2011 4th International Congress on Image and Signal Processing

The Effects of Data Reduction on LiDAR-based Digital Elevation Models

Jaco Immelman
Department of Geography
University of South Africa
Pretoria, South Africa

Lukas G.C. Scheepers
Department of Geography, Environmental Management and Energy Studies
University of Johannesburg
Johannesburg, South Africa

***Abstract*—LiDAR data enables highly accurate terrain representations, however, various applications are hampered by data handling efficiency; specifically lengthy processing times. To address this, both point density reductions and the use of various resolution grids are compared as data reduction methods to test their effects on the accuracy and handling efficiency of the derived Digital Elevation Model (DEM). A series of point densities of 1%, 10%, 25%, 50% and 75% were interpolated along a range of horizontal resolutions (1-, 2-, 3-, 4-, 5-, 10-, and 30- m). Results indicate that resolution reduction provides the most efficient DEMs in terms of their data handling. DEMs generated at a 3 m resolution using all of the data points deviated less than 6% from the $_{1m}$DEM$_{100\%}$, while significantly only taking 10% of the processing time. Resolution reduction provided sufficient accuracies for varying terrain complexities.**

Keywords-digital elevation model; data reduction; data handling efficiency; resolution

I. INTRODUCTION

Topography is a spatial aspect that finds relevance in many sub-disciplines, both in- and outside geography. Accurate elevation data provides reliable topographic information for various applications. Light Detection and Ranging (LiDAR) has been described as the most effective technique for elevation data acquisition, providing accurate and reliable elevation data [1, 2]. The use of LiDAR as an active remote sensing technique provides increased accuracy in the data collection phase, allowing for accurate datasets with high point densities [3].

Higher point densities increase the likelihood that all terrain variances will be mapped, thereby increasing the accuracy of DEMs. However, the high point density of LiDAR is often associated with increased storage requirements and a lack of sufficient processing capacity. Data reduction techniques, such as point density reduction and the use of low resolution grids, aim to reduce the storage requirements and lengthy processing times associated with inefficient datasets. Random point reduction has shown that datasets can be reduced significantly while still allowing for accurate terrain parameter extraction [4; 5]. Studies show that as little as 50% of the original dataset can be used without severely affecting the accuracy of the terrain representations [3; 4; 5; 6].

Point reduction should however be matched to the grid resolution. Ideally, the number of cells should equal the number of points in the dataset [7]. Alternatively, [8] suggests that the horizontal resolution for grid DEMs should vary between 2.25 and 3.25 meters. Interestingly, studies [3; 5 and 9] following the principle that the number of cells should equal the number of points were also within the range suggested by [8]. This range of grid resolutions provides an indication for the most accurate DEM, rather than the most efficient DEM. When determining grid resolution for the most efficient DEM, minimum and maximum resolution thresholds exist; especially when selecting the grid size for specific point densities [6].

Point density, interpolation method and terrain morphology should be carefully considered when selecting the optimal grid resolution [1; 3, 4, 5; 10, and 11]. Various authors [3; 4; 5; 6] have found that original LiDAR datasets can be reduced with as much as 50%. Critically, terrains in these studies are only moderately complex [3], vary by 8m [6] to 30m [5], or their complexity was not discussed at all [5]. Data reduction thus needs to be investigated over a wider selection of terrain complexities.

To date, research has focused on the effects of data reduction on the accuracy of the reduced datasets instead of the effect of data reduction on the data handling efficiency. This study investigates the effects of combined data reduction techniques on the accuracy and handling efficiency of LiDAR data. The objectives for the study are to determine the effects of point density reductions, coupled with various horizontal resolution selections, on (a) DEM accuracy; (b) the handling efficiency of derived DEMs; and (c) to investigate the effect of terrain complexity on reduced datasets in terms of accuracy and data handling efficiency.

II. METHODS AND MATERIALS

A. Data

LiDAR data of the Rio Tinto Palabora Mine are used in this study. The survey was conducted with an aircraft mounted with an Optech ALTM 3100 EA laser scanner, scanning at 70 kHz. The average spacing between the points is 0.79 cm with a mean data density of 1.26 points per m^2. The data provided was already filtered into ground returns and non-ground returns, however, the filtering algorithm used was not provided. The filtered data was assumed to be the most accurate data

978-1-4244-9306-7/11/$26.00 ©2011 IEEE

available, even though the process of filtering often introduces some form of error into the dataset [12].

B. Study Area

The study area covers the extent of the Rio Tinto Palabora Mine (Fig. 1), south of the town of Phalaborwa, in the Limpopo province of South Africa. The extent of the area is 15007.29 Ha. The area is undulating, with relative terrain variances. Because of opencast mining activities, terrain variances range from 11 m up to 534 m. Five study sites were selected to cover a range of terrain complexities. To determine the terrain complexity for each area, terrain descriptions such as range of elevation, mean elevation, average slope, and the standard deviation of the average slope were calculated (Table I). Fig. 2 illustrates the terrain morphology, as generated by LAStools.

C. Methodology

In order to investigate the effects of data reduction on the accuracy of the derived DEMs, point reduction and low resolution grids were used as data reduction techniques. The original ground return files were imported into a database (PostgreSQL), where files were reduced to 75%, 50%, 25% 10% and 1% of the original dataset using random point reduction. The percentages selected were based on previous work of [4; 5; and 6]. A test set (10% of the data points) was also extracted randomly to determine the accuracy of the DEMs derived from reduced data sets.

DEMs were created from each of the reduced datasets of the five study sites using Inter Distance Weighted (IDW) interpolation. Simple interpolators (such as IDW) perform well in areas where the point density is high [4; 13, and 14] and adds to the data handling efficiency [3]. Interpolation and processing was completed in Quantum GIS 1.6.0.

The Interpolation plug-in was used to interpolate the reduced datasets at 1-, 2-, 3-, 4-, 5-, 10- and 30-meter resolution. It is important to note that the study did not optimize the interpolation routine for the generation of DEMs, as the scope of the study was to investigate the effects of data reduction.

Elevation models derived from reduced data sets were compared to a model derived from all the data points, using the highest resolution ($_{1m}\mathrm{DEM}_{100\%}$). The study accepted the $_{1m}\mathrm{DEM}_{100\%}$ as the most accurate terrain representation that can be derived from the original dataset. DEM accuracy was measured using a test set, comprising of 10% of data points. The test set was subtracted from the generated DEMs and the mean of the absolute values was used to evaluate DEM accuracy. The deviation of derived DEMs was determined by comparing the accuracies of the derived DEMs to the most accurate model ($_{1m}\mathrm{DEM}_{100\%}$). Processing times were recorded and compared in terms of the time that it took to generate each of the DEMs, as a percentage of the time that the highest accuracy DEM processed.

TABLE I TERRAIN DESCRIPTION

Study sites description	**Terrain descriptive statistics**					
	Min (m.a.s.l)	***Max (m.a.s.l)***	***Mean (m.a.s.l)***	***Std Dev (m.a.s.l)***	***Mean Slope (deg)***	***Std Dev Slope (deg)***
Study site 1	395.93	454.56	413.79	8.76	6.69	7.32
Study site 2	350.45	373.26	361.41	4.26	5.51	5.39
Study site 3	-99.71	422.11	190.01	137.20	49.24	20.42
Study site 4	381.66	460.09	420.32	20.32	16.66	11.35
Study site 5	279.31	401.98	122.05	28.44	18.30	14.09

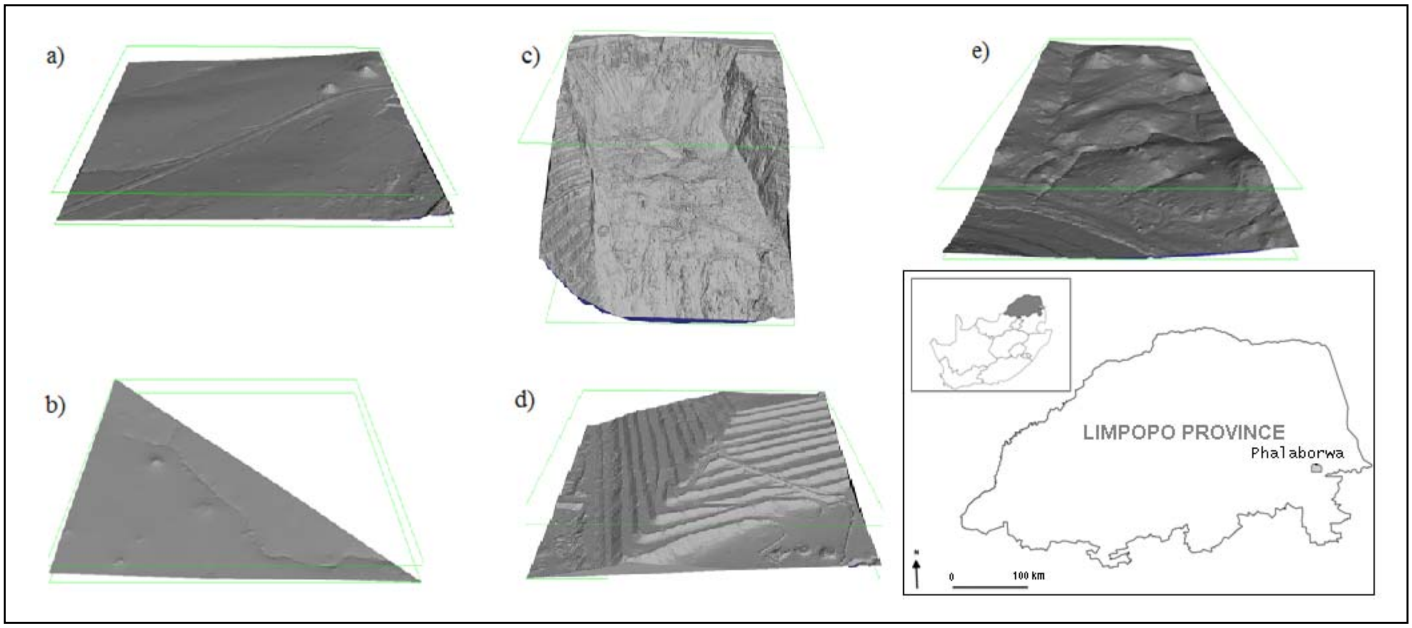

Figure 1: LASview illustration of the study sites. a) Study site 1, b) Study site 2, c) Study site 3, d) Study site 4, e) Study site 5 and location of the study area

RESULTS AND DISCUSSION

As expected, the accuracy of DEMs decreased as the datasets were reduced with severe differences between datasets that were conservatively and aggressively reduced. Low resolution grids had a negative impact on the accuracy of the derived DEMs; with the lowest resolution DEMs providing the least accurate results. Fig. 3 shows a drastic increase in deviation between the original DEM (${}_{1m}\mathrm{DEM}_{100\%}$) and the DEMs generated at the lowest resolution (30 meters) for all of the point densities tested. Similarly, the 1% dataset varies significantly for all the study sites. DEMs derived from 1% of the dataset, and DEMs derived at a 10-, and 30- meter resolution were not considered in the results of this study

As a comparison to previous work suggesting that datasets can be reduced up to 50%, the deviation of reduced DEMs were compared to DEMs created using 50% of the dataset, at a one meter horizontal resolution (${}_{1m}\mathrm{DEM}_{50\%}$). Terrain representations that were derived from 50% of the data deviated by an average of 12% from the original DEM. While previous studies did not focus on the deviation of the reduced DEMs from the original DEM, the difference between the RMSE values of the 50% and 100% datasets was between 5.4% [3] and 9.1% [4]. The deviation of the 50% point reductions in this study varied between 7% and 14%, which is similar to the deviation of previous studies. Importantly, the deviation of previous studies was obtained from areas with only moderate terrain complexities. Even though the results of this study show some variance between areas of different terrain complexity, the ${}_{1m}\mathrm{DEM}_{50\%}$ still show deviations of less than 15% for all terrain complexities tested. While differences between deviations of simple and complex terrain exist, no clear correlation could be determined due to a lack of measure for terrain complexity. If terrain complexity was classified according to a terrain complexity index, comprising of the range of elevation, mean slope and the standard deviation of the slope, it could be possible to determine the correlation between terrain complexity and DEM accuracy.

Table II illustrates that conservative resolution reduction is a feasible data reduction technique. If all data points are used to generate a DEM with a lower resolution (${}_{2m}\mathrm{DEM}_{100\%}$), the derived DEMs deviate very little from the ${}_{1m}\mathrm{DEM}_{100\%}$ and are more accurate than point density reduction (${}_{1m}\mathrm{DEM}_{50\%}$). However, while grid resolution could be lowered conservatively, low resolution grids (5m) vary in their accuracy. Moderately complex terrain (study site 1 and 2) could be represented adequately by lower resolution grids (5m), whereas the accuracies of representations of more complex terrains (study site 3, 4 and 5) were inconsistent when using ${}_{5m}\mathrm{DEM}_{100\%}$.

Reference [5; 3, and 9] used a two meter resolution as the highest resolution while concluding that half of the points in the dataset represented the terrain adequately. In contrast, the results of this study showed that the accuracies of ${}_{2m}\mathrm{DEM}_{50\%}$ were inconsistent. This could be attributed to the difference in terrain complexity between the studies, since the ${}_{2m}\mathrm{DEM}_{50\%}$ of more complex terrain (study sites 3, 4 and 5) show greater deviations from the ${}_{1m}\mathrm{DEM}_{50\%}$ (17.40%, 37.67% and 72.98% respectively).

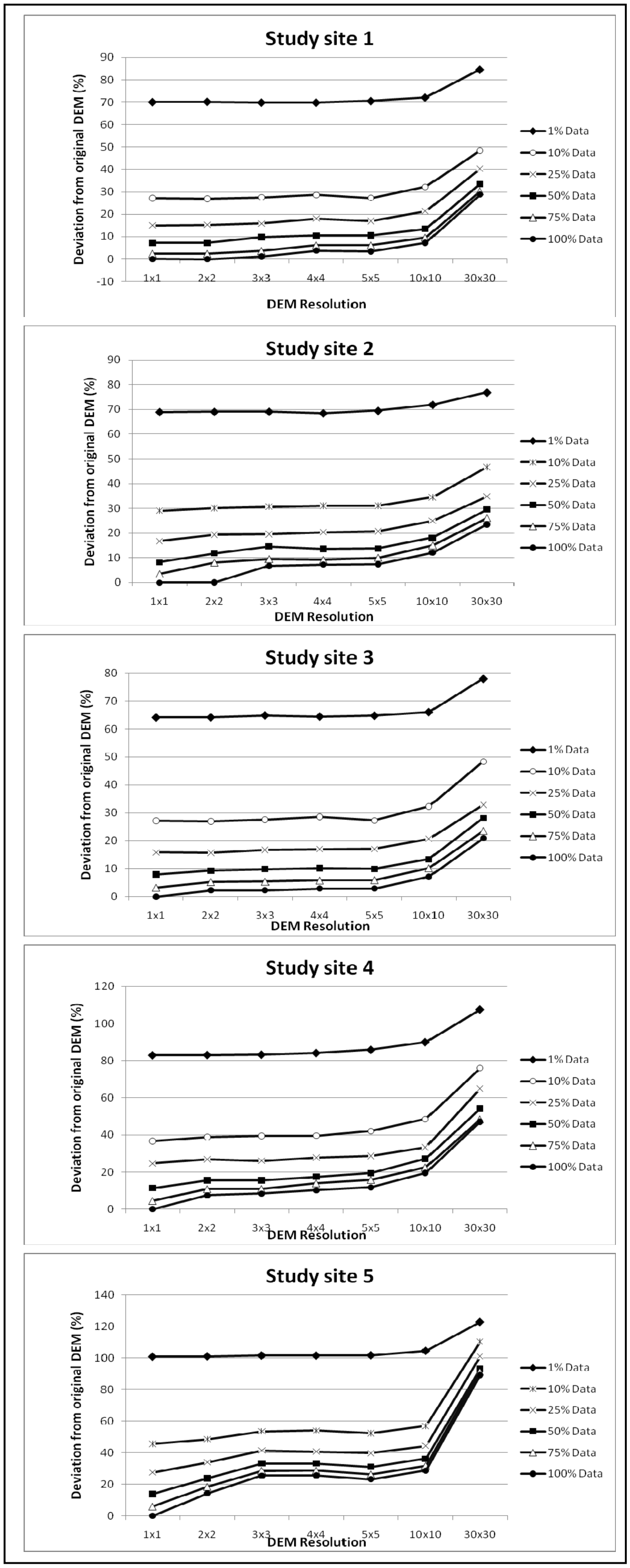

Figure 2: Deviation of DEMs derived from reduced datasets

Originally, grid resolutions of 1-, 2-, 5-, 10-, and 30- m were selected, based on the grid resolution used by [6]. However, it becomes apparent (Table II) that while a two meter grid resolution is sufficient for 100% and 75% datasets, five meter grid resolution provides less consistent results; especially ${}_{5m}DEM_{75\%}$. The ${}_{2m}DEM_{100\%}$ and ${}_{2m}DEM_{75\%}$ could be seen as an adequate representation, with less than 5% deviation from ${}_{1m}DEM_{50\%}$ (except ${}_{2m}DEM_{75\%}$ for study site 5). In contrast, a five meter grid resolution is adequate for only half of the DEMs generated from 100% and 75% of the data. Even without determining the factors affecting the decrease in accuracy, we can determine that a five meter resolution will not always provide adequate terrain representations. A five meter grid resolution thus lies outside the threshold resolution (an accuracy of less than 5% deviation from the ${}_{1m}DEM_{50\%}$). Three and four meter resolution DEMs were thus included in the study.

Interestingly, both ${}_{3m}DEM_{75\%}$ and ${}_{4m}DEM_{100\%}$ are more accurate than the ${}_{1m}DEM_{50\%}$, for all of the study sites except study site 5 (Table III). The three and four meter DEM of study site 5 (for all data percentages) shows a greater deviation than the ${}_{5m}DEM_{100\%}$, and should be regarded as an anomaly. Table III illustrates that, in terms of accuracy, lower resolution grids generated from high point densities (100% and 75%) are more accurate than high resolution grids, generated from lower point densities (such as 50% data density).

More importantly, when comparing the processing times of the ${}_{3m}DEM_{100\%}$ to the processing times of the ${}_{1m}DEM_{50\%}$, a three meter grid resolution provides more accurate results at a fraction of the processing time (Table IV). In fact, even compared to the ${}_{2m}DEM_{50\%}$ used in previous studies, the ${}_{3m}DEM_{100\%}$ provides a DEM that is more accurate (for various terrain complexities), with better handling efficiency.

TABLE II DEVIATION OF DERIVED DEMS

Devation from original DEM (%)[a]	**Point density reduction**	**Resolution reduction**		**Resolution and point density reduction**			
	50% data 1m resolution	***100% data 2m resolution***	***100% data 5m resolution***	***75% data 2m resolution***	***75% data 5m resolution***	***50% data 2m resolution***	***50% data 5m resolution***
Study site 1	7.2	-0.1 (-100.7)	3.4 (-52.8)	2.5 (-66.0)	6.2 (-14.0)	7.2 (-0.1)	10.7 (47.7)
Study site 2	8.2	0.1 (-99.15)	7.5 (-9.3)	8.0 (-2.3)	10.1 (22.0)	11.8 (43.7)	13.8 (67.11)
Study site 3	7.9	2.3 (-71.1)	2.9 (-63.4)	5.3 (-32.9)	6.0 (-24.6)	9.3 (17.4)	9.9 (25.1)
Study site 4	11.3	7.5 (-34.0)	11.7 (3.7)	10.8 (-4.8)	11.7 (39.4)	15.6 (37.7)	19.5 (72.6)
Study site 5	13.8	14.4 (4.4)	23.2 (68.8)	18.4 (33.5)	23.2 (90.9)	23.8 (73.0)	31.1 (126.1)

a. Values in brackets indicate the deviation from the 1mDEM50%.

TABLE III DEVIATION OF DERIVED DEMS

Deviation from original DEM (%)[b]	**Point density reduction**		**Resolution reduction**			**Resolution and point density reduction**		
	50% data 1m resolution	***50% data 2m resolution***	***100% data 2m resolution***	***100% data 3m resolution***	***100% data 4 resolution***	***75% data 2m resolution***	***75% data 3m resolution***	***75% data 4m resolution***
Study site 1	7.2	7.2 (-0.1)	-0.1 (-100.7)	1.2 (-83.9)	3.9 (-46.6)	2.5 (-66.0)	1.2 (-83.2)	3.7 (-49.1)
Study site 2	8.2	11.8 (43.7)	0.1 (-99.15)	6.7 (-18.6)	7.3 (-11.4)	8.4 (-2.4)	9.5 (15.3)	9.2 (11.3)
Study site 3	7.9	9.3 (17.4)	2.3 (-71.1)	2.3 (-71.3)	2.9 (-63.3)	5.3 (-32.9)	5.4 (-31.4)	5.87 (-26.0)
Study site 4	11.3	15.6 (37.7)	7.5 (-34.0)	8.4 (-25.8)	10.2 (-9.9)	10.8 (-4.8)	5.3 (-3.8)	13. (22.6)
Study site 5	13.8	23.8 (73.0)	14.4 (4.4)	25.5 (85.2)	25.6 (85.9)	18.4 (4.4)	28.6 (107.7)	28.7 (108.4)

b. Values in brackets indicate the deviation from the 1mDEM50%.

TABLE IV PROCESSING TIMES OF DERIVED DEMS

Processing time compared to $_{1}DEM_{100}$ (%)	**Point density reduction**		**Resolution reduction**			**Resolution and point density reduction**		
	50% data 1m resolution	***50% data 2m resolution***	***100% data 2m resolution***	***100% data 3m resolution***	***100% data 4m resolution***	***75% data 2m resolution***	***75% data 3m resolution***	***75% data 4m resolution***
Study site 1	66.5	16.6	33.5	15.2	8.6	24.9	10.6	6.3
Study site 2	55.6	19.4	37.5	12.1	6.0	18.5	8.4	3.0
Study site 3	43.2	10.8	21.6	9.7	5.8	16.9	9.6	4.1
Study site 4	92.9	11.4	22.9	10.0	5.7	18.6	8.6	4.3
Study site 5	45.8	6.3	23.6	5.83	4.86	27.1	4.38	3.61

When only using point density reduction the accuracy of $_{1m}DEM_{50\%}$ deviates between 7.2% and 13.8%. Lower grid resolutions and higher point densities have a smaller effect on the accuracy of the derived terrain representations. For instance, $_{3m}DEM_{100\%}$ deviates between 1.2% and 5.9% over varying terrain complexities (study site 5 excluded). Resolution reduction combined with conservative point reduction ($_{3m}DEM_{75\%}$) also provides considerably lower deviations, varying between 1.2% and 9.5%.

Significantly, while providing DEMs with smaller deviations from the $_{1m}DEM_{100\%}$, data handling efficiency improves drastically when using resolution reduction. On average, $_{1m}DEM_{50\%}$ processed 49.3% faster than the $_{1m}DEM_{100\%}$. Processing times of the $_{3m}DEM_{100\%}$ were on average 8.8% of the total processing time (91.2 % faster than the $_{1m}DEM_{100\%}$) and the $_{4m}DEM_{100\%}$ and $_{3m}DEM_{75\%}$ processed 4.6% and 6.9% of the total processing time respectively.

I. CONCLUSION

The high data volumes associated with LiDAR often require some form of data reduction to increase data handling efficiency. Critically, the terrain should still be represented accurately. References [3, 4, 5 and 6] provide evidence that LiDAR datasets can be reduced without severely affecting the accuracy of the derived elevation models. DEMs derived from 100% of the data, at a two, three and four meter resolution provided more accurate results than DEMs derived from 50% of the data at one meter resolution. Results also showed that DEMs using a lower grid resolution can represent terrain of varying complexity adequately.

The results allow us to recommend the use of resolution reduction, or point density reduction coupled with resolution reduction as data reduction technique. Point density reduction requires more aggressive reduction to achieve better data handling efficiency, which can lead to inaccurate terrain representations. In contrast, the use of grid resolution reduction provided accurate terrain representations, requiring less aggressive reduction to achieve optimal data handling efficacy.

While a combined data reduction technique provides computationally efficient results, the accuracy of the derived DEMs can vary. When using a combined data reduction technique, our results suggest a threshold of $_{3m}DEM_{75\%}$, while the threshold for using only resolution reduction is $_{4m}DEM_{100\%}$. Importantly, users should have an understanding of the compromise between optimal accuracy and optimal data handling efficiency. Various applications might require different accuracies, thus the selection of data reduction technique should be determined by the acceptable level of accuracy. The resolutions and data percentages used in this study were based on previous studies. A further investigation using smaller point density and resolution intervals is necessary to further the understanding of the threshold data density.

Moreover, it would appear that areas with only moderately complex terrain provided more accurate results, but no correlation was found between the effects of data reduction on moderate and complex terrains. Thus a future research should investigate the relationship between terrain complexity and data reduction, using an index for terrain complexity.

ACKNOWLEDGMENT

We would like to express our gratitude to the Palabora Mining Company for the use of the data and to the Southern Mapping Company for supplying the data.

REFERENCES

[1] X. Liu, “Airborne LiDAR for DEM generation: some critical issues,” Progress in Physical Geography, vol. 31, no. 1, pp. 31–49, February 2008.

[2] M. E. Hodgson and P. Bresnahan, “Accuracy of airborne LiDAR-derived elevation: empirical assessment and error budget,” Photogrammetric Engineering & Remote Sensing, vol. 70, no. 3, pp. 331-339, March 2004.

[3] X. Liu, Z. Zhang, J. Peterson and S. Chandra, “The effect of LiDAR data density on DEM accuracy,” Proceedings of the International Congress on Modelling and Simulation, pp. 1363-1369, 2007 [MODSIM07, Christchurch, New Zealand].

[4] E. S. Anderson, J. A. Thompson and R. E. Austin, “LiDAR density and linear interpolator effects on elevation estimates,” International Journal of Remote Sensing, vol. 28, no. 18, pp. 3889-3900, September 2005.

[5] X. Liu and Z. Zhang, “LiDAR data reduction for efficient and high quality DEM generation,” The International Archives of the Photogrammtery, Remote Sensing and Spatial Information Sciences, vol. XXXVII, part B3b, Beijing, 2008.

[6] E. S. Anderson, J. A. Thompson, D. A. Crouse and R. E. Austin, “Horizontal resolution and data density effects on remotely sensed LiDAR-based DEM,” Geoderma, vol. 132, pp. 406-415, 2006.

[7] Y. H. Chou, P. S. Liu and R. J. Dezzani, “Terrain complexity and reduction of topographic data,” Journal of Geographic Systems, vol. 1, pp. 179-198, 1999.

[8] I. V. Florinsky and G. A. Kuryakova, “Determination of grid size for digital terrain modelling in landscape investigations- exemplified by soil moisture distribution at micro scale,” International Journal of Geographic Information Sciences, vol. 9, pp. 421-432.

[9] X. Liu, Z. Zhang, J Peterson and S. Chandra, “Large area DEM generation using airborne LiDAR data and quality control,” Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, pp. 79-85, 2008.

[10] C. W. Bater and N. Coops, “Evaluating error associated with LiDAR derived DEM interpolation,” Computers and Geosciences, vol. 35, pp. 289-300, February 2009.

[11] F. J. Aquilar, F. Aguera, M. A. Aquilar and F. Carvajal, “The effects of terrain morpology, sampling density and interpolation methods on grid DEM accuracy,” Photogrammetric Engineering & Remote Sensing, vol. 71, No. 1, pp. 805-816, July 2005.

[12] G. T. Raber, J. R. Jensen, S. R. Schill, and K. Schuckman, “Creation of digital terrain models using an adaptive LiDAR vegetation point removal process,” Photogrammetric Engineering & Remote Sensing, vol. 68, no. 12, pp. 1307-1315, December 2002.

[13] T. A. Ali, “On the selection of interpolation method for creating a terrain model (TM) from LiDAR data,” Proceedings of the American Congress on Surveying and Mapping (ACSM) Conference 2004, Nashville, TN, USA

[14] V. Chaplot, et al, “Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density,” Geomorphology, vol. 77, 126-141, February 2006