The Effect of LiDAR Data Density on DEM Accuracy

Liu, X. [1,2], Z. Zhang [2], J. Peterson [2] and S. Chandra [2]

[1]Australian Centre for Sustainable Catchments and
Faculty of Engineering and Surveying
University of Southern Queensland
Toowoomba, Qld 4350, Australia
[2]Centre for GIS, School of Geography and Environmental Science
Monash University, Clayton, Vic 3800, Melbourne, Australia
Email: Liux@usq.edu.au

Keywords: DEM, LiDAR, data reduction, accuracy

EXTENDED ABSTRACT

Digital Elevation Models (DEMs) play an important role in terrain related applications, and their accuracy is crucial for DEM applications. There are many factors that affect the accuracy of DEMs, with the main factors including the accuracy, density and distribution of the source data, the interpolation algorithm, and the DEM resolution. Generally speaking, the more accurate and the denser the sampled terrain data are, the more accurate the produced DEM will be. Traditional methods such as field surveying and photogrammetry can yield high accuracy terrain data, but are very time consuming and labour intensive. Moreover, in some situations such as in densely forested areas, it is impossible to use these methods for collecting elevation data. Light Detection and Ranging (LiDAR) offers high density data capture. The high accuracy three dimensional terrain points prerequisite to very detailed high resolution DEMs generation offers exciting prospects to DEM builders. However, because there is no sampling density selection for different area during a LiDAR data collection mission, some terrains may be oversampled thereby imposing increases in data storage requirements and processing time. Improved efficiency in these terms can accrue if redundant data can be identified and eliminated from the input data set. With a reduction in data, a more manageable and operationally sized terrain dataset for DEM generation is possible (Anderson *et al.*, 2005a).

The primary objective of data reduction is to achieve an optimum balance between density of sampling and volume of data, hence optimizing cost of data collection (Robinson, 1994). Some studies on terrain data reduction have been conducted based on the analysis of the effects of data reduction on the accuracy of DEMs and derived terrain attributes. For example, Anderson *et al.* (2005b) evaluated the effects of LiDAR data density on the production of DEM at different resolution. They produced a series of DEMs at different horizontal resolutions along a LiDAR point-density gradient, and then compared each of these DEMs to a reference DEM produced from the original LiDAR data, this having been acquired at the highest available density. Their results showed that higher resolution DEM generation is more sensitive to data density than is lower resolution DEM generation. It was also demonstrated that LiDAR datasets could withstand substantial data reductions yet still maintain adequate accuracy for elevation predictions (Anderson *et al.*, 2005a)

This study explored the effects of LiDAR point density on DEM accuracy and examined to scope for data volume reduction compatible with maintaining efficiency in data storage and processing. Something of the relationship between data density, data file size, and processing time also emerges from this study.

The study area (113 km²) falls within the Corangamite Catchment Management Authority (CCMA) region, (south western Victoria, Australia). LiDAR data points were first randomly selected and separated to two datasets: 90% for training data and 10% for check points. Training datasets were used for subsequent reduction to produce a series of datasets with different data density, representing the 100%, 75%, 50%, and 25%, 10%, 5%, 1% of the original training dataset. Reduced datasets were used to produce correspondent DEMs with 5 m resolution. Results show that there is no significant difference in DEM accuracy if data points are reduced to 50% of the original point density. Processing time for DEM generation can thus be reduced to half of the time needed when using the original dataset.

1. INTRODUCTION

Digital Elevation Models (DEMs) play an important role in terrain-related applications, the success of which refers, among other things to accuracy. Of the many factors that affect the accuracy of DEMs, the accuracy, density and distribution of the source data, the interpolation algorithm, and the DEM resolution or grid size are the main factors (Gong *et al.*, 2000; Kienzle, 2004; Li *et al.*, 2005; Fisher and Tate, 2006). Generally speaking, the more accurate and the denser the sampled terrain data are, the more accurate the derived DEM will be. Competent application of traditional methods such as field surveying and photogrammetry yields high accuracy terrain data, but they are time consuming and labour intensive. In some terrains, for example, in densely forested areas, it is even impossible to use these methods for collecting elevation data. Light Detection and Ranging (LiDAR) provides an alternative high density and high accuracy three dimensional terrain point data acquisition. The principles of LiDAR and using LiDAR data to produce high quality DEMs have been well documented (Lohr, 1998; Wehr and Lohr, 1999; Lloyd and Atkinson, 2006; Liu *et al.*, 2007). LiDAR data accuracy and density are such that reliable and high accuracy, high resolution DEM generation can be confidently contemplated.

The primary objective of data reduction is to achieve an optimum balance between density of sampling and volume of data, hence optimizing the cost of data collection (Robinson, 1994). If the input data are not strictly regular in distribution, much depends on choice of and access to a suitable interpolation method. Tests of alternative approaches (Zimmerman *et al.*, 1999) show that none of them is of universal applicability. If sampling data density is high, the IDW (inverse distance weighted) method performs well (Ali, 2004; Blaschke *et al.*, 2004; Chaplot *et al.*, 2006). Because LiDAR data have high sampling density, the IDW method is sufficient for interpolating LiDAR data (Anderson *et al.*, 2005a; Liu *et al.*, 2007).

Using appropriate interpolation, very detailed high resolution DEMs with high accuracy can be generated from high density LiDAR data. However, because there is no scope to match data acquisition density by terrain type during a LiDAR data collection mission, some oversampling is usually inevitable. As a result, the data storage requirement and processing times will be higher than necessary. Strategies for handling the large volumes of terrain data without sacrificing accuracy are required (Kidner and Smith, 2003) if

efficiency is to be considered (Bjørke and Nilsen, 2002; Pradhan *et al.*, 2005). Through informed reduction in data (i.e. ratio of the information content to the volume of the dataset) (Chou *et al.*, 1999), a more manageable and operationally sized terrain dataset for DEM generation is possible (Anderson *et al.*, 2005a).

Some studies on terrain data reduction have been conducted based on the analysis of the effects of data reduction on the accuracy of DEMs and derived terrain attributes. For example, Anderson *et al.* (2005b) evaluated the effects of LiDAR data density on DEM production at a range of resolutions. They produced a series of DEMs at different horizontal resolutions along a LiDAR point density gradient, and then compared each DEM produced with different LiDAR data density at a given horizontal resolution, to a reference DEM produced from the original LiDAR data (the highest available density). Their results show that higher resolution DEMs are more sensitive to data density than lower resolution DEMs. It was demonstrated that LiDAR datasets could withstand substantial data reductions yet maintain adequate accuracy for elevation predictions (Anderson *et al.*, 2005a).

This study explored the effects of LiDAR data density on the accuracy of DEMs and examined to what extent a set of LiDAR data can be reduced for improving storage and processing efficiency in a moderate complex terrain area. It also attempted to examine the relationship between data density, data file size, and processing time. The choice of suitable DEM resolution based on terrain data density for the generation of an efficient DEM was discussed in detail.

2. MATERIAL AND METHODS

2.1. Study Area

The study area is in the region of Corangamite Catchment Management Authority (CCMA) in south western Victoria, Australia. Terrain types vary between the comparatively treeless basins of internal drainage on Victoria Volcanic Plains (VVP) to dissected terrains north and south. The plains have high priority for a range of research projects pertaining to environment management issues addressed in the catchment management strategy plan. In this study, a 113 km² sub-catchment area of moderate terrain complexity and covered by LiDAR data was selected as the test site, shown in Figure 1.
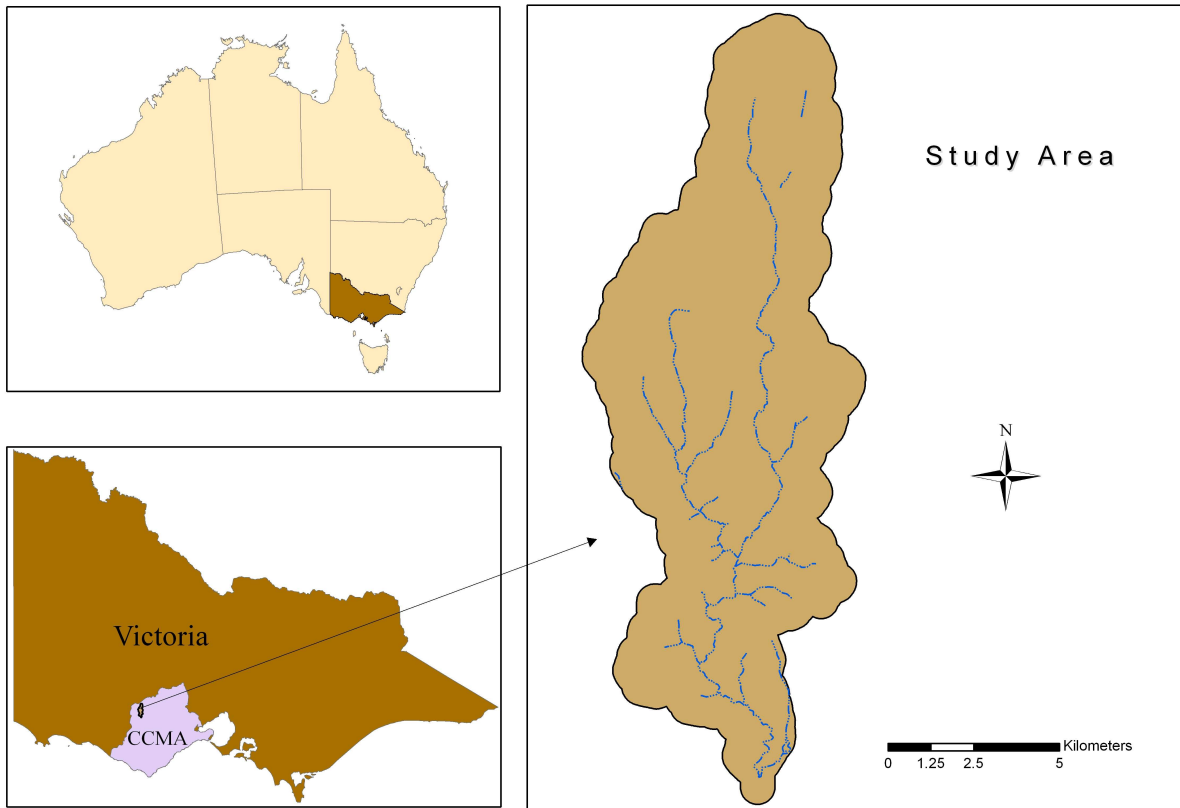
Figure 1. Study area: terrain height varies between 303m and 139m

2.2. Data

LiDAR data for an area of 6900 km², covering most part of VVP in the CCMA region, were collected over the period of 19 July 2003 to 10 August 2003. The primary purpose of this LiDAR data collection was to facilitate more accurate terrain pattern representation for the implementation of a serious of environment related projects. The LiDAR data were delivered by AAMHatch Pty Ltd as tiles in ASCII files containing x, y, z coordinates and intensity values. The data have been classified into ground and non-ground points by using data filter algorithms across the project area. Manual checking and editing of the data led to further improvement in the quality of the classification. The resulting data products used for DEM generation are irregularly distributed ground 3D points, with an average spacing of 2.2 m. The accuracy of LiDAR data was estimated as 0.5 m vertically and 1.5 m horizontally (AAMHatch, 2003).

2.3. Methods

Using the Geostatistical Analyst extension of ArcGIS 9.1, LiDAR data points were first randomly selected and separated to two datasets: 90% for training data and 10% for check points. Training datasets were used for subsequent reduction to produce a series of datasets with different data density, representing the 100%, 75%, 50%, 25%, 10%, 5%, 1% of the original training dataset. Reduced datasets were used to produce a series of DEMs. In this study, all the DEMs were generated with 5 m resolution, thus isolating the effects of DEM resolution. The reason for separating training data as 90% and test data as 10% of the original dataset is to ensure the high density of the training dataset and provision of enough test dataset check points. In this case, the average density of training data is about 2.4 m (space interval), nearly same as the original dataset. In the test dataset, a total of 465,136 points can be used as check points to assess

the accuracy of each of the range of DEMs produced.

With the IDW interpolator, a DEM was produced for each of the seven datasets. Data density, file size and processing time for generating each of these DEMs were listed in Table 1.

To assess the accuracy of DEMs generated from reduced LiDAR datasets with different data density, independent elevation checking is conducted by comparing elevation values of test data with correspondent elevation values interpolated from the DEM were calculated for each generated DEM. Root mean square errors (RMSEs) and standard deviation σ for each DEM were calculated to evaluate the overall accuracy of the DEM. The RMSE and σ were calculated with:

$$RMSE = \sqrt{\frac{\sum (E_{DEM} - E_{\mathrm{Re}f})^2}{n}} \qquad (1)$$

$$\sigma = \sqrt{\frac{\sum (E_{DEM} - \overline{E})^2}{n-1}} \qquad (2)$$

where, E_{DEM} is the elevation value from the DEM, and $E_{\mathrm{Re}f}$ is the correspondent reference elevation value from check points. n is the number of check points. $\overline{E}$ is a calculated mean error with: $\overline{E} = \sum (E_{DEM} - E_{\mathrm{Re}f})/n$.

Details of results are listed in Table 2.

Table1. Data density, file size and processing time for DEM generation

Reduced datasets	Data density (points/per m²)	Point data file size (MB)	Processing time for DEM generation
100%	0.037	444.0	14.31 h
75%	0.028	360.0	12.31 h
50%	0.018	240.0	7.00 h
25%	0.009	120.0	3.12 h
10%	0.004	48.6	57 m
5%	0.002	24.4	27 m
1%	less than 0.001	4.92	8 m

Table 2. Accuracy assessment of DEMs with different data density.

Reduced datasets	Maximum elevation difference (m)	Root mean square error (m)	Standard deviation (m)
100%	2.856	0.184	0.194
75%	2.858	0.188	0.196
50%	2.933	0.194	0.202
25%	4.489	0.212	0.220
10%	5.742	0.262	0.268
5%	7.923	0.326	0.331
1%	11.839	0.641	0.643

3. RESULTS AND DISCUSSION

As expected, data density reduction (i.e. increased space interval between data points) influences DEM accuracy: errors increased as data density decreased (Table 2). The larger distance between sampling points adversely affects the accuracy of generated

DEMs (Anderson *et al.*, 2005a). Compared with the DEM produced from the total LiDAR training dataset, however, there is no significant decrease in accuracy for the DEM generated from the 50% training dataset. This can be seen from Figure 2 in terms of both RMSE and standard deviation. Processing time for generating the DEM is only the

half of the time needed for generating the DEM using the total LiDAR training dataset. It should be note that the processing time listed in Table 2 may vary with the types of computer and software to be used. Point file size is also nearly the half of the total point file size. Therefore, for this study area and LiDAR dataset, at the given resolution of 5 m, the "efficient dataset" is the one with 50% of the original data density.
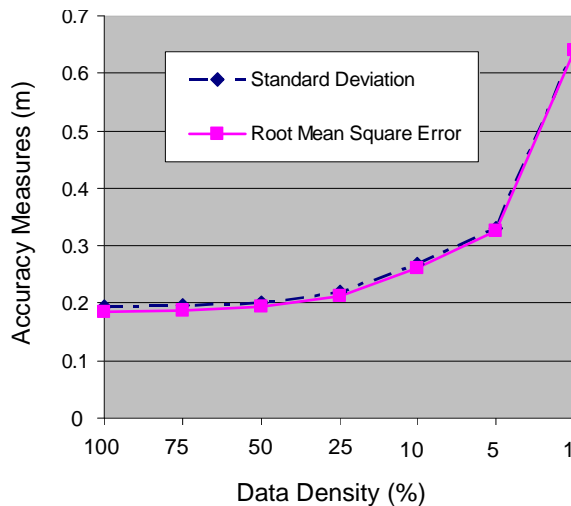

Figure 2. Data reduction and DEM accuracy

This study shows that LiDAR dataset (density) reduction can increase the efficiency of DEM generation in terms of file size and processing time. However, to what extent a dataset can be reduced depends on the original data density, terrain characteristics, interpolation method for DEM generation, and target DEM resolution (grid size). In this study, the effects of LiDAR data reduction on the accuracy of DEM were evaluated for a terrain with a range of relative relief attributes. The IDW interpolation method and 5 m resolution were selected for all DEM generation. Further comparison with different interpolation methods and DEM resolution needs to be implemented if comprehensive guidelines are to be assembled.

It is inappropriate to generate a high resolution DEM with very sparse terrain data: any surface so generated is more likely to represent the shape of the specific interpolator used than that of the target terrain because interpolation artefacts will abound (Florinsky, 2002; Albani *et al.*, 2004). The source data density constrains the resolution of DEM (Florinsky, 1998). On the other hand, generating a

low resolution DEM from high density terrain data will devalue the accuracy of the original data.

Clearly, the choice of the adequate resolution of a DEM is constrained by terrain input data density. McCullagh (1988) suggested that the number of grid cells should be roughly equivalent to the number of terrain data points in covered area. The grid size of a DEM can be estimated by:

$$S = \sqrt{\frac{A}{n}} \qquad (3)$$

where n is the number of terrain points and A is the covered area (Hu, 2003). This means that the DEM resolution should match the sampling density of the original terrain points.

In this study, for the DEM with 5 m resolution, each grid has 0.75 points in average for the DEM from the 50% dataset. And so the DEM resolution roughly matches the source data density. This is one reason why the accuracy of the DEM still can be maintained while LiDAR data density could be reduced to 50% of its original data density.

The ideal method for the assessment of the accuracy of a DEM is to compare the produced DEM with a "true" terrain surface. These kinds of "true" terrain surfaces are not available in practice. Although artificial terrain surfaces have been used to evaluate the effects of terrain complexity and interpolation methods on the accuracy of DEMs (Zhou *et al.*, 2006; Yilmaz, 2007), they are obviously not applicable for assessing the effects of sampled terrain data density. Using a DEM of relatively higher accuracy as a reference is an option, but access to such a DEM cannot be assumed when a new DEM-generation project is being implemented. The most commonly-used method is to compare interpolated elevation values from the DEM with a group of check points or with a subset of original points withheld from the generation of the DEM (Desmet, 1997; Yang and Hodler, 2000). Using RMSE or other accuracy measures the overall accuracy of the DEM can be evaluated. From a statistical point of view, the greater the number of check points, the more reliable the results. However, costs rise with number of check-points used. In this study, 10% of the original LiDAR data points (randomly selected) provided a sufficient number of check points (over 460,000 points) for evaluating the accuracy of DEMs produced at the different density levels.

The results of this study demonstrate possibilities for LiDAR data-set reduction before input to DEM generation routines. However, as mentioned, effective data reduction is affected by many factors and so further research is needed before a comprehensive set of guidelines can be assembled. It should also be noted that because not all data elements contribute optimally to the accuracy of produced DEM, the identification of feature-specific points (representing terrain features with more significant information content than other points) should be considered as an element in data volume reduction. Data reduction should be conducted in such a way that critical elements are kept while less important elements are removed (Chou *et al.*, 1999).

4. CONCLUSION

LiDAR technology offers high accuracy and high density 3D terrain data capture for detailed representation of terrain surfaces. However, without sampling selection of high density data during input data preparation for DEM generation, the storage requirements and processing times can be inflated due to data redundancy. Terrain data-point reduction mitigates the data redundancy and improves data-processing efficiency in terms of both storage and processing time. With guided data reduction, an efficient dataset can be identified for DEM generation.

This study showed that LiDAR data can be reduced to a certain level without significantly decreasing the accuracy of the output DEM. For a moderate complex terrain, LiDAR dataset with an average spacing of 2.4 m can be reduced to 50% of its original data density without degradation of the quality of the DEM. The accuracy of DEM produced using the 50% data reduction input data has no significant difference compared with the DEM produced from total original dataset. Such data reduction can lead to significant decrease of both data file size and processing time for DEM generation without compromising the DEM quality.

5. ACKNOWLEDGEMENT

6. REFERENCES

AAMHatch (2003), Corangamite CMA airborne laser survey data documentation, AAMHatch Pty Ltd, Melbourne, Australia.

Albani, M., B. Klinkenberg, D. W. Andison and J. P. Kimmins (2004), The choice of window size in approximating topographic surfaces from digital elevation models, *International Journal of Geographical Information Science*, 18(6), 577-593.

Ali, T. A. (2004), On the selection of an interpolation method for creating a terrain model (TM) from LIDAR data, In: *Proceedings of the American Congress on Surveying and Mapping (ACSM) Conference 2004*, Nashville TN, U.S.A.

Anderson, E. S., J. A. Thompson and R. E. Austin (2005a), LiDAR density and linear interpolator effects on elevation estimates, *International Journal of Remote Sensing*, 26(18), 3889-3900.

Anderson, E. S., J. A. Thompson, D. A. Crouse and R. E. Austin (2005b), Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM, *Geoderma*, 132(3-4), 406-415.

Bjørke, J. T. and S. Nilsen (2002), Efficient representation of digital terrain models: compression and spatial decorrelation techniques, *Computer and Geosciences*, 28, 433-445.

Blaschke, T., D. Tiede and M. Heurich (2004), 3D landscape metrics to modelling forest structure and diversity based on laser scanning data, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(8/W2), 129-132.

Chaplot, V., F. Darboux, H. Bourennane, S. Leguédois, N. Silvera and K. Phachomphon (2006), Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density, *Geomorphology*, 77(1-2), 126-141.

Chou, Y. H., P. S. Liu and R. J. Dezzani (1999), Terrain complexity and reduction of

topographic data, *Geographical Systems*, 1(2), 179-197.

Desmet, P. J. J. (1997), Effects of interpolation errors on the analysis of DEMs, *Earth Surface Processes and Landforms*, 22(6), 563-580.

Fisher, P. F. and N. J. Tate (2006), Causes and consequences of error in digital elevation models, *Progress in Physical Geography*, 30(4), 467-489.

Florinsky, I. V. (1998), Combined analysis of digital terrain models and remotely sensed data in landscape investigations, *Progress in Physical Geography*, 22(1), 33-60.

Florinsky, I. V. (2002), Errors of signal processing in digital terrain modeling, *International Journal of Geographical Information Science*, 16(5), 475-501.

Gong, J., Z. Li, Q. Zhu, H. Shu and Y. Zhou (2000), Effects of various factors on the accuracy of DEMs: an intensive experimental investigation, *Photogrammetric Engineering and Remote Sensing*, 66(9), 1113-1117.

Hu, Y. (2003), *Automated extraction of digital terrain models, roads and buildings using airborne LiDAR data*, (PhD Thesis), Department of Geomatics Engineering, The University of Calgary, Calgary, Alberta, Canada, pp. 206.

Kidner, D. B. and D. H. Smith (2003), Advances in the data compression of digital elevation models, *Computer and Geosciences*, 29, 985-1002.

Kienzle, S. (2004), The effect of DEM raster resolution on first order, second order and compound terrain derivatives, *Transactions in GIS*, 8(1), 83-111.

Li, Z., Q. Zhu and C. Gold (2005), Digital Terrain Modeling: Principles and Methodology, CRC Press, Boca Raton, London, New York, and Washington, D.C.

Liu, X., Z. Zhang, J. Peterson and S. Chandra (2007), LiDAR-derived high quality ground control information and DEM for image orthorectification, *GeoInformatica*, 11(1), 37-53.

Lloyd, C. D. and P. M. Atkinson (2006), Deriving ground surface digital elevation models from LiDAR data with geostatistics, *International Journal of Geographical Information Science*, 20(5), 535-563.

Lohr, U. (1998), Digital elevation models by laser scanning, *Photogrammetric Record*, 16, 105-109.

McCullagh, M. J. (1988), Terrain and surface modelling systems: theory and practice, *Photogrammetric Record*, 12(72), 747-779.

Pradhan, B., S. Kumar, S. Mansor, A. R. Ramli and A. R. B. M. Sharif (2005), Light detection and ranging (LiDAR) data compression, *KMITL Journal of Science and Technology*, 5(3), 515-523.

Robinson, G. J. (1994), The accuracy of digital elevation models derived from digitised contour data, *Photogrammetric Record*, 14(83), 805-814.

Wehr, A. and U. Lohr (1999), Airborne laser scanning - an introduction and overview *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(4), 68-82.

Yang, X. and T. Hodler (2000), Visual and statistical comparison of surface modeling techniques for point-based environmental data, *Cartography and Geographic Information Science*, 27(2), 165-175.

Yilmaz, H. M. (2007), The effect of interpolation methods in surface definition: an experimental study, *Earth Surface Processes and Landforms*, 32(9), 1346-1361.

Zhou, Q., X. Liu and Y. Sun (2006), Terrain complexity and uncertainties in grid-based digital terrain analysis, *International Journal of Geographical Information Science*, 20(10), 1137-1147.

Zimmerman, D., C. Pavlik, A. Ruggles and M. P. Armstrong (1999), An experimental comparison of ordinary and universal Kriging and inverse distance weighting, *Mathematical Geology*, 31(4), 375-389.