Resolving the Broccoli Problem: Identifying Optimal Computational Algorithms for the Accuracy Assessment of Tree Delineations from Remotely-sensed Data

Jon Murray[*1], Dr. David Gullick[1], Prof. George Alan Blackburn[1],
Dr. James Duncan Whyatt[1], & Dr. Christopher Edwards[2].

[1] *Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ*
[2] *School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA.*
*Corresponding author: Tel: +44 1524 652 01 Email: j.murray3@lancaster.ac.uk

January 12, 2018

Summary

For many different investigative purposes, trees and forests are aerially scanned using light detection and ranging (LiDAR). Often, this also requires the manual measurement of ground reference (GR) plots within the LiDAR scan. Upon analysis, there is regularly a mismatch between the alignment of GR and LiDAR tree locations, crown areas and tree heights. This anomaly is frequently overlooked and under-reported in the current literature. This study investigates the suitability of match pairing algorithms for the quantification of misalignment errors between two datasets representing GR and LiDAR data, and recommends an algorithm for accurately quantifying match-pairing differences.

KEYWORDS: Tree delineation, accuracy assessment, aerial LiDAR, matched pairing, algorithms.

1. Background

Light detection and ranging (LiDAR) investigations from aerial platforms are frequently used for the investigation of forests, woodlands or individual trees. These remote sensing (RS) operations are used to capture various forest attributes, including; data for forest inventory and mensuration, investigations into carbon accounting, the assessment of ecological habitats and for a range of environmental modelling purposes using geo-spatial analyses (Eysn et al., 2012; Jakubowksi et al., 2013; Wu et al., 2016). For these and other similar investigations, it is common practice to establish ground reference (GR) plots within the area that is scanned as part of the LiDAR flightline. The GR plots would typically be directly measured in as high level of detail as it is possible to achieve using manual methods, with the intention of enabling data training or validation of the indirectly measured RS data in attempts to minimise errors (Chen et al., 2006). Typically, this process requires the absolute position of the tree being recorded as where the stem emerges from the ground, attempting to avoid any ground assessment errors of determining where the aerial parts of the tree are located within space, albeit assessed only from the ground viewpoint position (Mills et al., 2010). When using LiDAR for individual tree crown (ITC) delineation, methods may vary according to the scale of the investigation and the intended use of the data. Frequent delineation approaches include; manual delineation, watershed or inverse watershed delineation, and local maxima delineation methods. For coniferous trees with their regular shape and form, ITC delineation is a readily repeatable exercise. However, broadleaved trees with their uniquely shaped crowns that differ in relation to their location and exposure to available sunlight, which causes unpredictable changes to the tree shape (Loehle, 1986), offer a more complex ITC delineation problem: this is the Broccoli problem.

Fundamentally, many of the ITC delineation methods use the calculation of high peaks and valley troughs within the LiDAR data to designate the central location, crown area and height of the RS trees.

However, as Valubuena (2014) describes, there are frequently differences between the location of the tree as measured in the GR data capture, and the location of the same tree when it is identified in the RS data. An accepted status quo regarding identifying the match between two data sets appears to be the acceptance of a 'reasonable' match, on the grounds of the assessment of indicative variables, or by applying arbitrary thresholds. For example; a match pairing will be accepted if a single tree, as represented in the GR and LiDAR datasets, is spatially located within 5 metres of one another (Næsset, 2002; Hyyppa et al., 2012; Listopad et al., 2011). It has been stated that there is a requirement for a standardised methodology to evaluate the success tree delineation methods (Zhen et al., 2016), although it is understood that no discrete index for quantifying the success of the match is currently available.

2. Aim and Objectives

The aim of this work was to develop a methodology for the assessment of match pairing algorithms to determine ITC delineation accuracy. This approach was developed for use with the complex canopies of broadleaved trees as situated within a woodland setting. This was achieved through the development of a synthetic environment for the creation and testing of computationally modelled trees. Matched pairing is a mathematical theory used to assess the correspondence between an object in one data set, with an object in another data set, based upon predetermined, observable characteristics. Different match pairing algorithms were used to quantify the degree of match pairing alignment between two data sets representing, for example; GR to LiDAR data, or LiDAR to LiDAR data. Developmental testing included the progressive increase of data noise into the synthetic environment to identify which algorithms are best suited to quantifying the matched pairing differences.

3. Method

When trees are scanned in LiDAR they are measured and recorded in 3D space, however, ITC delineation is fundamentally a 2D problem where tree locations, crown area and crown extent are expressed in the 2D Euclidean plane either as an organically shaped polygon, or as a formal 2D shape, typically a circle. A computational environment was created where a single, modelled tree was generated. The known tree characteristics included the tree's location, crown area and height. These characteristics were considered representative of the data captured for a typical broadleaved tree for GR data. A series of changes to the location, crown area and height were applied, thereby creating a second tree which represented a new dataset. The second tree was considered to exhibit the characteristics of a broadleaved tree that has been ITC delineated from aerially scanned LiDAR data. As the differences between the two trees was already identified, this was used to form a known 'divergence' model. A series of match pairing algorithms were tested against the two trees to identify which algorithm gave the most accurate quantification of the differences between the two trees when compared with the divergence model. The algorithms used were the calculation of the Hausdorff distance (HD), Hausdorff distance with multivariate inputs (HD++), and the Hungarian combinatorial optimisation algorithm, also known as the Hungarian method (HM). An assignment ranking index was developed that was based upon the accumulated difference values between the matched pairings. For example, if a pair of trees from two data sets matched, the difference in values between the location, height and area were scored to indicate the quality of the pairing. The number of agreed pairings between the two datasets was also recorded. To test the efficacy of the algorithms, data noise was added incrementally on a scale from 1-5 with randomised changes in tree location, crown area and height. This was achieved as a normal distribution can be described by a Gaussian function given both a mean and a standard deviation (Figure 1), where the properties of each tree were modified by +/- of a random sample from the appropriate distribution at each noise level (Table 1). Repeat testing also increased the modelled tree population from one in each dataset to a population of ten, and finally to populations of 500, thereby creating a synthetic environment that replicated trees in a complex woodland setting.
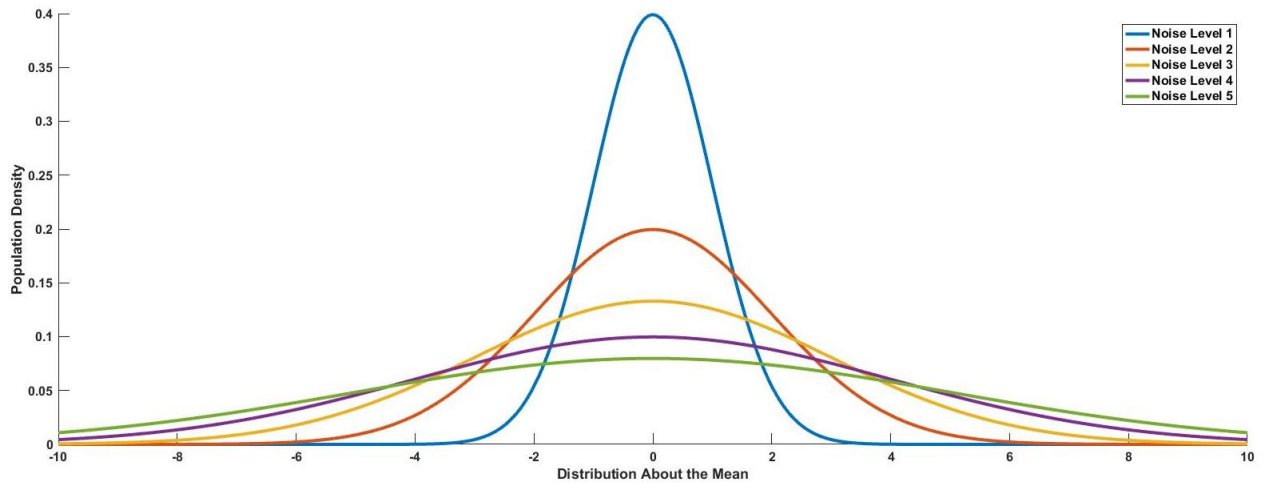
Figure 1   Gaussian curves demonstrating the effect on the data population by increasing standard deviations.

Table 1   Introduction of data noise following modification of the normal distribution and standard deviation (SD) effect on the data population relative to data noise levels.

| Data Noise Level	Population (%) by Standard Deviation (SD)
1	SD1 = 68% +/-1, 95% +/-2, 99% +/-3
2	SD2 = 68% +/-2, 95% +/-4, 99% +/-6
3	SD3 = 68% +/-3, 95% +/-6, 99% +/-9
4	SD4 = 68% +/-4, 95% +/-8, 99% +/-12
5	SD5 = 68% +/-5, 95% +/-10, 99% +/-15

4. Results

Following the expansion of the tree populations to 500 trees, and the introduction of the highest level of data noise, level 5, the metrics HD, HD++ and HM were compared against the known divergence model to produce r^2 and root mean squared error values (RMSE). The pairing success and assignment rank was also calculated. The two datasets were representative of GR and LiDAR data. The algorithm HD paired 59% of trees between GR and LiDAR, with an assignment rank of 30969, $r^2 = 0.165$ and RMSE = 1056.8. The algorithm HD++ paired 42.2% of all available trees, with an assignment rank of 44510, $r^2 = 0.075$ and RMSE = 2235.2. The HM algorithm paired 100% of the trees, with an assignment rank of 17942.2, $r^2 = 0.898$ and RMSE = 195.5 (Figure 2).

Hausdorff Distance (HD)
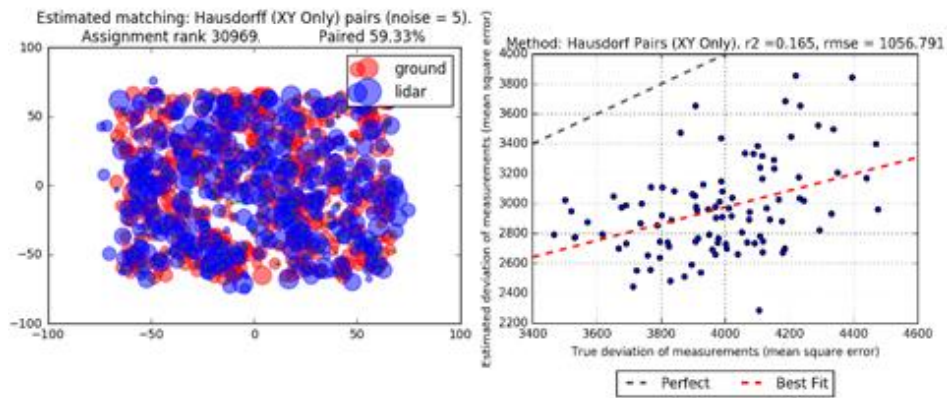
Hausdorff Distance with Multivariate Inputs (HD++)
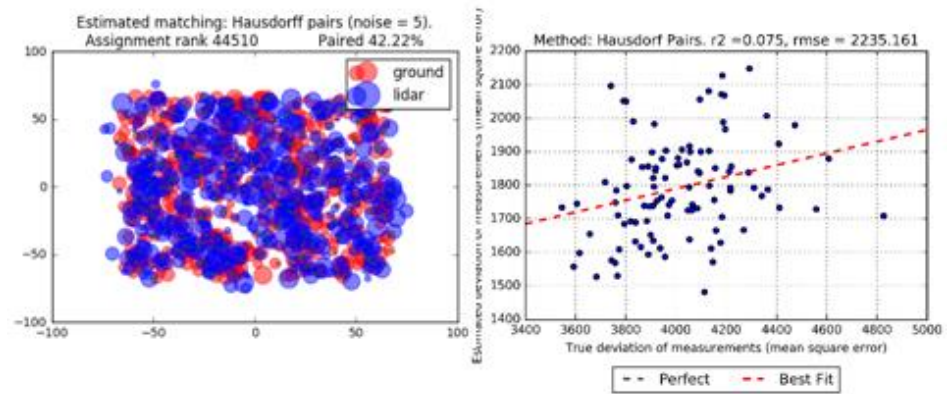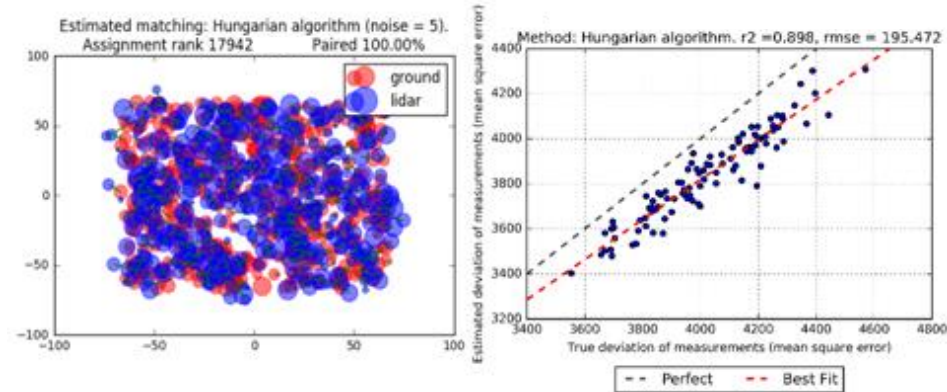

Hungarian Method (HM)

Figure 2   Three algorithms tested for their suitability in identifying the matched pairing alignment between two datasets with high levels of data noise, which represent individual tree delineations from remotely sensed data. The HD and HD++ algorithms achieve low numbers of matches between the two datasets and perform poorly against the expected divergence model (r2, RMSE). The HM algorithm achieves a 100% paired agreement between the two datasets and high correspondence with the expected divergence model (r2, RMSE).

5. Conclusion

Our findings demonstrate it is possible to use specific algorithms to quantify the matched paring success between two distinct data sets. We demonstrate in a synthetic environment with known amounts of data noise, that the algorithms HD, HD++ and HM can all quantify the matched paring task with varying degrees of success. However, it is evident that the HM algorithm is best suited to quantifying the success of matched pairing, and in dealing with the structural complexities of ITC delineation in a broadleaved

woodland. These results also indicate that the HM algorithm could be used to quantify matched paring between real-world GR and LiDAR data. We envisage further work will include the refinement of using the HM algorithm, the normalisation of the assignment ranking and the use of the HM algorithm with real-world acquired data.

6. Acknowledgements

7. Biography

Jon Murray is currently writing up his PhD. Entitled, "The Remote Sensing of Trees" at Lancaster University, and is a research associate at the University of Manchester. Jon's research interests involve using remote sensing to investigate tree structure, forestry, agriculture and environmental management.

8. References

Chen, Q., Baldocchi, D., Gong, P. & Kelly, M. 2006. Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. *Photogrammetric Engineering and Remote Sensing,* 72(8)**,** 923-932.

Eysn, L., Hollaus, M., Schadauer, K. & Pfeifer, N. 2012. Forest Delineation Based on Airborne Lidar Data. *Remote Sensing,* 4(3).

Hyyppa, J., Yu, X. W., Hyyppa, H., Vastaranta, M., Holopainen, M., Kukko, A., Kaartinen, H., Jaakkola, A., Vaaja, M., Koskinen, J. & Alho, P. 2012. Advances in Forest Inventory Using Airborne Laser Scanning. *Remote Sensing,* 4(5)**,** 1190-1207.

Jakubowksi, M. K., Guo, Q., Collins, B., Stephens, S. & Kelly, M. 2013. Predicting Surface Fuel Models and Fuel Metrics Using Lidar and Cir Imagery in a Dense, Mountainous Forest. *Photogrammetric Engineering & Remote Sensing,* 79(1)**,** 37-49.

Listopad, C., Drake, J. B., Masters, R. E. & Weishampel, J. F. 2011. Portable and Airborne Small Footprint Lidar: Forest Canopy Structure Estimation of Fire Managed Plots. *Remote Sensing,* 3(7)**,** 1284-1307.

Loehle, C. 1986. Phototropism of Whole Trees: Effects of Habitat and Growth Form. *The American Midland Naturalist,* 116(1)**,** 190-196.

Mills, S. J., Castro, M. P. G., Li, Z. R., Cai, J. H., Hayward, R., Mejias, L. & Walker, R. A. 2010. Evaluation of Aerial Remote Sensing Techniques for Vegetation Management in Power-Line Corridors. *Ieee Transactions on Geoscience and Remote Sensing,* 48(9)**,** 3379-3390.

Næsset, E. 2002. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. *Remote Sensing of Environment,* 80(1)**,** 88-99.

Valbuena, R. 2014. Integrating Airborne Laser Scanning with Data from Global Navigation Satellite Systems and Optical Sensors. *In:* MALTAMO, M., NÆSSET, E. & VAUHKONEN, J. (eds.) *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies.* Dordrecht: Springer Netherlands.

Wu, B., Yu, B., Wu, Q., Huang, Y., Chen, Z. & Wu, J. 2016. Individual Tree Crown Delineation Using Localized Contour Tree Method and Airborne Lidar Data in Coniferous Forests. *International Journal of Applied Earth Observation and Geoinformation,* 52**,** 82-94.

Zhen, Z., Quackenbush, L. J. & Zhang, L. J. 2016. Trends in Automatic Individual Tree Crown Detection and Delineation-Evolution of Lidar Data. *Remote Sensing,* 8(4)**,** 26.