Error in Spatial Ecology (PVM)

Brian Lees

Introduction

Uncertainty and error are unavoidable in spatial ecology mapping just as in other geographical applications. There are several common approaches, mapping polygons directly on air photographs, classification of digital remote sensed data or, often the most accurate, the spatial extension of good point data obtained by field survey (Predictive Vegetation Mapping or PVM). The focus of this paper is PVM. Huang and Lees (2004, 2005, 2007) looked at two main types of error in the sorts of models used to spatially extend good, point, vegetation data, inherent and operational. Inherent error is error resulting from inaccurate input data, while operational error is the error which results from the modelling itself. All spatial data, whether they are paper maps or digital layers, in vector or raster format, having categorical or numerical values, contain errors to some extent. This is due to not only instrument and human errors, but also the age of the data and the inherent complexity of the real world. Models tend to reduce the complexity of the real world using approximations and the cost of these approximations is increased error, uncertainty and less precision. Errors in input data can be transmitted through the modelling process and become manifest in the final products. Different models transmit error in different ways. The exact error propagation modes for most models are unclear and need careful analysis (Huang and Lees 2005). Van Niel et al. (2004) did this for error in elevation data, a common input to PVM, and Huang (Huang and Lees 2004) did this for PVM more generally. Huang and Lees (2007) and Hagen-Zanker et al. (2005) went further and looked at the effects of learning sample location and class error.

B. Lees (✉)
School of Physical Environmental and Mathematical Sciences,
University of New South Wales Canberra, ACT, Australia
e-mail: Brian.Lees@adfa.edu.au

© Springer Nature Singapore Pte Ltd. 2017

C. Zhou et al. (eds.), *Spatial Data Handling in Big Data Era*,
Advances in Geographic Information Science,
DOI 10.1007/978-981-10-4424-3_4

Once we properly identify the types and sources of error, track the error transmission through the modelling process, measure and quantify the magnitude of the error with suitable error indices, and visualise the error measurements for the users, sensible strategies need to be implemented to reduce or manage the error if possible. Common sources of error in predictive vegetation modelling are DTM inaccuracy, which flow through to slope, wetness estimates and aspect/insolation calculations, and soil/lithology data. Given the problems of circularity with soil mapping which is now often based on a model integrating vegetation, geology and slope data (McBratney et al. 2003), vegetation modellers often resort to geology maps rather than soil maps. However, geology maps are rarely produced at fine scale and often the classes depicted relate to the age of a unit rather than its varied lithology. These issues have been dealt with elsewhere.

This paper looks at one of the common sources of error in predictive vegetation mapping, the actual style of mapping, although the principles are more generally applicable to other forms of natural resource mapping. The style of mapping commonly used for vegetation is the area-class map (Mark and Csillag 1989; Bunge 1966). This is a form of thematic map often described as a choropleth map, but Mark and Csillag (1989) point out that this usage is incorrect and that the term 'choropleth map' should be restricted to situations where the bounding polygons are defined by other than the phenomena being mapped. For example, the administrative boundaries used by McHarg and Mumford (1969) for forest type polygons in his Potomac River Basin Study are correctly choropleth maps but give a quite misleading impression of the land cover around Washington because they are inappropriate for the phenomena being mapped and suffer badly, as a result, from the Modifiable Areal Unit Problem.

Background

In 1977 David Sinton wrote a seminal paper on 'the inherent structure of information as a constraint to analysis'. He used mapped thematic data of natural resources as a case study. Thematic mapping dates back to the early 17th Century, well before computer databases were available. However, thematic mapping of natural resource information of the type discussed by Sinton is properly area-class mapping although he didn't use that term. In these types of map the structure is of discrete polygons with the boundaries between categories or classes of the theme. These occur over contiguous regions of geographic space. The polygon structure can be seen as enclosing an Entity but Goodchild et al. (2009) suggest it can also be seen as is a Nominal Field. The structure both reflects the way we think and, to some extent, how we go about collecting data.

Sinton made the point that certain types of information are collected in such a way that there are steps which include generalisation inherent in the collection process. This generalisation of the information collected limits the types of analysis which can be carried out later.

This sounds like an obscure and abstract point, but it is not. For example, during the 1980's the Australian Government decided, as a result of considerable conflict between those supporting the logging of native forests and those opposing this on environmental grounds, to have a National Forest Inventory. This was intended to produce a master data set to avoid the conflict which had arisen between datasets produced by the competing interests. The first pass at joining up forest maps across state borders revealed that there were serious inconsistencies in the ways each state, and in some cases region, defined their forests. This should not have been a surprise as similar problems had been observed when attempts have been made to mosaic soil maps across administrative boundaries. Solving the problem by resorting to the original observations to recode the data proved impossible as they had not been retained.

Indeed, the practice in many field sciences was to observe, classify then record. In many cases the detail of the original observation only existed in field notebooks, which were not archived. In one legacy vegetation map close to home, of the Australian Capital Territory, it was found that the map was a mosaic based on three surveys, each by a different group, each at a different time, and each at a different nominal scale. The apparent homogeneity of one part and the apparent heterogeneity of another part were artifacts of this mosaicking. The map was, as a result, extremely misleading.

The situation described above is common. It is extremely costly to collect new data and the temptation is to make do with what is to hand. So, it is important to properly understand the process by which that data has been generated and stored.

Descriptions, Depictions, or Diagrams

The area-class map is a type of diagram. Maps and images, more generally, fall into this class of media. Often field scientists use a mixture of descriptive writing, symbolic equations and diagrams to communicate their ideas. These media appear to be processed in our brains through different pathways. The distinction between the way we believe these are processed are between model-based theories (including graphical ones), and sententially based theories. These appear to be dealt with by different parts of our brain. The issue has been presented as a choice between reasoning being semantic or syntactic respectively (Engelhardt 2002). In terms of syntax, written text and symbolic equations are essentially linear, while diagrams need not be. This suggests that the common practice of supporting a piece of difficult text with a graph or diagram works because we are delivering the message to our brain through two different pathways.

We can discriminate between the way we understand descriptions (written text and symbolic equations) and diagrams (Lemon and Pratt 1999). Most descriptions are Fregean (Peregrin 2000) where the symbol bears no resemblance to the thing it represents in terms of its properties. But, with analogical representations where relationships are explicit, there is such a resemblance. This is the first key difference between written text and symbolic equations (Fregean), and diagrams (analogical). The second key difference is that there is a sense in which text (and equations) are a sequential set of symbols. Text (and equations) rely for much of their meaning on this sequence. Usually what is written later in a sentence assumes that the reader knows what has been written beforehand. Because written text is processed by people in a sequential fashion it is hard for humans to store all the things that we have just read in our working memory, especially when the number of items appearing in the text is large. Diagrams can avoid this problem. However, sentential descriptions may allow a more accurate and precise description of a phenomenon than diagrams.

Only those diagrams in the bottom half of Table 1 can deal with spatial data spatially. This distinguishes tables and charts from diagrams of the sort we are concerned with. Why is a sentential representation not as effective as a diagram in helping us to solve a problem? (Larkin and Simon 1987);

- Diagrams can group together all the information that is used together. This effective use of display space is an important attribute as it thus avoids lengthy searches for the elements needed to make a problem-solving inference.
- Diagrams avoiding the need to match symbolic labels by using location to group information about a single element.
- Diagrams automatically support a large number of perceptual differences, which are extremely easy for humans to understand. In addition to the relationships in space, the meanings of relative size and colour are all easily understood by humans.

Diagrams appear to operate best when cognitive reasoning, which must extract the structural information from the sentential data by laborious comparisons and computations, is supported by a visual representation from which the user can

Table 1 Types of diagram (modified after Lemon and Engelhardt in TwD 1997)

Type of diagram	Characteristics	Treatment of spatial data is
Columns/list	Partitioning in one dimension	Aspatial
Table/matrix	Partitioning in two dimensions	Aspatial
Time line	Metric in one dimension	Aspatial
Two-axis chart	Metric in the Cartesian plane	Aspatial
Bar chart	Partitioning in one dimension combined with Metric in one dimension	Aspatial
Picture or image	Realistic; planar metric combined with planar topology	Spatial
Map	Symbolic; planar metric combined with planar topology	Spatial
Network diagram	Planar topology	Spatial
3D/CAD graphics	Metric and topology in Cartesian 3-space	Spatial

easily perceive the structure of the data. The use of graphs, tables and bar charts in the scientific literature is the best example of this phenomena. Maps, on the other hand, are quite often 'stand-alone' with the only text accompanying them being peripheral to the image. Some types of map rely on only the users' familiarity with the style of representation to carry their message. The area-class map has been successful because the message that it carries is easy to comprehend.

Collecting Data

Sinton (1977) breaks down the process of creating useful data for area-class mapping from initial observation. He argues that one must record the Theme, the Location and the Time of the observation. At least three types of generalisation take place;

- Aggregation—the transformation of the structure from point to polygon usually involves a loss of detail in Spatial Location of the original observation.
- Classification—observations are categorized with other like observations. In practice 'like' becomes 'similar' and the breadth of what 'similar' means is an important source of error.
- Induction—where a series of samples are generalised to include locations assumed to have the same characteristics. Thus aggregating places with similar (but not the same) characteristics as the places sampled into Classes.

Sinton (1977) defines this more precisely (in the case of vegetation mapping).

- One of the Attributes of the data (Time) is held constant.
- A second Attribute (Theme) is permitted to vary in a controlled way.
- The third Attribute (Location) is measured for its variation within the second (controlled) Attribute.

In thematic mapping, time is rarely not held constant. For some themes, such as elevation, soils or geology, this is rarely a problem. But for other more dynamic themes, such as vegetation, land use, land cover or population, the freezing of the theme at the time of data collection does present a major source of error. Allowing the theme to vary 'in a controlled way' runs the risk of falling foul of the Modifiable Area Unit Problem (Openshaw 1983). In this, poorly represented classes tend to be incorporated into larger classes, and so disappear. Also in many cases a continuum of change is represented as a series of classes, which are essentially overlapping gaussians. This, of course, leads to the creation of errors of omission and commission in the theme. Allowing the location to vary from a point to a polygon leads to similar errors. Generalisation can be seen as inducing error with regard to the original observations.

So, the form of an area-class map certainly aids perception and processing of the information presented. But this is at the cost of considerable generalization.

Changing the Data Model

In a previous paper (Lees 1996b) we investigated a method of representing vegetation data which avoided many of the errors which inherent in the use of the area class map. In this earlier paper we investigated how the processing of data to suit this format perpetuates the use of an inappropriate data model and places an upper limit on the accuracy of spatial extension of point data by most predictive modelling techniques.

A number of projects based on the Kioloa data set to develop and test new predictive modelling tools were set up to use standard (forest) industry data as input, and forest types as output (Moore et al. 1991; Lees and Ritman 1991; Fitzgerald and Lees 1993, 1994, 1996). Using the same modelling methods, learning sample and independent variables, but not classifying the learning samples into communities or forest types beforehand, it was shown to be possible to achieve a significant improvement in predictive accuracy over area-class mapping. If one deals with the point observations of vegetation without classifying them beforehand, then another form of visualisation can be used. Because we are dealing with digital information, we are no longer constrained to produce a single 'map' as the data storage medium and visualisation medium. We can now store and retrieve a considerable amount of information about any point, either in geographic, environmental or spectral space. This is a database-oriented approach. If we retain geographic space as the most convenient operational data space for users of predictive modelling, we are also selecting the domain with least database complexity where each point exists in only one location in each of the other domains (Lees 1994; 1996a, b; Aspinall and Lees 1994). We can then model the spatial extension of each relevant attribute of each entity observed in the ground truth plots.

The Database Oriented Technique

In Lees (1996b) the original observations were recoded as fuzzy membership of the canopy, species by species. A simple Back Propagation neural net was set up for each species. In order to provide an ongoing comparison of methods, the input layer was the same as that described in Fitzgerald and Lees (1993, 1994), and used the same datasets as Lees and Ritman (1991). A 9/10/10 structure was used with one hidden layer of ten nodes and an output layer of ten nodes. No spatial or temporal context was used. Each output node represented a range of fuzzy memberships (0–0.1; 0.1–0.2 and so on). The highest number in the output range was taken to indicate the membership of that cell and the whole output range for each cell was treated as a distribution and the probability of the membership was calculated. So, for each species, it was possible to estimate its fuzzy membership of the canopy and the probability of that estimate. For the species chosen as an example, Eucalyptus maculata, and using only the fuzzy memberships, the RMS error was calculated to be 0.2324. Calculating a distribution based on the ten output nodes is, of course, not optimal, but it seemed more sensible than setting the network up with, say, a 9/10/30 structure.

The constraints we placed on our earlier work to make it match existing user expectations resulted in an upper limit to predictive accuracy, for all the models tested, of about 65%. Those themes which were more appropriately represented by this data structure (land/sea discrimination) could be predicted with up to 99% accuracy. The later work (Lees 1996b) which expressed species distribution as a fuzzy membership of the tallest stratum (in a forest), midstratum or understorey, and lower stratum or ground layer, allows the prediction of a series of data layers which represent surfaces of spatially varying fuzzy memberships appeared to offer a significant gain in accuracy. Further, the use of a simple neural net configuration enables both fuzzy membership and the probability of this membership to be estimated. This made it possible to track error through subsequent uses of the modelled estimates. Methods of comparing the two methods have subsequently been investigated and this paper presents some of these findings.

Just How Bad Is an Area-Class Map of Forest Types?

The method used in Lees (1996b) required the information to be stored in some form of a database and no integration of species distributions was suggested. For use in the field this meant that a small computer system would have to be utilised. A fieldworker wishing information about a point would have to scroll through, in the case of the Kioloa data set, the distributions of forty-one tree species, ninety-four shrubs and one hundred and eight understorey species. Together with an error map for each, a database of 486 coverages would be needed for the Kioloa area. It would certainly be possible to query the database to return the information relating to a point but the overview of the immediate local area would still be elusive.

The increase in accuracy gained is seriously offset by the indigestible nature of the output. For real, practical, use some form of synthetic product is needed. This inevitably leads back to a classification of the data into communities, forest types, and so on. Depiction of this sort of synthesis naturally leads to an area-class map. It was argued in Lees (1996b) that it was easy to show that this is a major source of error but there is no insight into what might be the most appropriate replacement. In this exercise we have examined whether this claim was as accurate as thought. Just how bad is an area-class map of forest types for the normal user of these products?

To evaluate this question Allison (1998) decided to move away from the simple measure of ‘overall accuracy’ and investigate other measures. The calculation of ‘overall accuracy’, or ‘proportion of samples correctly classified’, used in many of our earlier papers, ignores any difference between the accuracy of individual classes. She used the Kioloa data set and one of our early models based on it (Lees and Ritman 1991) for the evaluation. This model was one of the earliest we developed, and one of the least accurate. The data set used covers an area of coastal plain and beach which runs back into a sandstone ridge with elevations ranging from sea level to 285 m. Land cover varies from wet sclerophyll rainforest, highly disturbed dry and wet sclerophyll forests with a complex fire history parts of which have been logged in the past, through heath to cleared grassland. There are small pockets of rainforest with palms. PVM models based on this dataset using the ground data pre-classed into forest types typically give predictions of ‘Sea’ at better than 95%, ‘Paddock’ at better than 80% and the other forest types at accuracies of less than 65%.

The ‘Overall Accuracy’ calculation of the Lees and Ritman (1991) model gave only a moderate result of 47.64%, based on the error matrix (Table 2), ‘User’s Accuracy’ and ‘Producer’s Accuracy’ calculations (Table 3) and errors of omission and commission (Table 4) were calculated for each class. These quite effectively demonstrate the problem of treating the continuum of change within this type of forest as a set of overlapping gaussians, or classes. Classes which were clearly separable, such as ‘Sea’, were predicted with low values of errors of omission or commission and high values of both Producer’s and User’s Accuracy. The difficult ecotonal ‘classes’ such E.botryoides forest and Lower Slope Wet forest, which appear to be separable in geographic space, are not separable in environmental data space and have very large errors associated with them (Lees 1996b). Whilst it is tempting to interpret this as meaning that better results are impossible using classed data, it is worth examining the degree of error involved in these measures.

Forest maps are specifically targeted at a particular type of user, the forest manager. Traditionally, the major concern of such a user was to accurately predict the timber volume in a forest coupe. Multi-use forests have extended the range of concerns foresters must address, but it is possible to look at each use and categorise the level of unsuitability of the map for a particular purpose. In uses where species composition is important, one could generate a table such as this (Table 5).

Table 2 The error matrix for Lees and Ritman (1991)

	Reference data									Total
	1	2	3	4	5	6	7	8	9	
Predicted class										
1	167	15	15	35	35	4	2	7	0	280
2	13	3	1	7	8	2	2	8	0	44
3	13	1	1	4	1	1	1	1	0	23
4	32	20	14	126	43	73	60	45	8	421
5	52	18	8	46	70	6	6	17	2	225
6	6	0	2	17	5	9	3	0	0	42
7	22	8	10	16	19	2	12	18	3	110
8	1	1	1	0	0	0	1	48	1	53
9	0	0	0	0	0	0	0	2	259	261
Total	306	66	52	251	181	97	87	146	273	1495

Table 3 User's and producer's accuracy (Allison 1998)

Vegetation class	User's accuracy (%)	Producer's accuracy (%)
1 Dry Sclerophyll forest	59.64	54.58
2 E.botryoides forest	6.82	4.55
3 Lower slope wet forest	4.35	1.92
4 Wet (E.maculata) forest	29.93	50.20
5 Dry (E.maculata) forest	31.11	38.67
6 Rainforest ecotone	21.43	9.28
7 Rainforest	10.91	13.79
8 Paddocks and cleared	90.57	32.88
9 Sea	99.23	94.87

Table 4 Errors of omission and commission (Allison 1998)

Vegetation class	Error of omission (%)	Error of commission (%)
1 Dry Sclerophyll forest	45.42	40.36
2 E.botryoides forest	95.45	93.18
3 Lower slope wet forest	98.08	95.65
4 Wet (E.maculata) forest	49.80	70.07
5 Dry (E.maculata) forest	61.33	68.89
6 Rainforest ecotone	90.72	78.57
7 Rainforest	86.21	89.09
8 Paddocks and cleared	67.12	9.43
9 Sea	5.13	0.77

Table 5 Levels of disagreement (Allison 1998)

Level 0	No disagreement	Predicted vegetation type matches actual vegetation type exactly
Level 1	Low level disagreement	Predicted vegetation type incorrectly assigned, but both predicted and actual vegetation types contain at least one identical main indicator species
Level 2	Moderate disagreement	Predicted vegetation type main species occurs as associated species or vice versa
Level 3	High level disagreement	Only predicted associated species occurs (as associated species)
Level 4	Total disagreement	No predicted main or associated species found

It can be seen that disagreement at both levels 1 and 2 is not likely to be very significant to the sort of user described above, the prediction is 'fairly close' to the actuality. Level 4 disagreement, however, is simply wrong. Here the prediction is completely at odds with the actuality. Disagreement at level 3 is harder to categorise, it may be as serious as level 4 for some users. But a ranking of the levels of disagreement such as this allows us to categorise the misclassifications in Lees and Ritman (1991) as shown in Table 6.

Table 6 Number of samples at each level of disagreement (Allison 1998)

Level of disagreement	Number of samples	Percentage of total
0	695	47.64
1	190	13.02
2	312	21.39
3	122	8.36
4	140	9.56

We can then go on to quantify the errors by looking at our test sample (Table 6) or even plot these as a map of levels of disagreement. From Table 6 one could suggest that, for a vegetation scientist checking the model, the accuracy would certainly be only 47.64%, but for a forest manager the accuracy would, in practical terms, be between 60.66 and 82.05% as the predicted species does, in fact, occur at level 2 disagreement. This suggests that it can still be a usable product at these levels. Importantly, the latter level of accuracy is slightly higher than that achieved by Lees (1996b).

Conclusion

All of this suggests that some our efforts to identify error and to reduce it have, if not been misguided, ignored the practicalities of field use. Even with all the miniaturized electronic tools available for field use, the database model is too cumbersome a vehicle for easy comprehension of all the necessary information. The traditional form of delivery, and use, of vegetation and soils information, the area-class map, may indeed perpetuate the use of an inappropriate data model and place an upper limit on the accuracy of spatial extension of point data by most predictive models, but it remains a practically useful form of delivering spatial information for most users. Clearly a combination of both methods, that described by Lees (1996b) and the synthesized area-class map, would allow access to the advantages of both. The former is clearly best suited to the office system, and the latter to the field.

References

Allison, B. (1998). *The assessment of the accuracy of a vegetation map produced using a GIS*. Unpublished BSc(REM) honours thesis, Department of Geography, Australian National University, Canberra, p. 133.

Aspinall, R. J., & Lees, B. G. (1994). *Sampling and analysis of spatial environmental data. Advances in GIS research* (pp. 1086–1098). Southampton: Taylor and Francis.

Bunge, W. (1966). *Theoretical geography*. Royal University of Lund, Department of Geography; Gleerup.

Engelhardt, J. V. (2002). The language of graphics: a framework for the analysis of syntax and meaning. Universiteit van Amsterdam. Institute for Logic. Language and Computation, pp. 193.

Fitzgerald, R. W., & Lees, B. G. (1993). Assessing the classification accuracy of multisource remote sensing data. *Remote Sensing of the Environment, 47*(3), 1–25.

Fitzgerald, R. W., & Lees, B. G. (1994). Spatial context and scale relationships in raster data for thematic mapping in natural systems. *Advances in GIS research, 1994*(1), 462–475.

Fitzgerald, R. W., & Lees, B. G. (1996). Temporal context in floristic classification. *Computers & Geosciences, 22*(9), 981–994.

Goodchild, M., Zhang, J., & Kyriakidis, P. (2009). Discriminant models of uncertainty in nominal fields. *Transactions in GIS, 13*(1), 7–23.

Hagen-Zanker, A., Straatman, B., & Uljee, I. (2005). Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science, 19*(7), 769–785.

Huang, Z., & Lees, B. (2005). Representing and reducing error in natural-resource classification using model combination. *International Journal of Geographical Information Science, 19*(5), 603–621.

Huang, Z., & Lees, B. G. (2004). Combining non-parametric models for multisource predictive forest mapping. *Photogrammetric Engineering and Remote Sensing, 70*(4), 415–425.

Huang, Z., & Lees, B. G. (2007). Assessing a single classification accuracy measure to deal with the imprecision of location and class: Fuzzy weighted Kappa versus Kappa. *Journal of Spatial Science, 52*(1), 1–13.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science, 11*(1), 65–100.

Lemon, O., & Pratt, I. (1999). Putting channels on the map: Verisimilitude and spatial constraints in a semantics of geographic information systems. *Logic, Language, and Computation, 2,* 143–164.

Lees, B. G. (1994). Decision trees, artificial neural networks and genetic algorithms for classification of remotely sensed and ancillary data. In *7th Australasian remote sensing conference, remote sensing and photogrammetry association Australia Ltd. Floreat* (pp. 51–60).

Lees, B. G. (1996a). Sampling strategies for machine learning using GIS. In M. F. Goodchild, L. Steyart, B. Parks, M. Crane, C. Johnston, D. Maidment & S. Glendinning (Eds.), *GIS and environmental modelling: Progress and research issues* (pp. 39–42). Fort Collins, Co: GIS World Inc.

Lees, B. G. (1996b). Improving the spatial extension of point data by changing the data model. In M. Goodchild (Ed.), *Proceedings of the Third international Conference on Integrating GIS and Environmental Modeling Santa Fe, New Mexico*, National Centre for Geographic Information and Analysis, WWW; CD.

Lees, B. G., & Ritman, K. (1991). Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management, 15*(6), 823–831.

Mark, D. M., & Csillag, F. (1989). The nature of boundaries on 'area-class' maps. *Cartographica. Cartographica: The International Journal for Geographic Information and Geovisualization, 26*(1), 65–78.

McBratney, A. B., Santos, M. L. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1), 3–52.

McHarg, I. L., & Mumford, L. (1969). *Design with nature* (p. 208), New York: Wiley.

Moore, D. M., Lees, B. G., & Davey, S. (1991). A new method for predicting vegetation distributions using decision tree analyses in a geographic information system. *Environmental Management, 15*(1), 59–71.

Openshaw, S. (1983). *The modifiable area unit problem*. Norwich, England: GeoBooks.

Peregrin, J. (2000). Fregean logic and Russellian logic. *Australasian Journal of Philosophy, 78*(4), 557–574.

Sinton, D. F. (1977). The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. International advanced study symposium on topological data structures for geographic information systems, Dedham, Mass. In G. Dutton (Ed.), *Harvard papers on geographic information systems, Harvard University* (p. 1978). Mass: Camb.

TWD, On-line discussion. (1997). *Thinking with diagrams*. http://www.mrc-cbu.cam.ac.uk/projects/twd/online-discussion. Accessed August 5, 2005.

van Niel, K., Laffan, S. W., & Lees, B. G. (2004). Error and uncertainty in environmental variables for ecological modelling: Sensitivity of variables to error in source elevation data. *Journal of Vegetation Science, 15,* 747–756.