

QUANTIFYING UNCERTAINTY IN CLASSIFIED POINT CLOUD DATA FOR GEOSPATIAL APPLICATIONS

Sumit Sen^{1,2}, Nozer Turel²

¹ GISE Lab, Department of Computer Science, IIT Bombay, Mumbai 400076 India

² Genesys International, 73-A SDF III, SEEPZ, Mumbai 400096 India sumitssen@gmail.com

KEY WORDS: Uncertainty, Classified Point Cloud, LIDAR, Segmentation, Classification.

ABSTRACT:

Classified Point Cloud data are increasingly the form of geospatial data that are used in engineering applications, smart digital twins and geospatial data infrastructure around the globe. Characterized by high positional accuracy such dense 3D datasets are often rated very highly for accuracy and reliability. However such data pose important challenges in semantic segmentation, especially in the context of Machine Learning (ML) techniques and the training data employed to provide classification codes to every point in massive point cloud datasets. These challenges are particularly significant since ML based processing of data is almost unavoidable due to the massive nature of the data that. We review different sources of uncertainty introduced by ML based classification and segmentation and outline concepts of uncertainty that is inherent in such automatically processed data. We also provide a theoretical framework for quantification of such uncertainty and argue that the standards of accuracy of such data should account for errors and omissions during auto segmentation and classification in addition to positional accuracy measures. Interestingly, the ability to quantify accuracies of ML based automation for processing such data is limited by the volume and velocity of such data.

1. INTRODUCTION

1.1 Background

Laser scanning and lidar systems have evolved as efficient techniques for capturing spatial data in a fast, efficient and highly reproducible way. It has been widely used in many fields, such as cultural heritage documentation, reverse engineering, three-dimensional (3-D) object reconstruction and digital elevation model (DEM) generation, as it can directly obtain the 3-D coordinates of objects. Classified LIDAR data has ushered a new generation of geospatial data that arguably has high locational accuracy. Highly dense point clouds therefore constitute much of the new geospatial data acquisition and are seen as the key to successful engineering projects such as corridor development or development of digital twins (Shirowzhan et al 2020). The ASPRS has helped develop several standards related to classified lidar data and in particular, the LAS file format that includes a point data record format (ASPRS, 2013). A key feature of this record is the use of a classification code that is based on table 4.9 of this standard and can have values from 0-255 which can represent classes of entities from 'Created, never classified', 'Unclassified', 'Ground' 'Building' to user defined classes. Whereas point class data can adhere to high horizontal accuracy standards as well as vertical accuracy based on the revised ASPRS standards for positional accuracy, the classification code assigned to each point data record may need attention in terms of errors of classification. Furthermore, the classification of different segments of a point cloud into multiple homogeneous regions, becomes a more complex affair when multiple points are mistakenly segmented into a region. The bigger picture of such errors in the context of a data that is held to very high standards of positional accuracy and is widely seen to be satisfying engineering grade geospatial data requirements. Furthermore, such errors are to be seen in the context of different 'automatic', 'semi-automatic' and manual approaches to classification and segmentation of lidar point cloud data.

Laser scanning techniques can be broadly classified into three categories, namely, airborne laser scanning (ALS), terrestrial

laser scanning (TLS) and mobile laser scanning (MLS) or mobile lidar. Thus, there is wide heterogeneity in how lidar data is collected, processed, and the perspective each data acquisition may have.

1.2 Uncertainty, the Achilles heel of GIS

Goodchild in his seminal article of 1998 (Goodchild, 1998) discussed how the notion that spatial data could be treated by applying classical theories of measurement error, was simply too limiting. His article provided a framework to understand the issues of uncertainty in spatial data beyond positional accuracy including

- Errors: These relate to blunders, misinterpretations, misclassifications and a host of other possibilities. These range from use of wrong class codes to mixing different user defined classes in Lidar files.
- Scale problems: When local details are removed to achieve 'generalization' similar to what is done when a denser point cloud is reduced in size by reducing the point density.
- Fuzziness: Such uncertainties stem from classes and features being incompletely defined. These are very common in point cloud datasets where a large number of points are classified as 'Above Ground' or 'vegetation'
- Sampling: Sampling related uncertainties arise from missing data points which is often the case in MLS as well as ALS when features behind or underneath are not represented in the data thereby creating an uncertainty about the data points that have not been represented. Very often with point cloud datasets, partially captured features have to be discarded to ensure better visualization of the dataset.

1.3 Motivation

Maps are representations of reality (Monmonier,1977) and the era of digital twins Geospatial Augmented Reality is being used to power smart city environments (Shirowzhan et al 2020). It is also important to note the massive investment in laser scanning technologies also provide a sense of higher sophistication which lends to an assumed sense of higher reliability of the dataset. It is however to be noted that positional accuracy is only one part of the overall accuracy. This paper argues that the classification and segmentation processes used to process the point clouds account for a significant portion of the overall uncertainty and hence pushes the uncertainty attribute of such data. Given the massive number of point clouds being generated, it is imperative that the precision and recall of the Machine Learning based techniques used for both these processes are performed under significant productivity stress (Grilli et al 2017). Naturally, the false positives and false negatives of such processes aggravate the overall quality of the final product. We examine this problem both theoretically and using an example from the power utility sector.

The remaining of the paper is structured as follows. In the next section we outline the processes of segmentation and classification of lidar point clouds with reference to the challenges of maintaining high accuracy. We also discuss such challenges in terms of the training process and availability of training datasets. In the third section we discuss a theoretical framework to account for total accuracy of the data generation process that takes into consideration the current ISPRS guidelines (ASPRS, 2013) and the classification process. Finally we state our conclusions based on our observations and list some suggestions for future work in this area.

2. UNCERTAINTY IN CLASSIFICATION AND SEGMENTATION

For successful exploitation of point clouds and to better understand them, it is necessary to segment and then classify such data. The former refers to group points in subsets (normally called segments) characterized by having one or more characteristics in common (geometric, radiometric, etc.) whereas classification means the definition and assignment of points to specific classes (“labels”) according to different criteria.

2.1 Accuracy of the Segmentation process

Segmentation is the process of grouping point clouds into multiple homogeneous regions with similar properties whereas classification is the step that labels these regions. Grilli et al (2017) have discussed the main categories of such segmentation into

- (a) **Edge-based** Segmentation that relies on detection of borders or edges and grouping of points inside such edges
- (b) **Region growing** segmentation which uses bottom-up or top-down approaches that first identifies the seed point and the growth of the segment based on color, geometrical criteria besides others.
- (c) **Model fitting** based segmentation which groups points that conform to the mathematical representation of the primitive shape into a segment.
- (d) **Hybrid techniques** that use a mix of above techniques.
- (e) Beyond these techniques, there are **Machine Learning** based techniques that employ various clustering algorithms such as K-means and hierarchical clustering that help create the segments. The primary objective of the ML component is to

cluster points based on attributes and features. This can be achieved by minimizing the sum of squares of distances between a given point and a cluster centroid; or by creating a hierarchical decomposition of a dataset by iteratively splitting the point cloud dataset into smaller subsets based on geometrical and radiometric characteristics until each subset consists of only one object. Lu et al. (2016) presented such a hierarchical clustering algorithm which clusters any dimensional data and can be applied to mobile mapping, aerial and terrestrial point clouds.

Like the process of segmentation, there are no standard ways of evaluating the results of segmentation. Generally speaking, the uncertainty of the segmentation process can be quantified in terms of the success of segmenting all points in the cloud. A non-exhaustive list of metrics that have been used previously are listed in table 1. below.

Measure	Description
Segment-wise uncertainty	These include measure of Covariation or pairwise distances per segment identified and used extensively in medical image processing (Roy et al 2018)
Mean Voxel/Pixel wise uncertainty	Uncertainty is computed based on the number of Voxel/Pixel values that are wrongly segmented (Roy et al 2018)
Area Fit Index (AFI)	Utilizes the difference between the areas of the largest segment identified and the actual object overlapping with the segment. (Lucieer, 2004)
Average distance between a segment boundary pixel and the reference boundary	Is a distance metric that utilizes the distance of reference boundaries and the segment boundaries (Lucieer, 2004)
Closest Distance Metric (CDM)	A cost function based on similarity between boundary images (Prieto and Allen, 2003)
Percent of area lost/gained in the segments	Compares the segments to ground truth and calculates a normalized value of pixel lost/gained by segmentation (Marpu et al, 2010)

Table 1. Uncertainty measures for segmentation processes

2.2 Machine Learning process in classification of lidar point clouds

Once a point cloud has been segmented, each segment (group) of points can be labelled with a class thus to give some semantic to the segment (hence point cloud classification is often called semantic segmentation or point labelling). In the past, semantic segmentation of point clouds has mostly been investigated for laser scanner data captured from airplanes, mobile mapping systems, and autonomous robots. Some of the earliest work on point cloud classification dealt with airborne LiDAR data, with a focus on separating buildings and trees from the ground surface, and on reconstructing the buildings. Often the point cloud is converted to a regular raster heightfield, in order to apply well-known image processing algorithms like edge and texture filters for semantic segmentation (Hug and Wehr, 1997), usually in combination with maximum likelihood classification (Maas, 1999) or iterative bottom-up classification rules (Rottensteiner and Briese, 2002).

For many applications, point-cloud classification is a basic step in LiDAR processing sometimes separated from the segmentation step. Because of the complex combination of artificial and natural objects in cities, the automated classification of 3D point clouds can be a very challenging task, especially in urban areas. Broadly speaking there are three important approaches to classification of point clouds, viz, data-driven, model-driven, and hybrid (both data & model) driven. Data-driven methods, typically, use a bottom-up approach that begins with the extraction of primitives, e.g., planes, cylinders, cones, spheres or tori, followed by analyzing primitive topology in 2D or 3D space. The geometric elements of the primitives, such as lines and critical vertices of the structures, are extracted and grouped to form models. In contrast to data-driven methods, model-driven approaches involve a top-down strategy that usually begins with a hypothetical model library and then uses the point clouds to search for optimal solutions of model composition from the model library. Hybrid approaches adapt both approaches to achieve a high success rate for the classification process. Since most of the initial studies transformed the LiDAR data point cloud into image data, traditional supervised pixel-based classification techniques such as neural networks. The object-based approach relies on a user-defined hierarchical structure to classify the segmented objects, and the technique has proven to outperform the traditional pixel-based classifiers especially on airborne LiDAR data (Chen et al., 2009; El-Ashmawy, Shaker, & Yan, 2011; Minh and Hien, 2011; Sasaki et al., 2012). Various studies reported that an overall accuracy of over 80% can be achieved using object-based technique on LiDAR-derived surfaces (Yan et al, 2015). Classification algorithms from point clouds can be grouped into automatic and interactive based on the level of user interaction required. Furthermore, techniques that work on different sources of data such as (Airborne Laser scanning) considerably differ from those developed for MLS (Mobile Laser Scanning) and TLS (Terrestrial Laser Scanning), are not mutually inclusive (Yadav et al, 2018).

Techniques of classification such as Random Forests, Decision Trees (DT), Support Vector Machines (SVMs) and K-Nearest Neighbours (K-NN) have all been used for such lidar data classification tasks (Yan et al, 2015). More recently, Deep Learning (DL) and Convolutional Neural Networks (CNN) as well Reinforcement Learning (RL) have all been attempted and have shown improved results (Griffiths and Boehm, 2019). A summary of techniques of classification employed with point clouds is shown in table 2. We indicate levels of epistemic uncertainty associated with each technique as well.

It is important to note that all machine learning tasks are based on the premise of learning from a training dataset. In case of classification of lidar point clouds, such datasets are important source of Aleatoric uncertainty based on

1. Completeness (and adequacy) of the training data
2. Correctness of the training data (noise and errors)

While some of these are unavoidable uncertainties there has been much work done in the quantification of uncertainties in Machine Learning processes and categorization of what is avoidable and what is reducible. We shall examine these aspects in the next subsection.

Classification Technique	Description	Level of epistemic uncertainty
Random Forest	Used both in classification and clustering is based on collection of Decision trees	Low
Decision Trees	Used both in classification and clustering	Moderate
SVM	Support Vector Machines can be used both in classification and clustering	Moderate
K-NN	K Nearest Neighbour helps in both classification and clustering	Moderate
CNN	The commonly used platform for Deep Learning	Low-very Low
Deep RL	Recently developed platform that uses	Low-very Low

Table 2. Types of ML algorithms used for processing point cloud data and relative epistemic uncertainty reported. (Maxwell et al 2018; Griffiths, and Boehm, 2019)

2.3 Training the Machine

Machine learning research has always stressed the importance of distinguishing between (at least) two different types of uncertainty, often referred to as aleatoric and epistemic, in terms of any data generated by the ‘trained machine’ (Fu and Lee, 2013). Learning from data is inseparably connected with uncertainty. This is largely due to the fact that learning, understood as generalizing beyond a finite set of observed data, is necessarily based on a process of induction, i.e., replacing specific observations by general models of the data-generating process. Thus, a training dataset helps the learning of clustering patterns, decision trees, or CNN and then such a ‘learnt modeling’ is applied to a large data to complete segmentation and/or classification tasks. Naturally, such models are never provably correct but only hypothetical and therefore uncertain, and the same holds true for the predictions produced by a model. In addition to the uncertainty inherent in inductive inference, other sources of uncertainty exist, including incorrect model assumptions and noisy or imprecise data.

It is desirable and should be considered as a key feature of any machine learning method that is employed in tasks of high precision. While there are significant investments to raise the positional accuracy of point cloud data, it is important to invest in quantifying and reducing uncertainty of the ML derived processes of segmentation and classification.

It is important to understand the challenges of distinguishing between two very different sources of uncertainty: aleatoric uncertainty, which is due to statistical variability and effects that are inherently random, and epistemic uncertainty which is caused by a lack of knowledge (Kruse et al, 1991). Furthermore, classification using ML can be seen as a decision making problem and classification problem is often formalized within the framework of Bayesian decision theory. Detailed models have thus been developed through bayesian inference that

enable to quantify the uncertainty (Senge et al, 2014). Furthermore, it has been reported that total variance in machine learning processes can

indeed be decomposed as $\sigma_{combined} = \sigma_{aleatoric}^2 + \sigma_{epistemic}^2$ (Clements et al 2019)

Thus, we recognize that the combined uncertainty of the machine learning based classification outputs is explained by

$$\psi_{combined} = \psi_{aleatoric} \cdot \psi_{epistemic} \quad (1)$$

where $\psi_{epistemic}$ depends on the choice of the model, whereas $\psi_{aleatoric}$ is based on the nature and extent of the training data used. We use this understanding to develop an overall account of uncertainty for classified point cloud datasets.

THEORETICAL ACCOUNT OF UNCERTAINTY IN GENERATING CLASSIFIED POINT CLOUDS

While the different sources of uncertainty in lidar point clouds beyond positional uncertainty have been discussed in the context of the segmentation and classification process, we now focus on developing a combined framework of uncertainty by attempting to quantify uncertainty by developing indicators of uncertainty at each step of the development of a classified point cloud.

2.4 RMSE and uncertainty values based on positional accuracy

The accuracy of the position of features, including horizontal and vertical positions, with respect to horizontal and vertical datums. It is usually represented as Root Mean Square Error and is simply derived from the sum of the squares of the difference in the coordinates of the data points and the corresponding ground-truth. is expressed as

$$RMSE = \sqrt{\left[\frac{\sum ((E_{data,i} - E_{true,i})^2)}{n} \right]} \quad (2)$$

Thus, horizontal RMSE is expressed as

$$RMSE_{Hor} = \sqrt{\left[\frac{\sum ((RMSE_E)^2 + (RMSE_N)^2)}{n} \right]} \quad (3)$$

Similarly vertical RMSE is expressed as

$$RMSE_{Ver} = \sqrt{\left[\frac{\sum (RMSE_z)^2}{n} \right]} \quad (4)$$

These measures have been incorporated into various standards including the ASPRS guidelines (ASPRS, 2015), FGDC standards (www.fgdc.gov) besides others. Very often data collection projects specify minimum requirements for horizontal and vertical accuracy. However, it is highly likely that collected is collected at different levels of RMSE, especially when the point clouds from aerfgdcial lidar surveys and mobile or terrestrial lidars are combined. Such spatially varying uncertainties pose a unique challenge for quantifying uncertainty in the positional accuracy of a combined dataset. For example, the RMSE values of a combined dataset may be well below the ASPRS threshold (ASPRS, 2015) but a certain small section (probably surveyed using mobile lidar) is very high. It is therefore pragmatic to use localized RMSE threshold values such as

$$RMSE_{local} = \sqrt{\left[\frac{\sum_{LocalArea} (E_{data,i} - E_{true,i})^2}{n} \right]} \quad (5)$$

2.5 Accounting for uncertainty introduced in segmentation

Uncertainties introduced in the segmentation step include the exclusion of certain points from a segment or inclusion of unnecessary points. While it is not necessary to drop any points from the dataset, at this stage, the inability to group points from a segment together reduces the ability to classify the wrongly segmented points together with the object in our experience. Sparse points or highly dense datasets suffer from such challenges and hence become too ‘noisy’ for classification tasks.

While the parameters that can represent accuracy of the segmentation process have been discussed earlier, we recollect that CDM, AFI as well as the percent of area lost/gained in segments are all normalized measures and can be helpful in quantifying the increase in uncertainty. We can express this increase in uncertainty as

$$\psi_{seg} = \frac{1}{CDM_{normalized}} \quad (6)$$

2.6 Uncertainty due to machine learning processes

The machine learning process accounts for at least two categories of uncertainty that is introduced. If the segmentation step utilizes a ML based clustering algorithm, the increase in uncertainty of equation 6 could be subdivided into epistemic and aleatoric uncertainties (Hüllermeier and Waegeman, 2019). The characterization of such uncertainty is thus the same as that for the classification step discussed henceforward.

Epistemic uncertainty of the ML based classification is primarily dependent on the suitability of the ML algorithm for the learning task. As expressed in table 2, these are characteristic of each ML technique although the techniques of calculating such values have only evolved recently. Since the highest possible values of accuracy in machine learning

While ‘class balanced datasets’, clear distinction between different classes and possibly with cases with fewer number of classes are expected to yield better results in aleatoric sense (Fu and Lee, 2013). Thus, it is possible to obtain more specific components of aleatoric uncertainty (i) based on coverage (and hence volume) and quality of the training dataset and (ii) the nature of the classes that are used for the ML (how distinct are the features that determine membership to a class). It is also important to note that higher uncertainty in segmentation (equation 6) is expected to result in higher aleatoric uncertainty

and vice-versa. Thus, we claim $\psi_{aleatoric} \propto \psi_{seg}$

Also if χ , Ω represent the size of the training data and the quality respectively on a normalized scale (where 0 represents the best case scenario and 1 is the worst), then

$$\psi_{aleatoric} = K \cdot \chi \cdot \Omega \cdot \psi_{seg} \quad (7)$$

where K is otherwise constant.

2.7 Other sources of uncertainty

Many other sources of uncertainty are introduced in the context of point cloud data which include imprecision of the geographic referencing, incompleteness of the scanning whereby certain parts or segments may be missing or partial, fuzzy nature of boundaries of the segments and objects being captured (and later classified). We denote such uncertainty as ψ_{multiple} and postulate that this can be further decomposed into measurable components.

2.8 Combined measures of Uncertainty

It is now possible to combine the multiple sources of uncertainty to account for an overall uncertainty value for the processed lidar point cloud. It is possible to add all uncertainties including RMSE and the different ψ values. It is also possible to use weightages for each. However, it may also be pragmatic to use the different values without combining them as this can help identify the source of uncertainty and help reduce it.

2.9 Examining uncertainty in power utility datasets

Power utilities have turned to using lidar point clouds to manage their assets and ensure preventive maintenance. We use data obtained from one such project to highlight the aspect of uncertainty in lidar point clouds as discussed earlier.

With high positional accuracy, the data shows very high positional accuracy in general ($RMSE_{\text{hor}} < 0.05$ ft, $RMSE_{\text{ver}} < 0.05$ ft). However, at segmentation level we notice some level of failures that causes noise (see figure 1). Note that the noise points are left alone by the classifier. If the failure occurs at classification level (see figure 2), the whole segment is wrongly labeled as a different class (in this case a vertical linear object is classified as a pole). In figure 3 we notice that the dense point cloud above the cables is classified as vegetation and it is possibly an epistemic uncertainty associated with the classifier. In some cases segmentation fails to identify the cable and hence classification fails to label them correctly. The semantic segmentation hence shows an uncertainty value > 0 and can range up to 0.2 (beyond which the classification is reattempted). The ability to locate the power cable in a continuous manner is critical for the application of this lidar point cloud and as seen in figure 4, the failure to have a continuous cable could result from uncertainty in the data capture, segmentation or the classification. These have all been accounted for in our theoretical discussion.

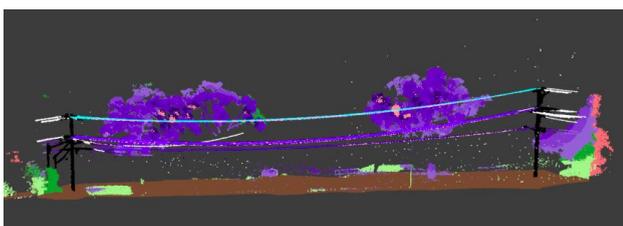


Figure 1. Salt and pepper noise in point cloud data resulting from poor segmentation.

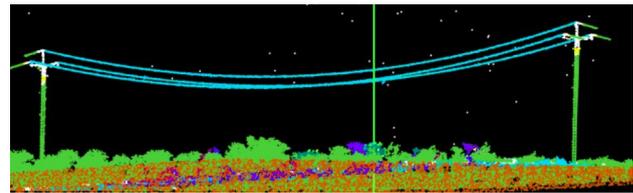


Figure 2. Misclassification of a linear object as a pole. It can be argued that the training data did not have any such vertical feature to be classified and hence the misclassification is a result of aleatoric uncertainty.

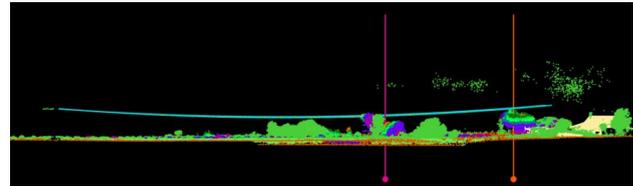


Figure 3. Misclassification of a dense segment of noise points above the cables as vegetation. Since the points are dense and hence would have been labelled by the classification algorithm in any case. Thus, it is easy to suspect epistemic uncertainty in this case.

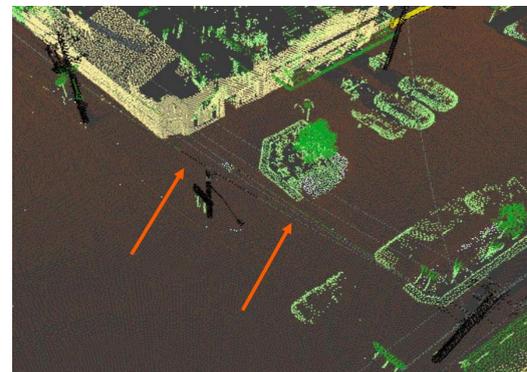


Figure 4. Failure to classify cables could be caused because of the uncertainty in the collecting data (incomplete capture), faulty segmentation as well as classification errors.

2.9.1 Computing uncertainty and reclassification

It is important to note that the cost of creating classified point cloud data includes a major component of segmentation and classification. As the cost of lidar surveys outpace that of traditional or photogrammetry based surveys, it is important to remember that even with very low RMSE values (< 0.05 ft), there are significant uncertainties that remain associated with lidar point clouds. Due to the dense point clouds and massive datasets, it is convenient (almost compulsory) to use ML based techniques to generate classified point clouds. The cost of processing data in terms of manually validating classified power utility data for the example above can exceed one person-day for less than 10 km. Furthermore, in worst cases the classification is required to be repeated and hence cause higher expenses and project delays.

Since manual classification forms the basis of most training datasets as well as validation of the classified point clouds, human errors and omissions are critical to the framework of point cloud data processing. It is important to account for all investments in creating the cloud point data while stating its quality and not relying only on positional accuracy.

The volume and velocity of point cloud data processing by what is commonly termed as spatial big data analysis (Shirowzhan et al 2019) makes the quantification of uncertainty a challenging task (and an expensive one). However, it is important to ensure that the data adheres to acceptable quality standards as shown in the running example.

3. CONCLUSIONS AND FUTURE WORK

We have reviewed multiple sources of uncertainty in lidar point clouds and have shown that a holistic approach to spatial data quality is necessary. The specific need to account for uncertainty introduced at the segmentation and classification stage has been highlighted. The understanding of epistemic and aleatoric uncertainties to reduce classification errors in the machine learning based classification of point clouds is an important aspect of this holistic approach.

Our experience with classification shows that there multiple aspects of uncertainty associated with the acquisition and processing of lidar point clouds. With higher precision of the laser based sensing mechanism and prescribed standards of accuracy, it is important to recognize that manual 3D segmentation is a complex procedure which requires a skilled user, patience and an acute eye for detail. Although there are multiple available training datasets, these are not necessarily contextual geospatial applications at hand. We demonstrate through the power utility example that the cost of using lidar point data is closely related to the uncertainties associated with the uncertainties and hence to overall spatial data quality. With higher investments in lidar based spatial data, it is therefore important to revisit Goodchild's article (Goodchild, 1998), and work towards specification of uncertainties endemic to generation of classified point cloud data.

We believe that this area of spatial data quality is an emerging topic of research. Some of the areas for future exploration (but not limited to) are:

- Develop a detailed framework to statistically account for uncertainties in the machine learning framework, especially in deep learning. Such work can benefit from the progress made by the data science communities in providing epistemic and aleatoric values of uncertainty. Of particular interest are the epistemic uncertainties of a classifier, especially using Monte Carlo Dropout (MCD) (Miok et al 2019).
- More importantly, it is important to develop metrics for spatial data quality that considers the different sources of uncertainty rather than positional accuracy alone. Thus, the theoretical framework stated in this paper could be the basis of a spatial data quality metric and could be used to understand the cost of lidar data projects (Hummel et al 2011)
- Training datasets used to classify point cloud data are often not inline with classification requirements. Thus, it is important to understand the role of inadequate training in the quality of classified point datasets and hence will be useful to compare the relative uncertainty of data produced by classifiers that have been trained using different datasets (Fu and Lee, 2013; Wang, et al 2019).
- The aspects of semantics in 'semantic segmentation' and the use of shared vocabularies are important to resolve ambiguities in 'user defined' classes of lidar

point clouds. Research related to geospatial ontologies and probabilistic ontologies (Sen, 2008) can provide solutions to resolve such challenges.

ACKNOWLEDGEMENTS

We acknowledge the support of Genesys International and colleagues for the use case and data related to the power utility example.

REFERENCES

- ASPRS, 2015. ASPRS positional accuracy standards for digital geospatial data. *Photogrammetric Engineering and Remote Sensing*, 81(3), pp.1-26.
- ASPRS, 2013. LAS specification, version 1.4 – R13 (PDF). American Society for Photogrammetry and Remote Sensing. 2013-07-15 available at https://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf
- Chen, Y., Su, W., Li, J., Sun, Z., 2009. Hierarchical object oriented classification using very high resolution imagery and LiDAR data over urban areas. *Advances in Space Research* 43 (7), 1101–1110.
- Clements, W.R., Robaglia, B.M., Van Delft, B., Slaoui, R.B. and Toth, S., 2019. Estimating risk and uncertainty in deep reinforcement learning. arXiv preprint arXiv:1905.09638.
- Clements, W.R., Robaglia, B.M., Van Delft, B., Slaoui, R.B. and Toth, S., 2019. Estimating risk and uncertainty in deep reinforcement learning. arXiv preprint arXiv:1905.09638.
- El-Ashmawy, N., Shaker, A., , Yan, W. Y., 2011. Pixel vs object-based image classification techniques for LiDAR intensity data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5/W12), 43–48.
- Fu, J. and Lee, S., 2013. Certainty-based active learning for sampling imbalanced datasets. *Neurocomputing*, 119, pp.350-358.
- Goodchild, M.F. "Uncertainty: the achilles heel of gis?" *Geo Info Systems*, vol. 8, no. 11, pp. 50–52, 1998.
- Griffiths, D. and Boehm, J., 2019. A review on deep learning techniques for 3D sensed data classification. *Remote Sensing*, 11(12), p.1499.
- Grilli, E., Menna, F. and Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, p.339.f
- Hug, C. and Wehr, A., 1997. Detecting and identifying topographic objects in imaging laser altimeter data. *International archives of photogrammetry and remote sensing*, 32(3 SECT 4W2), pp.19-26.
- Hüllermeier, E. and Waegeman, W., 2019. Aleatoric and epistemic uncertainty in machine learning: A tutorial introduction. arXiv preprint arXiv:1910.09457.

- Hummel, S., Hudak, A.T., Uebler, E.H., Falkowski, M.J. and Megown, K.A., 2011. A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. *Journal of forestry*, 109(5), pp.267-273.
- Kruse, R., Schewecke, E. and Heinsohn, J., 2012. *Uncertainty and vagueness in knowledge based systems: numerical methods*. Springer Science & Business Media.
- Lu, X., Yao, J., Tu, J., Li, K., Li, L. and Liu, Y., 2016. Pairwise Linkage for POint Cloud Segmentation. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 3(3).
- Lucieer, A., 2004. *Uncertainties in Segmentation and their Visualisation* (Doctoral dissertation, Utrecht University and International Institute for Geo-Information Science and Earth Observation (ITC)).
- Maas, H.G., 1999, June. The potential of height texture measures for the segmentation of airborne laser scanner data. In *Fourth international airborne remote sensing conference and exhibition/21st Canadian symposium on remote sensing* (Vol. 1, pp. 154-161).
- Marpu, P.R., Neubert, M., Herold, H. and Niemeyer, I., 2010. Enhanced evaluation of image segmentation results. *Journal of spatial science*, 55(1), pp.55-68.
- Maxwell, A.E., Warner, T.A. and Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), pp.2784-2817.
- Minh, N. Q., Hien, L. P., 2011. Land cover classification using LiDAR intensity data and neural network. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 29 (4), 429–438.
- Miok, K., Nguyen-Doan, D., Zaharie, D. and Robnik-Šikonja, M., 2019, September. Generating Data using Monte Carlo Dropout. In *2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP)* (pp. 509-515). IEEE.
- Monmonier, M.S., 1977. *Maps, Distortion, and Meaning*. Resource Paper No. 75-4.
- Mosavi, A., Ghamisi, P., Faghan, Y. and Duan, P., 2020. *Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics*. arXiv preprint arXiv:2004.01509.
- Mosavi, A., Ghamisi, P., Faghan, Y. and Duan, P., 2020. *Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics*. arXiv preprint arXiv:2004.01509.
- Prieto, M.S. and Allen, A.R., 2003. A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10), pp.1265-1273.
- Rottensteiner, F. and Briese, C., 2002, September. A new method for building extraction in urban areas from high-resolution LIDAR data. In *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* (Vol. 34, No. 3/A, pp. 295-301). Natural Resources Canada.
- Roy, A.G., Conjeti, S., Navab, N. and Wachinger, C., 2018, September. Inherent brain segmentation quality control from fully convnet monte carlo sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 664-672). Springer, Cham.
- Sasaki, T., Imanishi, J., Ioki, K., Morimoto, Y., Kitada, K., 2012. Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data. *Landscape and Ecological Engineering* 8 (2), 157–171.
- Sen, S., 2008. Framework for probabilistic geospatial ontologies. *International Journal of Geographical Information Science*, 22(7), pp.825-846.
- Senge, R., Bösner, S., Dembczyński, K., Haasenritter, J., Hirsch, O., Donner-Banzhoff, N. and Hüllermeier, E., 2014. Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty. *Information Sciences*, 255, pp.16-29.
- Senge, R., Bösner, S., Dembczyński, K., Haasenritter, J., Hirsch, O., Donner-Banzhoff, N. and Hüllermeier, E., 2014. Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty. *Information Sciences*, 255, pp.16-29.
- Shirowzhan, S., Tan, W. and Sepasgozar, S.M., 2020. Digital Twin and CyberGIS for Improving Connectivity and Measuring the Impact of Infrastructure Construction Planning in Smart Cities.
- Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S. and Vercauteren, T., 2019. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338, pp.34-45.
- Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S. and Vercauteren, T., 2019. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338, pp.34-45.
- Yadav, M., Lohani, B. and Singh, A.K., 2018. Road surface detection from mobile lidar data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, pp.95-101.
- Yan, W.Y., Shaker, A. and El-Ashmawy, N., 2015. Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, 158, pp.295-310.