Geodata for the Literary Atlas of Europe", in Understanding Different Geographies, Kriz, K., Cartwright, W. E. and Kinberger, M. (eds), Heidelberg: Springer-Verlag.

Sea Power Centre. (2005) "Gallipoli as a Joint Maritime Campaign", *Semaphore*, Issue 4, March 2005.

Travers, T. (2001) "Liman von Sanders, the Capture of Lieutenant palmer, and Ottoman Anticipation of the Allied landings at Gallipoli on 25 April 1915", *The Journal of Military History*, 65, October 2001, Society for Military History, pp. 965-979)

Velsley, C., D. (1997) "Gallipoli 1915", Master of Arts Thesis, California State university Dominguez Hills, USA. http://proquest.umi.com.ezproxy.lib.rmit.edu.au/pqdweb?index=5&sid=1&srchmode=2&vinst=PROD&fmt=6&startpage=-1&clientid=16532&vname=PQD&RQT=309&did=738217001&scaling=FULL&ts=1230726095&vtype=PQD&rqt=309&TS=1230726950&clientId=16532

Web page accessed December 31 2008.