Computers, Environment and Urban Systems 41 (2013) 289–298

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Application of random sets to model uncertainty of road polygons extracted from airborne laser points

Liang Zhou *, Alfred Stein

University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 13 January 2012
Received in revised form 29 June 2012
Accepted 30 June 2012
Available online 25 July 2012

Keywords:
Uncertainty modelling
Road extraction
Random sets
Laser scanning
Dependency

ABSTRACT

High point densities obtained by today's laser scanning systems enable the extraction of features that are traditionally mapped by photogrammetry or land surveying. While significant progress has been made in the extraction of roads from dense point clouds, little research has been performed on modelling uncertainty in extracted road polygons. In this paper random sets are used to model this uncertainty. Based on the accuracy reported by the data provider, positional errors in laser points are simulated first by a Markov Chain Monte Carlo method. An algorithm is developed next to detect the positions of road polygons in the simulated data and integrating the random sets for the uncertainty modelling. This algorithm is adapted to point data with different densities and variable distributions. Uncertainty modelling includes modelling of the dependence between the vertices of a road polygon. Road polygons constructed from vertices with different truncated normal distributions along with their uncertain line segments are represented by random sets, and their parameters are estimated. The effect of distributions on the area of the mean set is analysed and validated by a set of reference data collected from GPS measurements and image digitising. Results show that random sets provide useful spatial information on uncertainties using their basic parameters like the core, mean and support set. The study shows that random sets are well-suited to model the uncertainty of road polygons extracted from point data.

1. Introduction

LiDAR is a terrain and urban information acquisition technique based on laser technology. The use of LiDAR data for urban modelling and visualisation has received much attention recently. Specifically, digital city modelling is benefiting from realistic visualisations. Advantages such as short data acquisition and processing times, relatively high accuracy and point density, and low costs have caused LiDAR to be preferred over traditional aerial photogrammetric techniques.

Laser scanners nowadays can acquire point clouds with densities of 20–50 pts m^{-2} from airborne platforms. These point densities enable the use of laser scanning data for various mapping tasks. Studies on the use of laser points typically focused on the applications such as DTM generation (Kraus & Pfeifer, 2001), 3D building modelling (Brenner, 2005; Oude Elberink & Vosselman, 2009; Pu & Vosselman, 2009), and change detection (Matikainen, Hyyppä, & Hyyppä, 2003). In the context of road furniture and forest inventories, algorithms for the detection of pole-like objects have been developed (Brenner, 2009; Pfeifer, Gorte, & Winterhalder, 2004; Rutzinger, Pratihast, Oude Elberink, & Vosselman, 2010). Recent research shows that objects such as traffic islands and pavements can be extracted from airborne point clouds (Zhou & Vosselman, 2012). Positional accuracy of the extracted objects is improved by fitting a sigmoid-shaped surface. Uncertainty, however, still exists in the modelled road polygons due to the limitation of designed model and various point densities and distributions in different study areas. As the shapes of road polygons are diverse in urban areas, the designed mathematical models have their limitations in precisely modelling these diverse shapes. Hence, properly modelling this uncertainty in the extracted road polygons will be interesting for data users and stimulate other applications in urban environments.

As a basic GIS operation, the area of a polygon is calculated from the coordinates of the vertices representing its boundary. Therefore uncertainty in the coordinate values leads to uncertainty in area calculation (Van Oort, Stein, Bregt, De Bruin, & Kuipers, 2005). Traditional probability theory has been used for modelling error propagation in spatial objects. Positional uncertainties are mainly caused by measurement errors (Zhang & Goodchild, 2002). Analytical approaches have been developed for modelling uncertain points (Thapa & Bossler, 1992) and lines (Shi & Liu, 2000). Models for uncertain polygons are mostly based on models for points and lines, as the uncertain location of the outline of a polygon is specified by the joint probability distribution function

* Corresponding author. Tel.: +31 53 4874 345; fax: +31 53 4874 335.
E-mail addresses: zhou19153@itc.nl (L. Zhou), stein@itc.nl (A. Stein).

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.006

of its primitive points (Zhao, Stein, & Chen, 2010). Research has been conducted to model the uncertainty of individual polygons (Bondesson, Stahl, & Holm, 1998; Chrisman & Yandell, 1988; Griffith, 1989; Liu & Tong, 2005; Magnussen, 1996; Næset, 1999; Shahin, 1997). Prisley, Gregoire and Smith (1989) derived variance and covariance equations from an area equation that required the coordinate values of each polygon's centroid. Based on the work of Chrisman and Yandell (1988), Van Oort et al., 2005 derived variance and covariance equations, which are independent of polygon centroid coordinates. In most of the above studies, assumptions were made that the uncertainty in the area of a polygon originated from the uncertainty of the vertices defining its outline. Liu and Tong (2005) developed an error model to describe the influence of line uncertainty on the uncertainty of polygon area where the effect of uncertainty in parametric curves on the uncertainty of polygon area, however, was not explicitly modelled. The main reason is that an error model for parametric curves is complicated as the error may be either in the endpoints or in the parameters (Chrisman & Yandell, 1988). Since spatial objects derived from remote sensing data may have gradual transition boundaries (Stein, Hamm, & Ye, 2009), estimating the probability density function (pdf) for each polygon vertex may be difficult. In addition, georeferencing of remote sensing data and manual digitization may introduce correlations between boundary points (Heuvelink, Brown, & Van Loon, 2007).

Alternatively, uncertain spatial objects can be modelled by random sets. Random sets were originally developed for the study of randomly varying geometrical shapes (Stoyan & Stoyan, 1994) and for image segmentation (Epifanio & Soille, 2007). Zhao et al. (2010) used random sets to model the uncertainty of wetland derived from satellite images. It quantifies extensional uncertainty of spatial objects and models the broad boundaries extracted from images (Zhao, Stein, Chen, & Zhang, 2011). Its parameterization is adapted for monitoring seasonal dynamics of wetland variation and interannual changes of wetland inundation extents (Zhao, Stein, & Chen, 2011). In this way, a natural entity with uncertainties is treated as random sets in population space. The objects derived by digitizing, thresholding or segmentation of image data were modelled as a set of observations in sample space. A statistical analysis helps to understand the characteristics of the sample.

The aim of this paper is to examine the feasibility of employing random sets to model the uncertainty of road polygons derived from airborne laser scanning data. This is realised in three steps. First, the uncertainties in airborne laser points are simulated and their effect on the uncertainties of derived polygons is explored. Second, random sets theory is used to model the uncertainties of derived polygons. Third, a statistical analysis is employed to estimate the characteristics of the derived polygons. The study is applied on a set of six road polygons from the city of Enschede, The Netherlands.

2. Data description

Airborne laser scanning data used in this study were acquired with a FLI-MAP 400 system (Fugro Aerial Mapping., 2011) with forward, nadir, and backward looking scan directions. The system consists of an airborne laser scanner, two digital cameras and two video cameras. The dataset contains 15 strips with a point density of 20 pts m^{-2} and was recorded at a flight height of 275 m above ground in Enschede, The Netherlands. The systematic errors (offsets between strips) are in the order of 4–8 cm for the X, Y coordinates and 2–3 cm for the Z coordinates. The stochastic platform positioning error is approximately 2–3 cm for X, Y and Z coordinates. Planimetric standard deviations of 2 cm have been achieved with a little additional calibration following Vosselman (2008). This is superior to the accuracy reported in the platform specification of 5 cm accuracy in both horizontal and vertical directions at the 95% confidence interval. Fig. 1 highlights the six road polygons digitised from the orthophoto, being the subject of the study. More details are provided in Sections 3 and 4.

3. Methodology

The quality of objects derived from remote sensing data depends upon properties of the input data and the processing steps. Uncertainty modelling of the output objects should include errors in the input data as well as errors caused by data processing methods. In this study the positional errors in laser points were assumed to have a bivariate normal distribution and were simulated by using Markov Chain Monte Carlo (MCMC) simulations (Besag, 2001). A previously developed method (Zhou & Vosselman, 2012) was used to detect and model the road polygons from the error-contaminated data. Every extracted road polygon consists of a sequence of vertices and the uncertainty of each vertex is represented by its uncertain line segment. Random sets were used to model the uncertainties in the extracted road polygons and results of statistical analysis were discussed. The impact of viable distances between two neighbouring vertices and their dependency on the uncertainty of derived road polygons is analysed by means of different experiments.

3.1. Simulating the positional errors in laser points

Markov Chain Monte Carlo methods consist of a class of algorithms for sampling from probability distributions. To conduct an MCMC simulation, the probability distribution of random variables should be known. Hunter et al. (1996) found that the standard deviation supplied by the data producer is useful to simulate random errors. In this research, it is assumed that the horizontal error of each LiDAR point has a bivariate normal distribution with a mean of zero and a standard deviation (SD) known from data provider. Since the vertical error of LiDAR points has less influence on the positional uncertainty of a polygon, we do not take it into account. The Metropolis–Hastings algorithm is applied on the MCMC method for obtaining a sequence of random samples from a probability distribution. As the laser points have been calibrated to a high degree of accuracy, a planimetric standard deviation of 2 cm was achieved. For each laser point, a sequence of 1000 random samples was selected by the Metropolis–Hastings algorithm from a bivariate normal distribution $X \sim \mathcal{N}(\mu, \Sigma)$, where $X = [x, y]$, $u = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{Bmatrix} 4 & 0 \\ 0 & 4 \end{Bmatrix}$ (Fig. 2).

3.2. Mapping road polygons from simulated data

In urban areas, curbstones often separate the road surface from the adjacent pavement. These curbstones are mapped using a three step procedure. First, the locations with small height jumps near the terrain surface are detected. Second, midpoints of high and low points on either side of the height jump are generated, and are put into a sequence to obtain a polygon describing the approximate curbstone location (Zhou & Vosselman, 2012). A sigmoidal function (Eq. (1)) is then fitted to the simulated points near the polygon to increase the accuracy. We use this function to describe the height Z as a function of the location x perpendicular to the road side. In this equation W is the slope parameter, x_{ip} is the inflection point, and Z_t and Z_b are the top and bottom height (Fig. 3a). Further, the longitudinal shape of the curbstone is assumed to follow a cubic polynomial $x_{ip} = a_0 + a_1 y_{ip} + a_2 y_{ip}^2 + a_3 y_{ip}^3$ where the y-direction is taken parallel to the local road side

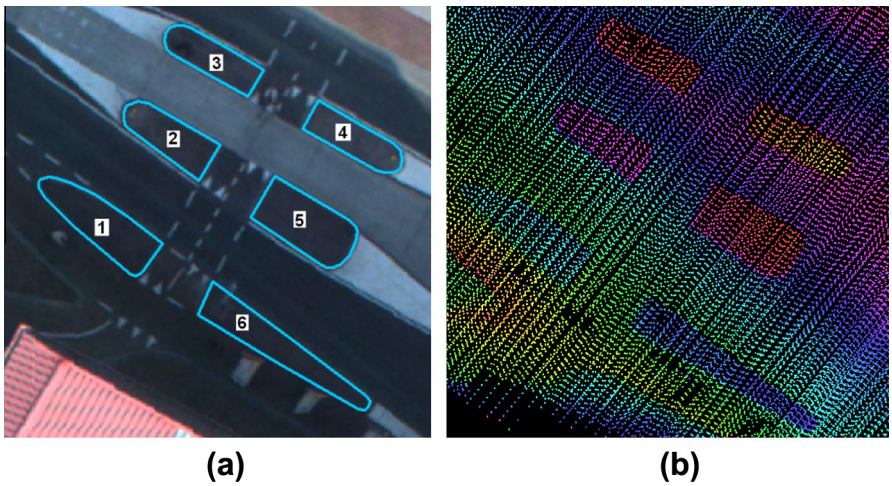

Fig. 1. Digitised road polygons (blue) from orthophoto (a), colour-coded laser points by height with cycle length of 0.3 m (b).

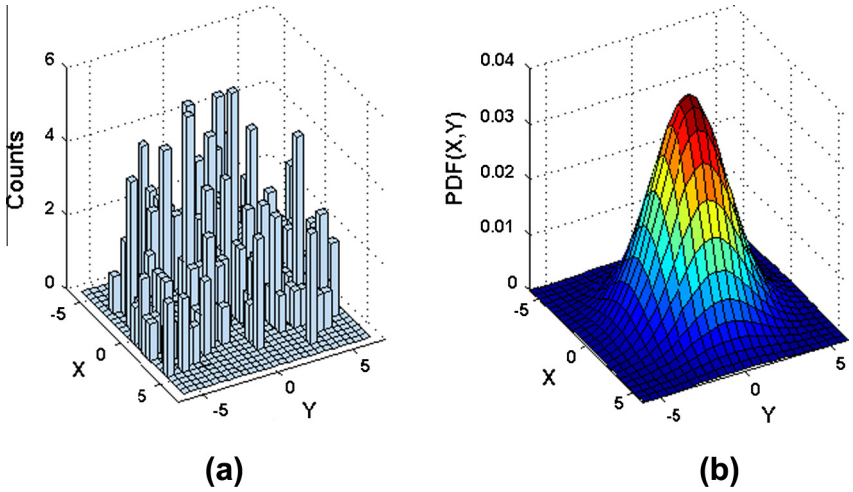

Fig. 2. Thousand random samples selected by Metropolis algorithm (a), theoretical bivariate normal distribution with mean of zero and variance of 4 cm (b).

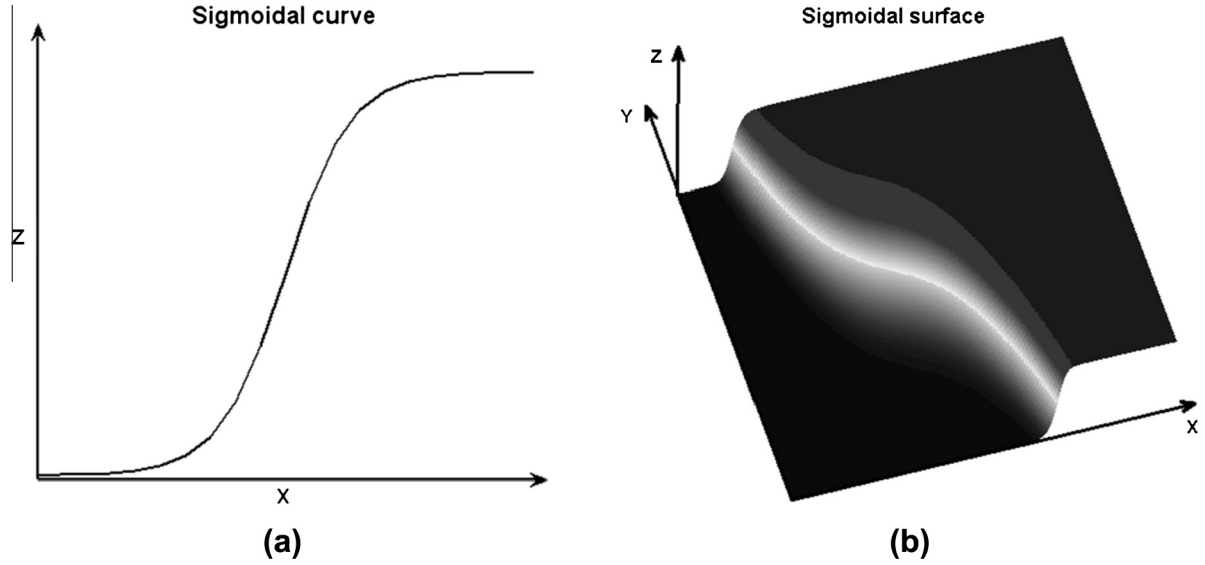

Fig. 3. Sigmoidal curve (a) and sigmoidal curve surface (b).

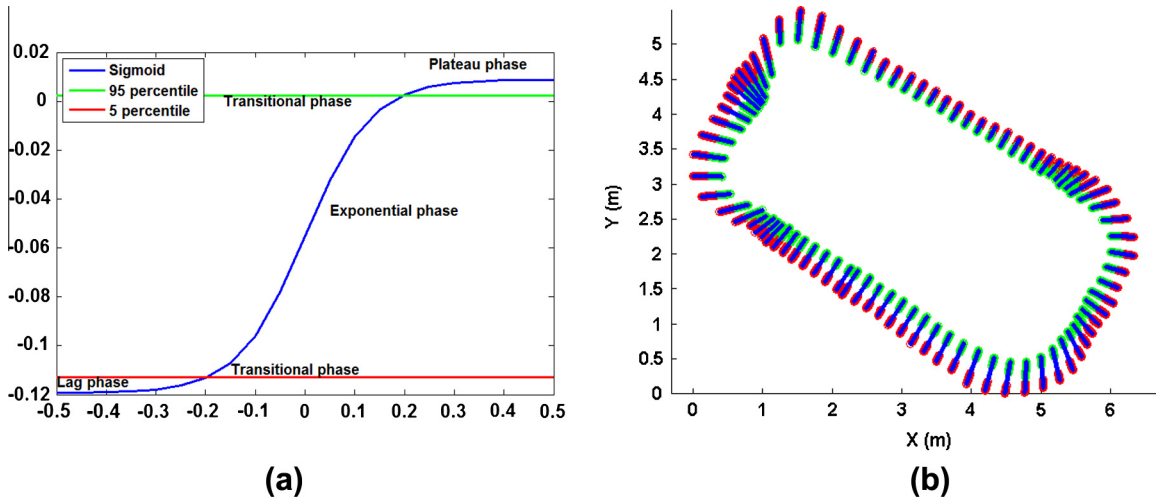

Fig. 4. Five percent (red line) and 95% (green line) height of a sigmoidal curve (a) and recorded 1000 positions of them, blue line segments are the uncertain line segments of each vertex (b).

direction. This polynomial is used to substitute x_{ip} in Eq. (1). The resulting Eq. (2) describes a 2.5D surface (Fig. 3b).

$$Z = Z_b + \frac{Z_t - Z_b}{1 + e^{w(x_{ip} - x)}} \quad (1)$$

$$Z = Z_b + \frac{Z_t - Z_b}{1 + e^{w\left(a_0 + a_1 y_{ip} + a_2 y_{ip}^2 + a_3 y_{ip}^3 - x\right)}} \quad (2)$$

The method uses polygons reconstructed by means of the detection method and simulated points near the ground surface as input data. A window is created and moved along each polygon with a small step size, thus determining the number of vertices on each polygon. A smaller step size leads to more vertices. By fitting the sigmoidal surface to the simulated laser points in the moving window, parameters of a sigmoidal curve in the centre of this window are estimated. Even though these parameters are accurately estimated by least squares, the position of a vertex is still uncertain, because it may appear between the two transitional phases in (Fig. 4a), with boundaries that are also not crisp. In our study, we assumed that the position of the vertex varies along the sigmoidal curve from 5% to 95% of its height (Fig. 4a). For each vertex, the positions at the 5% (red[1] points in Fig. 4b) and at the 95% (green points in Fig. 4b) height of a sigmoidal curve are estimated and recorded 1000 times by simulated laser points. The uncertainty of each vertex is within the line segment (blue) between the innermost and outermost end points (Fig. 4b) and this line segment is defined as the uncertain line segment of a given vertex.

Random data types can be either random point, random line or random regions (Zhao, Chen, & Stein, 2009). Random regions can be regarded as random sets and used for spatially modelling areal geographic entities, such as clouds, dunes, field patches and lakes. We consider a polygon constructed from the uncertain vertices as an element of the random set. The distribution of all the elements of a random set is described by its covering function as described below.

3.3. Constructing random sets from extracted polygons

3.3.1. Random sets

We equate the road polygons with a random set X. A random set X on the Euclidean space R^n associates a probability value to each element $x \in R^n$. It quantifies how likely it is that x belongs to X. The random set X on R^n is a function $P_x: R^n \to [0,1]$, called the covering function of the random set, taking values between 0 and 1. The set $X_\alpha = \{x \in r^n | P_x(x) \geqslant \alpha\}$ is the α-level set of X, whereas the set $X_0 = \{x \in R^n | P_x(x) > 0\}$ is the support set of X and $X_1 = \{x \in R^n | P_x(x) = 1\}$ is the core set of X. We denote the set of all random sets in R^n by $\Re R^n$. The mean X_m of the random set has been defined in several ways (Stoyan & Stoyan, 1994). Following (Zhao et al., 2009), we choose the Vorob'ev expectation because it considers sets with a finite number of points. In a 2D space $I \subset R^2$, the mean area EA of the random set X is then defined as $EA(X) = \int_{R^2} P_x(x)dx$. According to the definition of the Vorob'ev expectation, the set X_m is equal to $X_m = \{x \in R^2 | 0 \leqslant \alpha_m \leqslant 1 | P_x(x) \geqslant \alpha_m\}$ where α_m is such that X_m has the area EA(X). If α_m is not unique, then it is set equal to the infimum of all such α_ms. When $\alpha_m = 1/2$, the mean and median are identical.

3.3.2. Random sets generated by extracted polygons

The uncertainty of each vertex is represented by a line segment described in Section 3.2. In order to construct random polygons by means of uncertain vertices, the position of each vertex is sampled from its uncertain line segment (Fig. 5a). Oversampling the positions of each vertex along its uncertain line segment increases the computation costs, whereas too few samples cannot truly represent the distribution of a random set. In this experiment, each uncertain line segment has been subdivided into 100 intervals of equal length (Fig. 5b). It is assumed that the probability of a vertex appearing on its uncertain line segment has a normal distribution.

Approaches for modelling the uncertainty in spatial objects have been discussed in (Clementini, 2005; Schneider, 1996; Tøssebro & Nygård, 2008), but none has modelled the dependency between adjacent vertices of a single object. Dependence exists between adjacent vertices during different mapping tasks like digitizing and land surveying. It should be considered during the uncertainty modelling. The shape and size of an object vary and dependence may change from one mapping task to another. No explicit methods exist to model these dependencies. In this study, we assumed that the probability distribution of one vertex located along its uncertain line segment is positively correlated with the distribution of its neighbours. For example, if the probability of a vertex appearing towards the exterior of a polygon is high, the remaining vertices are likely following the same tendency. In order to model this, the probability of a vertex located on its uncertain line segment is modelled as a truncated normal distribution

[1] For interpretation of colour in Figs. 1,2,4–13, the reader is referred to the web version of this article.

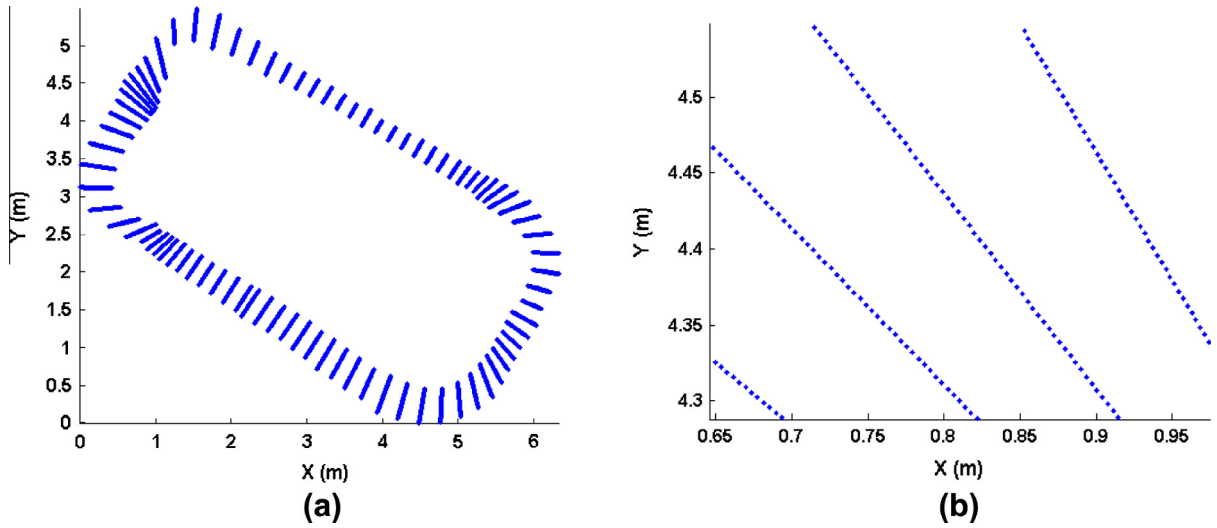

Fig. 5. Uncertain line segments (a) and equally divided uncertain line segments (b).

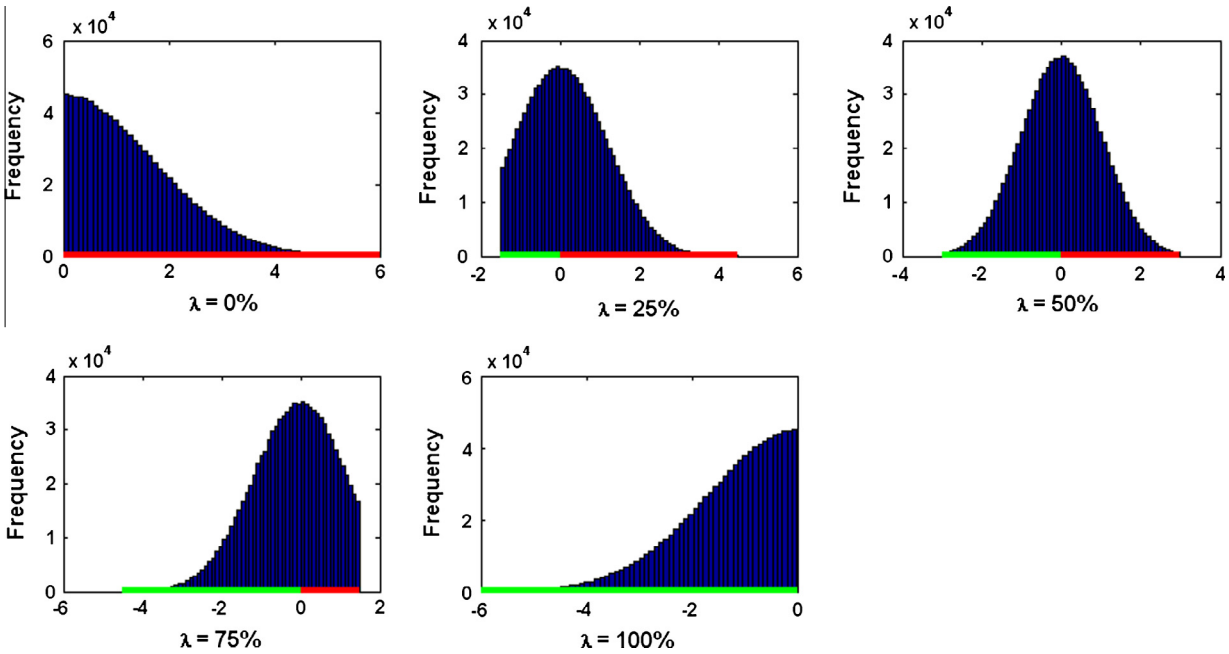

Fig. 6. Truncated normal distribution on an uncertain line segment with mean = 0 and standard deviation = 1, λ is taken as the ratio between the length on the left (green line) and right (red line) side of the mean.

depending on an assumed position of the mean. Because the uncertain line segment of each vertex has a different length, a range of ratio values (λ) between the length on left and right side of the mean is selected to define the positions of mean on each uncertain line segment and probability distribution along them will be scaled accordingly. Fig. 6 shows five assumed positions of means defined as the ratios 0%, 25%, 50%, 75% and 100%. The truncated normal distribution is simulated with mean = 0 and standard deviation = 1 and covers the uncertain line segment within a six sigma precision.

Five groups of 1000 polygons are constructed according to the five different truncated normal distributions. Each group is treated as a random set and the X_1, X_0 and X_m are calculated by a covering function. For each random set, the arithmetic mean of the polygon area is calculated. Then a covering function is built for each sampled vertex from its uncertain line segment. The core X_1 is made up by those vertices, which are covered by all random polygons (pink line in Fig. 7). The support X_0 is the largest random polygon that contains all random polygons (red line in Fig. 7). X_m is the Vorob'ev mean of a random set, which area is close or equal to the arithmetic mean of the random sets (green line in Fig. 7).

3.4. Validation

Two types of reference data were collected for polygon 5 to validate the modelling results. From a RTK-GPS field campaign in total eight GPS points are selected along the boundary of traffic island 5 (Fig. 1). The positions of these GPS points are preferred at the distinct points of this traffic island, which can largely represent its shape. The GPS points have a standard deviation below 3 cm. This accuracy was confirmed through re-visiting some points after several hours (Gerke, 2011). These GPS points are overlaid with orthophotos and used as a guide to digitise the complete outline

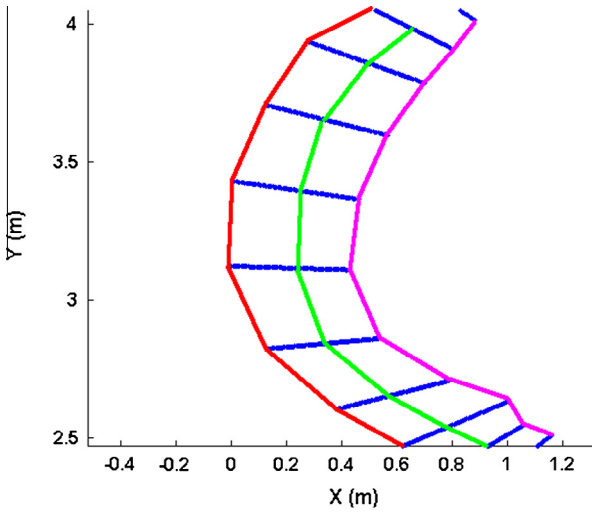

Fig. 7. The calculation of core (pink line), support (red line) and mean set (green line) for ratio of 50% (normal distribution).

of this traffic island. The orthophotos were produced from the aerial photos collected alongside the LiDAR data with a spatial resolution of 10 cm. The quality of the direct georeferencing was reported as the total system accuracy with a standard deviation of 5 cm. As the geographical accuracies of aerial photos were correlated to the quality of the LiDAR data, the produced orthophotos cannot be considered as a truly independent validation source. The areas of the digitised road polygons, however, can be used to evaluate the area variations of derived road polygons as they are not subject to the quality of the georeferencing. Good quality of registration between the GPS points and orthophotos can be observed in Fig. 8. These reference data presented in Fig. 8 are used as ground truth and compared with the output from uncertainty modelling.

4. Results

4.1. Effect of step size on the variation of polygons

To investigate the relationship between the number of vertices and the calculated polygon area, different step sizes for the moving window are used in the experiment. One traffic island (Fig. 8) was selected to analyse the relationship.

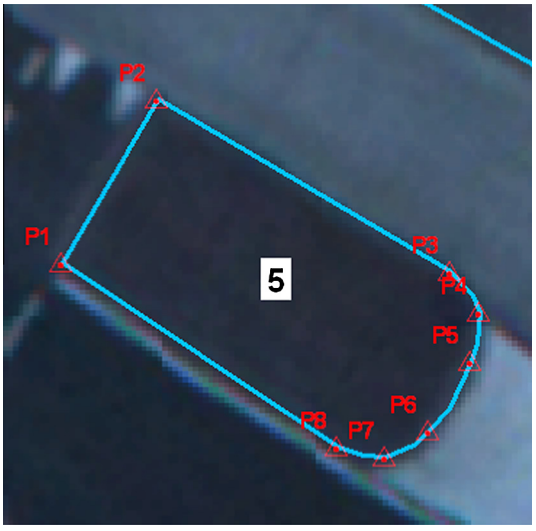

Fig. 8. Reference data in orthophoto, red points are collected by GPS, blue polygon is manually digitised on orthophoto by using GPS points as guide.

Table 1
Relationship between polygon area and step size.

Step size (cm)	X_0 (m^2)	X_m (m^2)	X_1 (m^2)	σ (m^2)	Vertices (n)
10	19.66	16.64	13.88	0.05	165
15	19.63	16.62	13.88	0.05	110
20	19.58	16.58	13.84	0.06	83
25	19.63	16.61	13.84	0.07	67
30	19.54	16.52	13.83	0.07	56
35	19.51	16.51	13.72	0.08	48
40	19.47	16.51	13.71	0.09	42

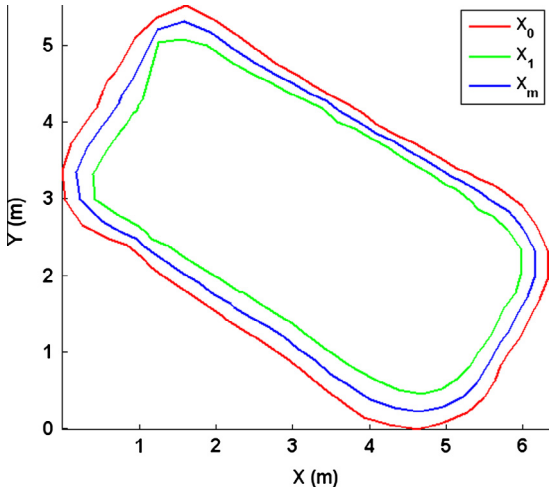

Fig. 9. Mean, core and support set of random sets with step size of 25 cm.

The step size of moving widow shown in Table 1 ranges from 10 cm to 40 cm. Random sets are built up regarding the differences of step size and the sets X_0, X_m and X_1 are calculated accordingly (Fig. 9). The dependency between adjacent vertices has a normal distribution along its uncertain line segment in this test. The general trend can be observed that the areas of X_1, X_0 and X_m increase slightly with the decrease of step size. If there are more vertices on a polygon, its boundary becomes smoother and the areas tend to become stable, as shown by the standard deviation (σ).

4.2. Impact of dependencies of vertices on the variation of polygons

Next, dependencies of vertices on each polygon are simulated. Five λ values are used to simulate the distribution of each vertex on its uncertain line segment. As observed from Table 1, the area of X_m becomes stable if the step size is less than 25 cm. Hence, the polygons with step size of 15 cm are taken as an example and random sets are generated by the polygons with different dependencies on their vertices.

Table 2
Different random sets generated by polygons with different dependencies on their vertices.

Dependency λ (%)	X_0 (m^2)	X_m (m^2)	X_1 (m^2)	σ (m^2)
0	19.66	18.34	13.88	0.08
25	19.66	17.92	13.88	0.08
50	19.66	16.66	13.88	0.08
75	19.66	15.47	13.88	0.08
100	19.66	15.07	13.88	0.08

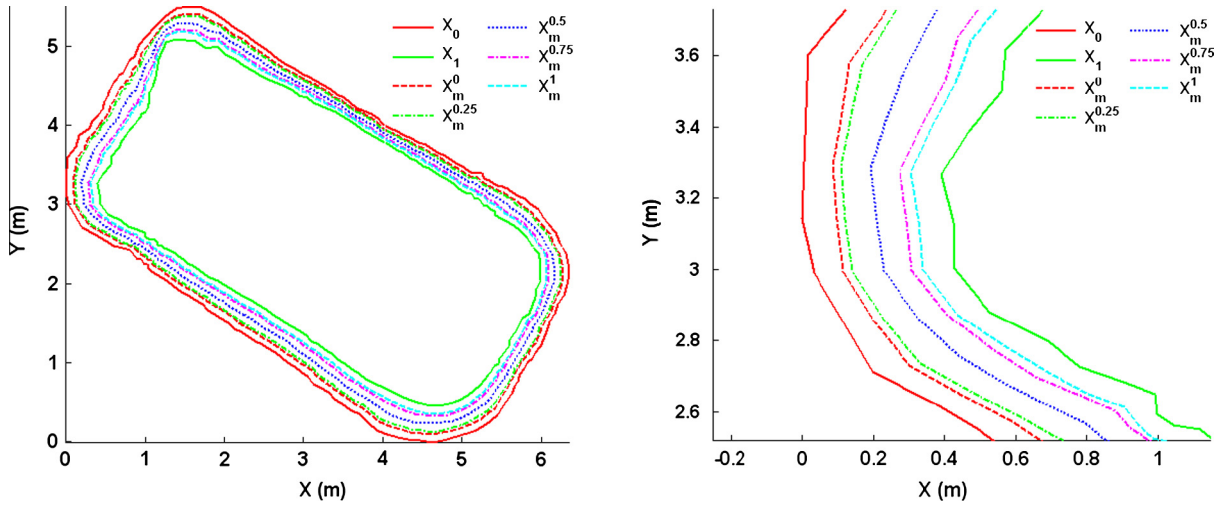

Fig. 10. Mean sets of random sets with different distributions on adjacent vertices.

Table 2 shows that the areas of X_1 and X_0 remain the same by increasing λ. The reason is that they are constructed by the innermost and outermost points on the uncertain line segment of each vertex. The changes of the distribution have a large impact on the area of X_m. Fig. 10 shows the position of mean on uncertain line segment has linear relation with the area of the mean set, if the mean is closer to the interior of the polygon (with higher λ value) the area of the mean set is smaller.

4.3. Applying algorithm to more polygons

The developed algorithm is further applied to polygons 2, 3 and 4 in Fig. 1. Fig. 11 shows that polygons 2 and 3 have an irregular shape in comparison with polygons 4 and 5 and contain sharp corners in the south-eastern parts. Such irregular shapes are caused by the variation of the point density and distribution, making uncertainty modelling more difficult. Even though modern laser scanners can acquire point clouds with high point density, due to the different perspective and incidence angle of the laser beam the point spacing is not guaranteed to be regular. The advantage of this algorithm is that the mapping algorithm is not constrained to the point density and distribution in the local area. The size of the moving window is adjusted by the point density, allowing to carry out uncertainty modelling in an area with variable point density. The internal and external boundaries given by X_1 and X_0 respectively are more irregular than X_m for polygons 2 and 3. So if X_m is considered as the true representation of a polygon then the result is acceptable, because it is calculated from the covering function, in which the shapes of X_1 and X_0 have only a little effect.

Table 3 shows that with the decrease of λ the area of X_m decreases for each polygon. Table 3 suggests that the area of X_m shows a linear relation with λ. If we refer to Fig. 12, however, the relation between these two variables is not linear. The general trend is that the area of X_m significantly changes from 25% to 75% of λ, whereas for the remaining intervals of λ the area of X_m varies slightly. The maximum area variation is 0.42 m^2.

4.4. Comparison between modelling outputs and references

The area of polygon 5 as digitised from orthophoto equals 16.74 m^2 which is higher than the area of each X_m in Table 1. By increasing the number of vertices, the area of mean set is approaching the reference but still less than it. This indicates that the distribution of a vertex along its uncertain line segment may not be normal. But if we compare the reference with the area of the mean set in Table 2, it suggests that the reference lies between the mean sets calculated by the vertices with λ ranging from 0.25 to 0.5 and close to 0.5. Fig. 13 shows the GPS points together with the output mean sets with different λ values. All P3–8 fall within both sides of the X_m with $\lambda = 0.5$, whereas P7 is the closest to the X_m with $\lambda = 0.25$. P1 and P2 are close to the support set, because they are on the boundary of the polygon where the local shape is right angular (Fig. 13). Such a shape cannot be properly modelled by the designed mathematical model, which assumes longitudinal shape of a polygon is a 3rd order polynomial. As a result the sharp corner on a polygon is rounded, leading to problems in the uncertainty modelling step.

5. Discussion

In this paper, we presented random sets method to model the uncertainty of objects derived from airborne laser points. Due to the nature of laser scanning, measurements are rarely acquired at the distinct positions of an object, such as its outlines, corners or edges. We show how these objects can be detected and modelled by means of developed road detection algorithms and mathematical models respectively, while uncertainty in the road polygons is analysed by random sets. The advantage of the road detection algorithm is that it is not limited to the point density and distribution, which have a large effect on the uncertainty in extracted polygons. By using the statistical parameters (X_0, X_m and X_1) of random sets, we demonstrate that random errors in the laser points and randomness of detection parameters have different effects on extracted features when objects have different extensional uncertainties.

Current methods for detecting and modelling the traffic polygons can use variable step sizes for the moving window, resulting in different vertices on the polygons. Results suggest that the area of the polygon tends to be stable when a smaller step size is selected. Selection of a small step size, however, may increase computation costs while confronting large objects. Furthermore, to make the detecting and modelling generic, it is assumed that the boundary of a traffic island can be modelled by a 3rd order polynomial. This may not hold for all situations. For example, the study polygon in Fig. 8 has actually two right angular corners, which cannot be properly modelled by such a polynomial. In order to get a more accurate boundary around these areas, either a more specific model needs to be considered or a data-driven approach without

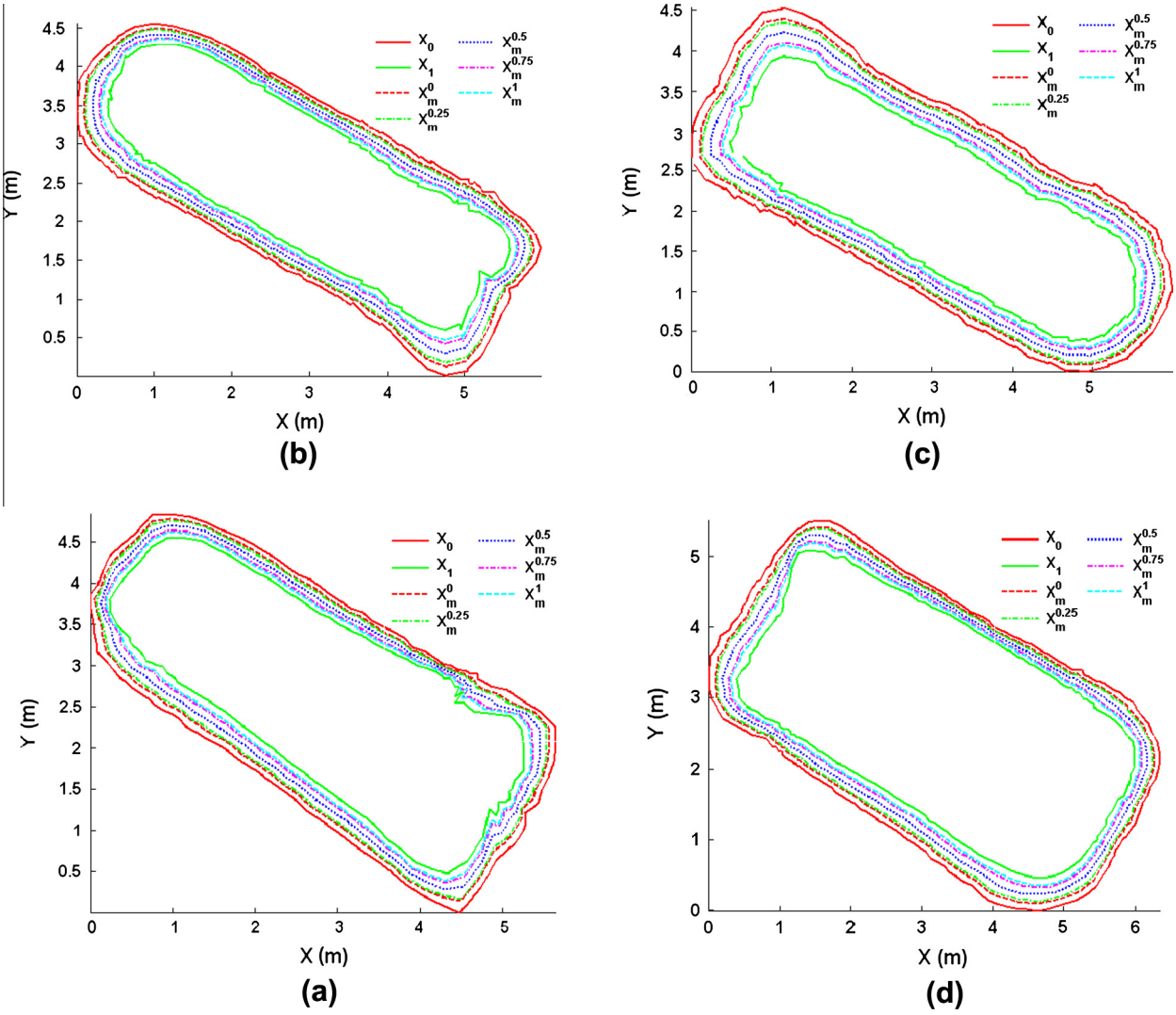

Fig. 11. Mean sets of random sets with different distributions for polygon 2 (a), 3 (b), 4 (c) and 5 (d).

Table 3
Relationship between mean set area and λ.

Polygon	X_1 (m^2)	$\lambda = 1$	$\lambda = 0.75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.00$	X_0 (m^2)
2	9.15	10.06	10.35	11.27	12.25	12.57	13.61
3	8.17	9.12	9.44	10.40	11.44	11.75	12.82
4	7.67	8.83	9.20	10.38	11.63	12.05	13.38
5	13.88	15.07	15.47	16.66	17.92	18.34	19.66

any geometrical model involved could be an optimal solution in further study.

Random set theory is selected in this study to model uncertainty. The empirical covering function for positions in random sets can also be regarded as a fuzzy membership value, thus indicating the possibility of covering a position by the object. Furthermore, α-cuts as an important concept of a fuzzy set, used as an efficient interpretation tool to describe the internal structure of a fuzzy boundary. It can be regarded as a nested random set in our approach. Therefore, random set can serve as an alternative approach to simultaneously obtain fuzzy membership functions and α-cuts.

Dependence between adjacent vertices varies due to the size and shape of an object and also changes from one mapping task to another. There is no explicit way to combine the dependence in uncertainty modelling. Current work assumed that each vertex has a truncated normal distribution along its uncertain line segment, which is subject to current detecting and modelling method.

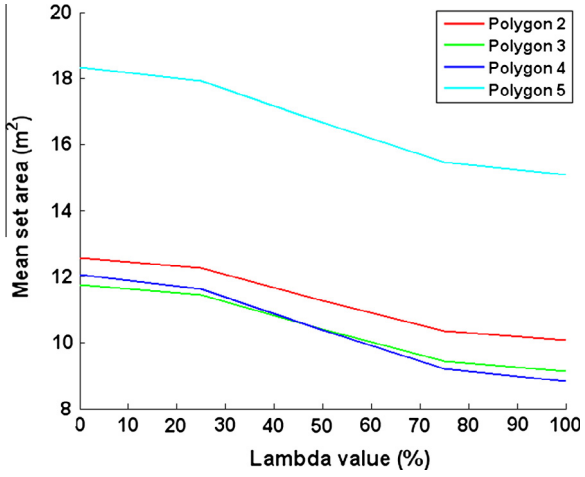

Fig. 12. 2D plot mean set area against λ.

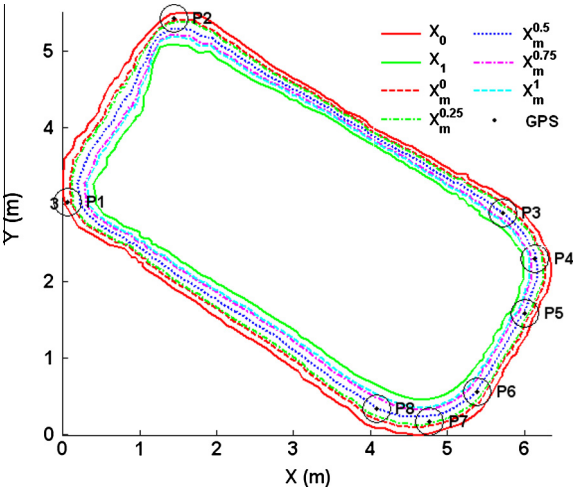

Fig. 13. GPS points overlaid with output polygons from uncertainty modelling step.

This uncertain line segment has been subdivided in 100 intervals of equal length. As the maximum length of all uncertain line segments in this study are less than 1 m, the sample distance is then less than 1 cm, which is sufficient for any current mapping purposes. Another assumption can be made when the detecting strategy changes. By simulating the distribution of a vertex along its uncertain line segment, the effect of the distribution on the uncertainty of a road polygon is properly modelled.

In dense urban areas, LiDAR can serve as a promising technology for data acquisition. As the data quality is essential for urban applications, uncertainty assessment of raw LiDAR data and the results of reconstruction are important. The uncertainty in raw LiDAR data arises both in planimetric as well as height features. We are currently investigating methods for modelling planimetric errors using the simulation method. Future work may focus on the combined effect of planimetric and height errors on the uncertainty of object reconstruction.

6. Conclusions

Random errors in LiDAR data can be properly modelled by simulation methods. By applying the developed road detection procedure to simulated data, the derived road polygons contain the uncertainty in the point data itself and also those obtained during data processing. Each road polygon is then treated as an element of a random set, which can spatially model objects with uncertain boundaries.

Random sets applied to road polygons are useful for modelling the uncertainty, which is not clearly addressed in either traditional analytical methods or other popular methods. A statistical analysis of random sets is encouraging for determining and characterising the variation of the uncertainty in detected road polygons.

In traditional methods the correlation between the adjacent vertices of an object is not explicitly modelled or taken into account. This study shows that simulating the correlation between the vertices and experimenting with different possibilities the model's sensitivity is clearly determined.

The reference data collected from fieldwork can assist in analysing the characteristics of uncertainty modelling outputs. These results show that random sets are well-suited to model the uncertainty of road polygons extracted from point data.

References

Besag, J. (2001). Markov chain Monte Carlo for statistical inference. *Technical Report, University of Washington, Seattle, USA.*

Bondesson, L., Stahl, G., & Holm, S. (1998). Standard errors of area estimates obtained by traversing and GPS. *Forest Science, 44*(3), 405–413.

Brenner, C. (2005). Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation, 6*(3–4), 187–198.

Brenner, C. (2009). *Extraction of features from mobile laser scanning data for future driver assistance systems. Advances in GIScience.* Berlin Heidelberg: Springer.

Chrisman, N. R., & Yandell, B. S. (1988). Effects of point error on area calculations: A statistical model. *Surveying and Mapping, 48*(4), 241–246.

Clementini, E. (2005). A model for uncertain lines. *Journal of Visual Languages & Computing, 16*(4), 271–288.

Epifanio, I., & Soille, P. (2007). Morphological texture features for unsupervised and supervised segmentations of natural landscapes. *IEEE Transactions on Geoscience and Remote Sensing, 45*(4), 1074–1083.

Fugro Aerial Mapping. (2011). FLI-MAP 400 Retrieved http://www.fugrowaterservices.com/downloads/aerial-mapping/FLIMAP400system_LR.pdf. (Accessed 25.10.11)

Gerke, M. (2011). Using horizontal and vertical building structure to constrain indirect sensor orientation. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 307–316.

Griffith, D. A. (1989). *Distance calculations and errors in geographic databases. Accuracy of spatial databases*. London: Taylor and Francis, pp. 81–90.

Heuvelink, G., Brown, J., & Van Loon, E. (2007). A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science, 21*(5), 497–513.

Hunter, G., Hock, B., Robey, M., & Goodchild, M. (1996). *Experimental development of a model of vector data uncertainty*. In Paper presented at the spatial accuracy assessment in natural resources and environmental science: Second international symposium, Fort Collins, Colorado.

Kraus, K., & Pfeifer, N. (2001). Advanced DTM generation from LIDAR data. *International Archives of Photogrammetry, Remote Sensing and Spatial, Information Sciences, 34*(Part 3/W4), 23–30.

Liu, C., & Tong, X. H. (2005). Relationship of uncertainty between polygon segment and line segment for spatial data in GIS. *Geo-Spatial Information Science, 8*(3), 183–188.

Magnussen, S. (1996). A coordinate-free area variance estimator for forest stands with a fuzzy outline. *Forest Science, 42*(1), 76–85.

Matikainen, L., Hyyppä, J., & Hyyppä, H. (2003). Automatic detection of buildings from laser scanner data for map updating. *International Archives of Photogrammetry, Remote Sensing and Spatial, Information Sciences, 34*(Part 3/W13), 218–224.

Næset, E. (1999). Effects of delineation errors in forest stand boundaries on estimated area and timber volumes. *Scandinavian Journal of Forest Research, 14*(6), 558–566.

Oude Elberink, S., & Vosselman, G. (2009). Building reconstruction by target based graph matching on incomplete laser data: Analysis and limitations. *Sensors, 9*(8), 6101–6118.

Pfeifer, N., Gorte, B., & Winterhalder, D. (2004). Automatic reconstruction of single trees from terrestrial laser scanner data. *International Archives of the Photogrammetry, Remote Sensing and Spatial, Information Sciences, 35*(Part B5), 111–114.

Prisley, S., Gregoire, T., & Smith, J. (1989). The mean and variance of area estimates computed in an arc-node geographic information system. *Photogrammetric Engineering and Remote Sensing, 55*(11), 1601–1612.

Pu, S., & Vosselman, G. (2009). Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(6), 575–584.

Rutzinger, M., Pratihast, A. K., Oude Elberink, S. J., & Vosselman, G. (2010). Detection and modelling of 3D trees from mobile laser scanning data. *International Archives of the Photogrammetry, Remote sensing and Spatial, Information Sciences, 38*(Part 5), 520–525.

Schneider, M. (1996). Modelling spatial objects with undetermined boundaries using the Realm/ROSE approach. *Geographic Objects with Indeterminate Boundaries, 2*, 141–152.

Shahin, F. (1997). Modeling positional uncertainty of linear features in geographic information systems. *An-Najah University Journal for Research, 11*, 23–38.

Shi, W., & Liu, W. (2000). A stochastic process-based model for the positional error of line segments in GIS. *International Journal of Geographical Information Science, 14*(1), 51–66.

Stein, A., Hamm, N., & Ye, Q. (2009). Handling uncertainties in image mining for remote sensing studies. *International Journal of Remote Sensing, 30*(20), 5365–5382.

Stoyan, D., & Stoyan, H. (1994). *Fractals, random shapes, and point fields: methods of geometrical statistics*. Chichester, UK: Wiley.

Thapa, K., & Bossler, J. (1992). Accuracy of spatial data used in geographic information systems. *Photogrammetric Engineering and Remote Sensing, 58*(6), 835–841.

Tøssebro, E., & Nygård, M. (2008). Representing uncertainty in spatial databases. In *The 2008 high performance computing & simulation conference*, Nicosia, Cyprus, pp. 141–152.

Van Oort, P., Stein, A., Bregt, A., De Bruin, S., & Kuipers, J. (2005). A variance and covariance equation for area estimates with a geographic information system. *Forest Science, 51*(4), 347–356.

Vosselman, G. (2008). Analysis of planimetric accuracy of airborne laser scanning surveys. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*(Part B3a), 99–104.

Zhang, J., & Goodchild, M. F. (2002). *Uncertainty in geographical information*. London: Taylor & Francis.

Zhao, X., Chen, X., & Stein, A. (2009). *A random sets model for spatial objects with uncertain boundaries*. In Paper presented at the 12th AGILE international conference on geographic information science, Leibniz University Hannover, Germany.

Zhao, X., Stein, A., & Chen, X. (2010). Application of random sets to model uncertainties of natural entities extracted from remote sensing images. *Stochastic Environmental Research and Risk Assessment, 24*(5), 713–723.

Zhao, X., Stein, A., Chen, X., & Zhang, X. (2011). Quantification of extensional uncertainty of segmented image objects by random sets. *IEEE Transactions on Geoscience and Remote Sensing, 49*(7), 2548–2557.

Zhao, X., Stein, A., & Chen, X. L. (2011). Monitoring the dynamics of wetland inundation by random sets on multi-temporal images. *Remote Sensing of Environment, 115*(9), 2390–2401.

Zhou, L., & Vosselman, G. (2012). Mapping curbstones in airborne and mobile laser scanning data. *International Journal of Applied Earth Observation and Geoinformation, 18*, 293–304. http://dx.doi.org/10.1016/j.jag.2012.01.024.