RASTERIZING AIRBORNE LASER SCANNING POINT CLOUDS BY BLOCK KRIGING

Wenxia Tan*, Jonathan Li , Yu Li

Department of Geography & Environmental Management, Faculty of Environment, University of Waterloo
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada -{twenxia, junli, y62li}@fes.uwaterloo.ca

KEY WORDS: Airborne laser scanning, Point clouds, Rasterization, Block Kriging, Geostatistics

ABSTRACT:

Airborne laser scanning (ALS) is increasingly becoming a standard method for the collection of dense elevation models, especially in 3D urban mapping. However, automation in processing of ALS point-clouds involves handling huge datasets, irregular point distribution, multiple views, and relatively low textured surfaces. Since raster data structure is the most commonly used data representation method and is relatively easy to store and process, there comes a need to convert the ALS point clouds into raster data format. Since the ALS point clouds are rarely at the same location as the centre of the discretization grid, approximation is therefore required. A simple and most often used method is selecting a known point value to represent the grid. Such a point-to- point transformation often leads to serious information loss. Transformation of the ALS point clouds to grid is known as a special case of the change of support, because it changes the data volume from point to area. In this paper, we present block Kriging to model this kind of change of support towards rasterization of the ALS point clouds. The mathematic and algorithmic formulations are illustrated. Results from the UW campus show that the proposed method can better preserve the information in the ALS point clouds than the point-to-point transformation. Quality assessment is designed and conducted to evaluate the performance of block Kriging. Detailed error analysis is also provided to illustrate the accuracy of the proposed method.

1. INTRODUCTION

Airborne laser scanning (ALS) or light detection and ranging (LIDAR) is a rapidly emerging technology in photogrammetry, remote sensing, surveying and mapping communities, which provides high accurate Earth's surface contour information for the generation of digital elevation models (DEMs), three-dimensional (3D) city models, and 3D vegetation mapping (Ackermann, 1999; Baltsavias, 1999; Wehr and Lohr, 1999; Haala, and Brenner, 1999; Shan, and Sampath, 2005; Koch *et al*., 2006). A typical ALS system consists of a platform (e. g., a helicopter or an aircraft) and an integrated sensor system including a laser scanner, a Global Positioning System (GPS) receiver, and an inertial measurement unit (IMU). Raw ALS data acquired by an ALS system is usually characterized as a set of sub-randomly distributed points in three-dimensional (3D) space, called 3D point clouds, with the x, y coordinates specifying the geographical location and the z coordinate the elevation. Recently, many methods have been proposed to use the data acquired by the ALS system to generate dense digital elevation models (DEMs) (Brovelli and Cannata, 2002; Sithole and Vosselman, 2004; Ma, 2005).

Generally, ALS data are represented with three basic data structures: point clouds, raster models, and triangulated irregular network (TIN). Point clouds contain all the original information, but the data volume is very huge and this makes the processing and application of ALS data difficult. While TIN is more flexible and fewer point is needed to be stored to present the terrain, it is not suitable for information extraction. Compared to the point clouds and TIN, raster models are the most common data representation method and are relatively easy to store and process. Moreover, most of the existing digital image processing algorithms can be used for raster data processing. To this end, there comes a need to convert the ALS point clouds into raster data format.

Although ALS enables point sampling at very small separation distances, subsequent prediction from points to a grid (altitude matrix) is subject to much uncertainty. This research focuses on the use of Kriging, an optimal technique for unbiased spatial prediction, to derive a digital surface model (DSM) from ALS point clouds. Although ordinary Kriging and universal Kriging have been used for this purpose (Stein, 1999), there is no systemic investigation on the effects of terrain morphology, sampling density, and different Kriging techniques for ALS point clouds on the accuracy of interpolated heights in a raster DSM. This research will focuses on this issue. As shown in Figure 1, irregularly distributed point clouds (left) are interpolated into gird (right).

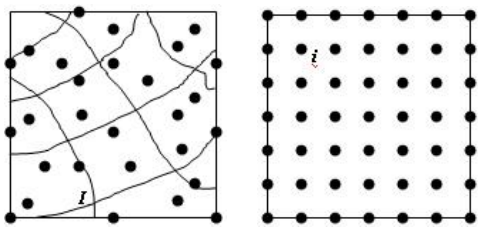

Figure 1. Rasterizing ALS points

2. BACKGROUND

2.1 Spatial interpolation and Geostatistics

Since the ALS point clouds are rarely at the same location as the centre of the discretization grid, spatial interpolation should be used. It is a procedure of estimating the value of a field

* Corresponding author. Wenxia Tan, PhD student,- twenxia@fes.uwaterloo.ca.

variable at unsampled sites within the area covered by sample locations. Normally, three kinds of approximation are often used. (1) If there is only one point in one grid, then the point is used to represent the value of the grid. (2) If more than one point in one grid, the average value of the points is used, or the median, maximum, or minimum one is chosen. (3) If no data point is in the grid, the nearest point to the grid centre is used. In conclusion, the above methods all use the value of a certain point to represent the grid of interest. This scheme usually leads to part of the natural dispersion not reflected by the data, and causes information loss. In order to better represent the data, geostatistic approach for spatial interpolation is considered by modelling the spatial correlation among the data points in a certain neighbourhood.

A fundamental assumption for geostatistical methods is that any two locations that are a similar distance and direction from each other should have a similar difference squared. This relationship is called stationary. If the spatial process is isotropy, spatial autocorrelation may depend only on the distance between two locations. The rate at which the correlation decays can be expressed as a function of distance. If the process is second order stationary, the covariance between any two random errors depends only on the distance and direction that separates them, not their exact locations. Semivariogram is a common tool to capture the second-moment structure of spatial data. It describes the variability of two locations of the data separated by a distance *h* (Oliver *et al*., 2005):

$$\gamma(s,h)=\frac{1}{2}\mathrm{Var}\left[Z(s)-Z(s+h)\right] \quad (1)$$

Where $\gamma(s,h)$ is the semivariogram, and $Z(s)$ is the data value at location s.

Wallace and Marsh (2005) used geostatistics to extract measures that characterize the spatial structure of vegetated landscapes from satellite image for mapping endangered Sonoran pronghorn habitat. They use variogram parameters to discriminate between different species-specific vegetation associations. Woolard and Colby (2002) used DEM generated from ALS data and spatial statistics to better understand dune characterization at a series of spatial resolutions.

Digital images are rich in data, but in many instances they are so complex as to require spatial filtering to distinguish the structures in them and facilitate interpretation. The filtering can be done geostatistically by Kriging analysis. It proceeds in two stages. The first involves modeling the correlation structure in an image by decomposing the variogram into independent spatial components. The second takes each component in turn and kriges it, thereby filtering it from the others.

In Lloyd and Atkinson (2002), inverse distance weighting, ordinary Kriging and Kriging with a trend model are assessed for the construction of DSMs from ALS data. Factorial Kriging is a geostatistical technique that allows the filtering of spatial components identified from nested variograms.

2.2 Block Kriging

The elevation of an area of interest (AOI) D can be modeled as the random field $\{Z(s): s \in D \subset R^2\}$. The ALS point clouds covering the AOI can be considered as the collection of independent observations at location $s = \{s_1, \ldots s_n\}$ on the random field, and denotes by the data vector, $Z(s) = \{Z(s_1), \ldots, Z(s_n)\}$.

Raster representation of the random field means to lattice the continuous domain B and then calculate a typical elevation for each grid (pixel). Let the region of a given pixel be *B* and the corresponding area be $|B|$. The elevation of the pixel can be predicted by calculating the average value of the random field in *B*:

$$Z(B)=\frac{1}{|B|}\int_B Z(\boldsymbol{s})d\boldsymbol{s} \quad (2)$$

Since the exact value of *Z(B)* cannot be calculate directly, we can only predict *Z(B)* using the observed data. Therefore, a window is set up around the grid *B*, raw data points

$$s_w=\{s_k \in s; s_k \in B\} \quad (3)$$

Then, the window *B* are used to predict *Z(B)*. With block Kriging predictor:

$$p(Z,Z(B))=\sum_{k=1}^{m}\lambda_k Z(s_k) \quad (4)$$

where λ_k are chosen to minimize the mean-squared prediction error between *p(Z,Z(B))* and real, unknown elevation of *B*:

$$E\left[\left(E\left[Z(B)\middle|\boldsymbol{Z(s)}\right]-Z(B)\right)^2\right] \quad (5)$$

Under this circumstance, *p(Z,Z(B))* is an unbiased prediction of *Z(B)*, the optimal weights $\{\lambda^k\}$ can be obtained by:

$$\lambda'=\left(\sigma(B,s)\right)+1\frac{1-1\Sigma^{-1}\sigma(B,s)}{1'\Sigma^{-1}1})\Sigma^{-1} \quad (6)$$

where the elements of the vector $\sigma(B, s)$ are $\mathrm{Cov}[Z(B), Z(s_k)]$. By discrentizing *B* into points, $\{\mu_j\}$, the point to block covariance can be approximated using

$$Cov[Z(B),Z(s)] \approx 1/N\sum_{j=1}^{n}C(u_j^{'},s) \quad (7)$$

Where $\sum$ is the matrix composed by the covariance of every two observed points. Due to the autocorrelation of spatial process, the elevation values in a small region always have a constant mean $u(s)$, the covariance between observations $Z(s_i)$, $Z(s_j)$ is:

$$\begin{aligned} &Cov(Z(s_i),Z(s_j)) \\ &=E[\{Z(s_i)-u(s)\}\{Z(s_j)-u(s)\}] \end{aligned} \quad (8)$$

As shown in figure 2, block Kriging is a method which uses the value of a block to represent the value of the grid.

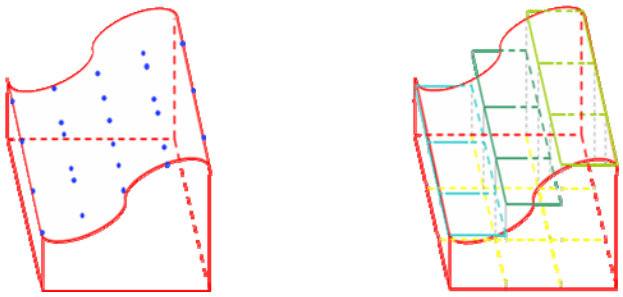

Figure 2. Figure placement and numbering.

3. OUR METHOD

3.1 Study area and data

The raw ALS dataset covered the main campus of the University of Waterloo (UW), Waterloo, Ontario was acquired by the Airborne Laser Terrain Mapper (ALTM) manufactured by Optech (http://www.optech.ca/) on March 11, 2006. The average flying height was 1,200 m above ground level and the flying speed was 66.9 *m* per second. The scan angle is 20º. The desired resolution is 0.908 *m*. The raw data contain more than seven millions point clouds.

A subset of the raw data (64 *m* by 64 *m*) is used in this study. It contains 5280 points. So, the point density in the scene is 1.28 point/m^2. Figure 3 shows the study area by Orthoimage which is part of the UW main campus. This area was selected since it offers an ideal site for studying the effectiveness of selected spatial statistics approach: it contains different kinds of objects such as trees, bare ground, buildings, and parking lots. Figure 4 shows the raw ALTM data point of the study area. It is shown that the point density is not uniformly distributed in the whole study area. The point density in the left lower part is bigger than other area. Figure 3 shows that the lower left of the study area was covered by trees. So the higher density was due to the multi-return from trees.

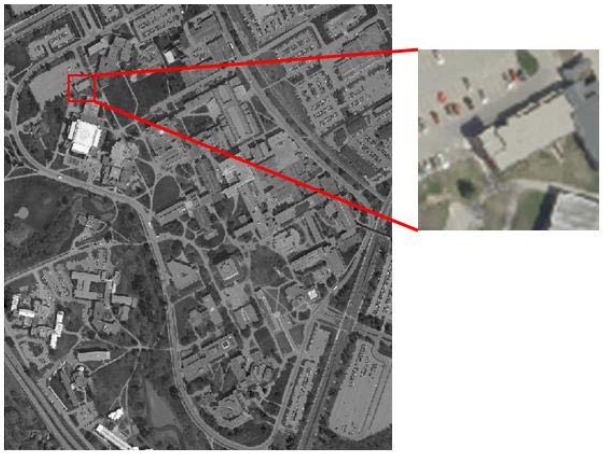

Figure 3. Orthoimage of the study area – UW main campus.

3.2 Exploratory data analysis

The first stage of Kriging requires a preliminary analysis of the raw data to determine which type of Kriging should be employed. The histogram shown in Figure 5 illustrates the distribution of the height values used in this investigation. The histogram displays a slightly skewed towards low elevations. So the distribution of the height is fairly normal.

Semivariogram is a measurement of the spatial autocorrelation between the data point. As the distance between two data points increases, the value of the semivariogarm increases accordingly. When the distance reaches a certain value, the value of the semivariogram will increase very slowly, and not exceed a certain value. This certain value is called range. Its range is a measure of the distance threshold under which the data is correlated.

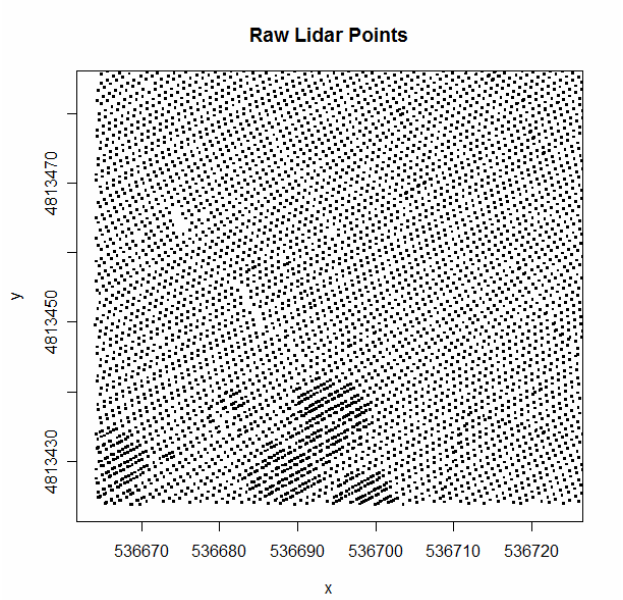

Figure 4. Raw ALTM points.

By measuring the distance between two locations and plotting the difference squared between the elevation values at locations, a semivariogram cloud is created. Shown in Figure 6, the x-axis is the distance between the locations, and on the y-axis is the difference of their elevation squared. Each dot in the semivariogram represents a pair of locations, not the individual locations on the map. We can see clearly from the semivariogram that in a distance of 25 *m*, the autocorrelation between the elevations are gradually decreased. So the range of the semivariogram is 25 *m*.

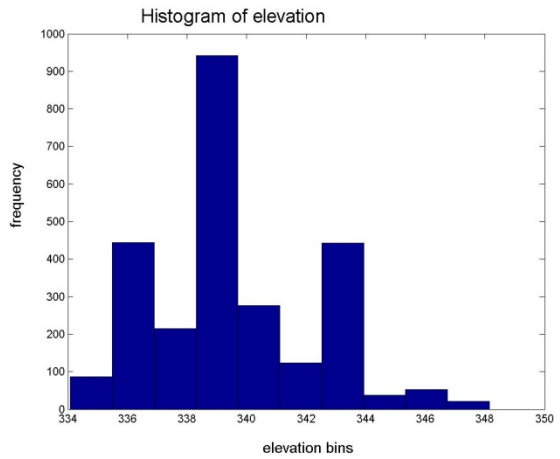

Figure 5. Histogram of elevation.

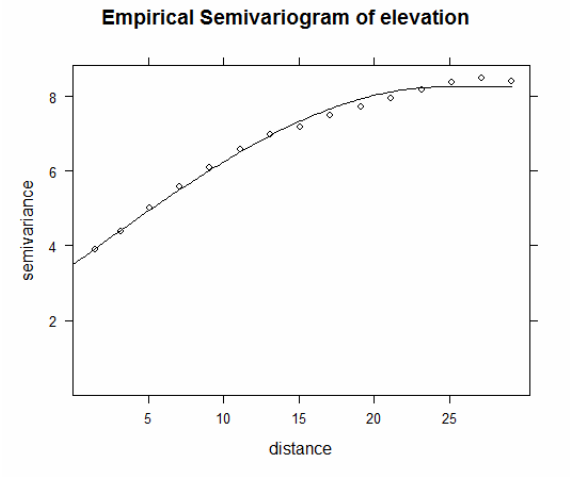

Figure 6. Semivariogram of elevation.

3.3 Block size and distance threshold

In the exploratory data analysis, the range of the semivariogram of the raw ALS data points was already estimated. It is 25 *m*, that's to say if the distance between two points exceeds 25 *m*, there should be no spatial autocorrelation between them. So when the elevation of a gird is estimated, only the data points in the distance of 25 *m* should be considered.

The block size is another important parameter for rasterizing the raw ALS point clouds. If small block size is used, then the result will be more concise. The total number of point cloud in the study area (64 *m* by 64 *m*) is 5280, so the point density is 1.23 points/m^2. On average, there is about one point in every square meter. In block Kriging, the covariance between the data points in the block and the points in the distance threshold are taken into account. Only when at least one point resides in the block, block Kriging could be used. In this study, the block sizes of 1 *m* × 1 *m*, 2 *m* × 2 *m*, 4 *m* × 4 *m*, are used, respectively.

3.4 Block Kriging

Figure 7 demonstrated the step by step procedure to compute the elevation of every gird. The input includes the raw ALS data point, the gird size, and the distance threshold. The distance threshold is used as a window for estimation. Only the data point in the window will be used for calculation. Then for every grid, First, calculate the covariance every two points the window. Second, calculate the covariance between the gird and every point. Third, the elevation is calculated by block Kriging by Equations (4) and (6) mentioned in Section 2.

4. RESULTS AND ACCURACY ASSESSMENT

Table 1 shows the error of ordinary Kriging and block Kriging in different block sizes. First, we could see that block Kriging did a much better job than ordinary Kriging. The standard error of block Kirging is less than 1 *m*, but the standard error of ordinary Kriging is about 4 *m*, which is unacceptable.

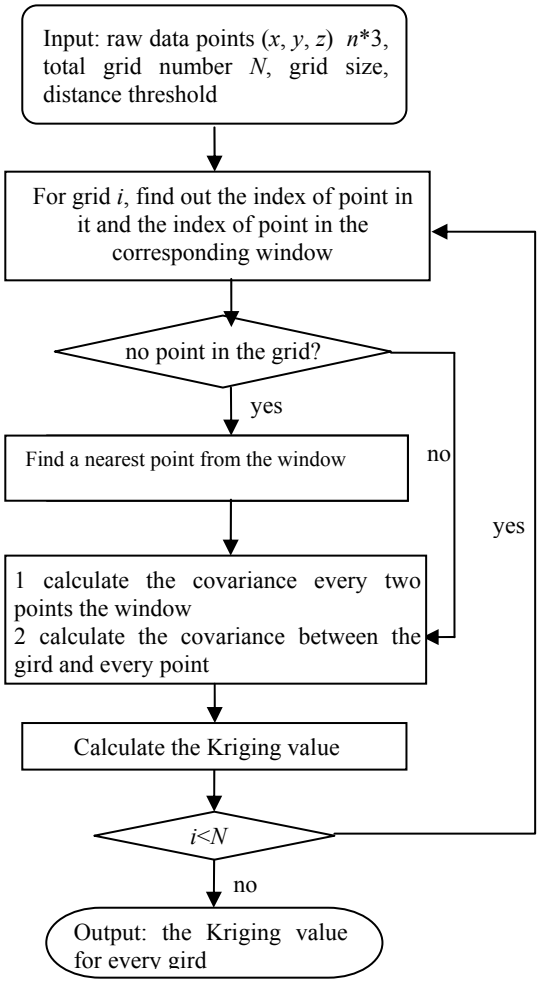

Figure 7. Flow chart of block Kriging computation.

Block size	Min	Median	Mean	Max
1 *m*	0.2098	0.3320	0.3353	0.7549
2 *m*	0.1631	0.2728	0.2796	0.6845
4 *m*	0.1274	0.1980	0.2169	0.5902
Ordinary Kriging	3.778	3.906	3.909	4.333

Table 1. Error assessment of different block sizes

Secondly, for block Kriging in different block size, the error is different. As the block size increases from 1 *m* to 2 *m* and 4 *m*, the mean standard error decrease from 0.33 to 0.27 and 0.21. This is because when bigger block size is used, more data points are considered, the estimated results are more likely to represent the true value. This could be further demonstrated in Figures 8 and 10, they are the 3D graph block Kriging. The block size is 1m by 1m and 2 *m* by 2 *m* respectively. As we can see, when the block size is 1 *m*, the result contains quite a lot of noise. When the block size increases to 2 m, the result is quite good.

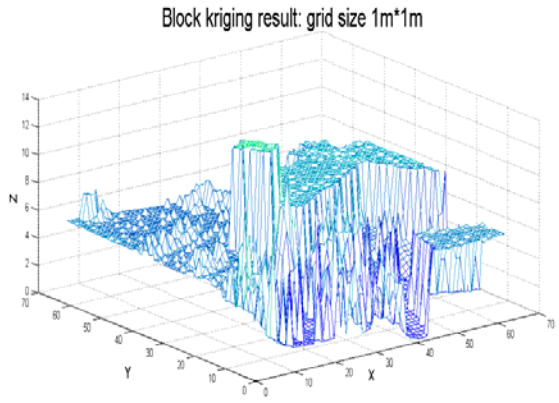

Figure 8. Block Kriging with a block size of 1 *m*.

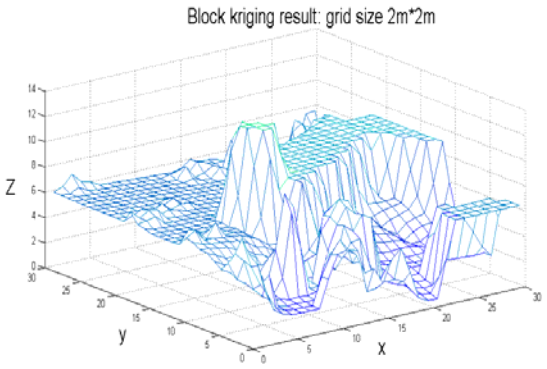

Figure 9. Block Kriging with a block size of 2 m.

Figures 9 and 10 show the Kriging results and the standard error in raster format, respectively. In Figure 10, each shaded square is centred at the intersection of the gird. Each grid is shaded with a different grey tone according to the elevation value estimated at that location. When the elevation is high, the colour is darker. From this figure, we could see Kriging could restore the scene when the change in elevation is continuous. But at the building edge, the boundary is blurred. That is due to the spatial averaging. Figure 11 is the corresponding standard error of Figure 10. This map allows evaluating the precision of estimation at any part of the region. The whiter squares correspond to areas with smaller error. When compared this error diagram with the raw data points shown in figure 4, the relationship can be easily found. The areas with high point density correspond to areas with low estimation errors. That's to say, as the point density increase, the estimation error decrease. Big error occurs at the boundary of the whole dataset, this was due to the lack of data at the boundary.

5. CONCLUSION

This paper provides a new approach to rasterize raw ALS point clouds. Block Kriging is used as an interpolation method to estimate elevations on a regular grid using irregularly spaced ALS point clouds. Firstly, the spatial structure of the data is analyzed by considering the spatial autocorrelation between data points. The spatial variability of the data is integrated into the estimation procedure of the semivariogram. Then, the data points in the block and other data points nearby are modelled, which leads to increased precision of the estimated elevation. So the standard error of block Kriging is much less than ordinary Kriging. Moreover, as the block size increase, more data points are considered when doing the Kriging, the error decreases. At the same time, the computation time increases quickly. So a balance between time and error should be made.

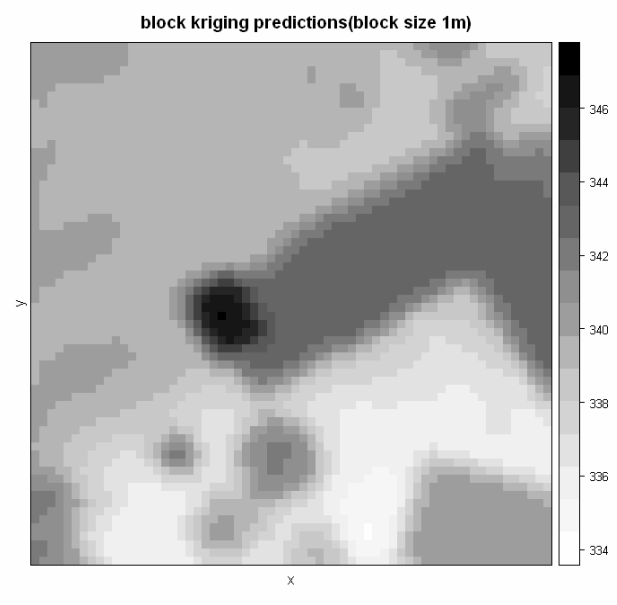

Figure 10. Raster representation of block Kriging of elevation (block size 1 *m*)

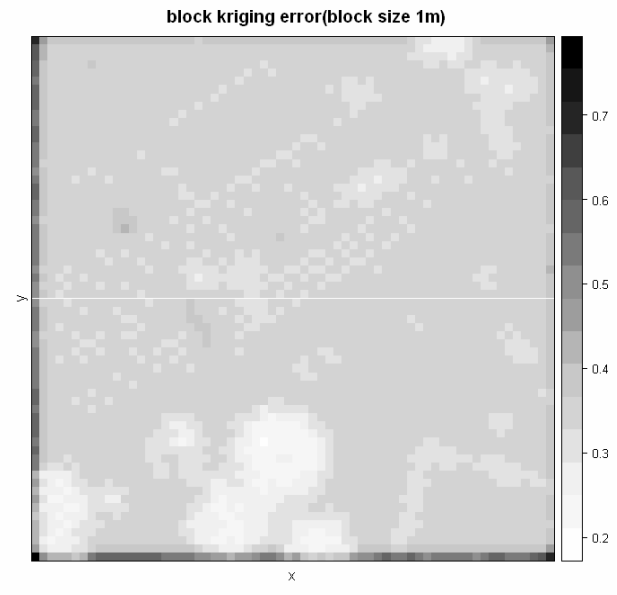

Figure 11. Raster representation of block Kriging standard error (block size 1 *m*)

Several points of the method proposed in this paper can be improved by further studies. For elevations at building edges, the change should not be continuous. So the problem about how to preserve edges in block Kriging needs further study. Moreover, the calculation time for block Kriging is quite long, for the dataset used in this study, it takes several minutes to do Kriging. A quick algorithm should be figured out for its use in large area.

ACKNOLEDGMENTS

The research was partially supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant.

REFERENCES

Ackermann, F., 1999. Airborne laser scanning: present status and future expectations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 64-67.

Baltsavias, E. P., 1999. A comparison between photogrammetry and laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 83-94.

Brovelli, M. A. and M. Cannata, 2004. Digital terrain model reconstruction in urban areas from airborne laser scanning data: the method and an example for Pavia (northern Italy). *Computers and Geosciences*, 30, pp. 325-331.

Haala, N. and C. Brenner, 1999. Virtual city models from laser altimeter and 2D map data, *Photogrammetric Engineering & Remote Sensing,* 65(7) , pp. 787–795.

Koch, B., U. Heyder, and H. Weinacker, 2006. Detection of Individual Tree Crowns in Airborne Lidar Data, *Photogrammetric Engineering and Remote Sensing*, 72(4) , pp. 357-364.

Lloyd, C. D. and P. M. Atkinson, 2002. Deriving DSMs from LIDAR data with Kriging. *International Journal of Remote Sensing*, 23(12), pp. 2519-2524.

Ma, R., 2005. DEM generation and building detection from LIDAR data. *Photogrammetric Engineering and Remote Sensing*, 71(7) , pp. 847-854.

Oliver, S. and Carol A. Gotway, 2005. *Statistical Methods for Spatial Data Analysis.* New York: CRC Press.

Shan, J. and A. Sampath, 2005. Urban DEM generation from raw Lidar data: A labeling algorithm and its performance, *Photogrammetric Engineering and Remote Sensing*, 71(2), pp. 217-226.

Sithole, G. and G. Vosselman, 2004. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(1-2), pp. 85-101.

Stein, M. L., 1999. Interpolation of Spatial Data*:* Some Theory for Kriging, Springer, 247 pp.

Wallace, C. S. A. and S. E. Marsh, 2005. Characterizing the spatial structure of endangered species habitat using geostatistical analysis of IKONOS imagery. *International Journal of Remote Sensing*, 26(12), pp. 2607-2629.

Wehr A. and U. Lohr, 1999. Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 68-82.

Woolard, J. W. and J. D. Colby, 2002. Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. *Geomorphology*, 48(1-3), pp. 269-287.