Application of small area variance estimates of forest parameters using earth observation auxiliary variables and a k-Nearest Neighbours technique.

Dylan Walshe[1], Daniel McInerney[1], Joao Paulo Pereira[1], and Kenneth A. Byrne[2]

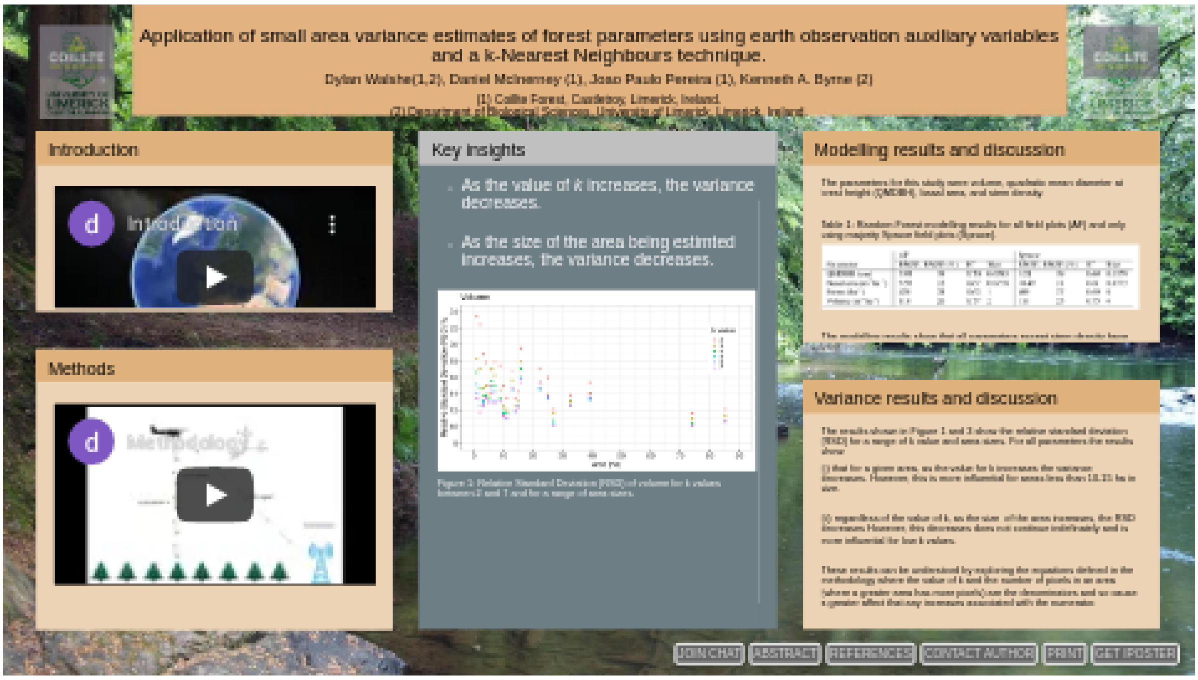
[1]Coillte
[2]University of Limerick

November 23, 2022

Abstract

Combining auxiliary variables and ground data of forest parameters using the model-based approach has become a frequently used methodology to produce synthetic estimates for small areas. These small areas arise as it may not be financially feasible to take ground measurements or some areas may be inaccessible to ground-based personnel. Until recently, these estimates have been calculated without providing a measure of the variance or in large homogeneous forested areas where the variability would be minimal. This paper uses a Random Forest algorithm to produce estimates of QMDBH (cm) and volume (m3 ha-1) and a k-nn technique to produce variance estimates in a previously unproven heterogeneous forest environment. The area of interest (AOI) was the commercial forest in the Slieve Bloom Mountains in the Republic of Ireland, where the majority species are Sitka spruce (Picea sitchensis (Bong.) Carr.) and Lodgepole pine (Pinus contorta Dougl.). Field plots were measured during the Summer in 2018 during which a LiDAR campaign was flown and Sentinel 2 satellite imagery captured, which were both used as auxiliary variables. Root mean squared errors (rmse) for the modeled estimates of QMDBH and volume were 4.4 cm and 111 m3 ha-1 and the r2 values were 0.65 and 0.76 respectively. An independent dataset of pre-harvest forest stands was used to validate the modeled variance estimates. The results showed that 73% of QMDBH measurements and 60% of volume measurements were within the confidence intervals of the estimated values. The mean percentage standard deviation for QMDBH and volume were 17% and 21% respectively. This application of variance estimation in such a heterogeneous forest landscape adds further weight to the applicability of the methodology in a range of forest landscapes. This finding is extremely important as it highlights the benefits of using earth observation data to produce estimates and confidence intervals for small areas in heterogeneous forest landscape.

Application of small area variance estimates of forest parameters using earth observation auxiliary variables and a k-Nearest Neighbours technique.



Dylan Walshe(1,2), Daniel McInerney (1), Joao Paulo Pereira (1), Kenneth A. Byrne (2)

(1) Coillte Forest, Castletroy, Limerick, Ireland.
(2) Department of Biological Sciences, University of Limerick, Limerick, Ireland.

PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

INTRODUCTION

[VIDEO] https://www.youtube.com/embed/fS9ypW77PlE?rel=0&fs=1&modestbranding=1&rel=0&showinfo=0

This video was made by using Google Earth Pro, Rayshader in R (Morgan-Wall, 2020) , and Openshot.

METHODS

[VIDEO] https://www.youtube.com/embed/k4gMj0VAH0k?rel=0&fs=1&modestbranding=1&rel=0&showinfo=0

KEY INSIGHTS

- As the value of *k* increases, the variance decreases.

- As the size of the area being estimted increases, the variance decreases.

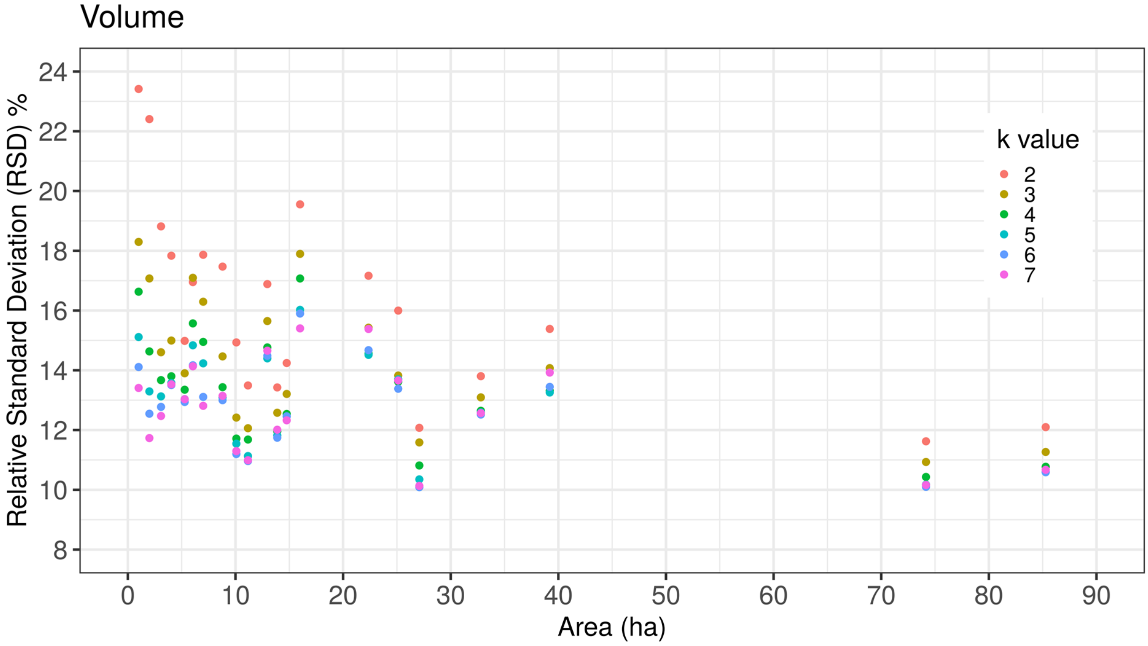


Figure 1: Relative Standard Deviation (RSD) of volume for k values between 2 and 7 and for a range of area sizes.

MODELLING RESULTS AND DISCUSSION

The parameters for this study were volume, quadratic mean diameter at brest height (QMDBH), basal area, and stem density.

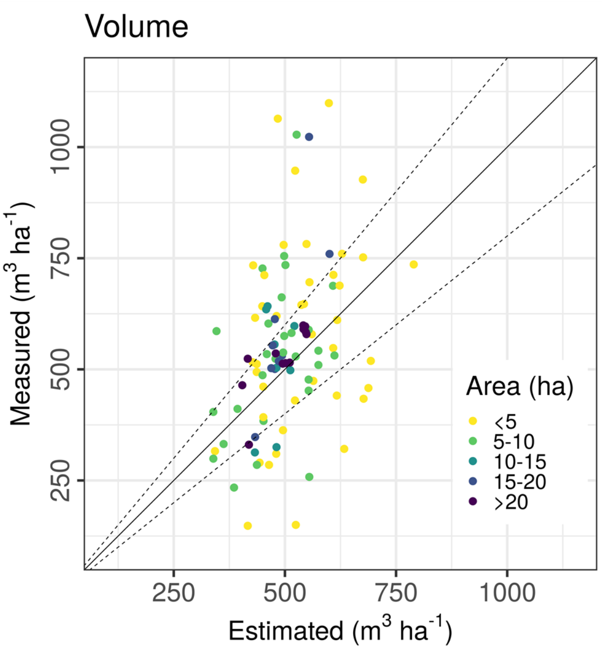
Table 1: Random Forest modelling results for all field plots (AF) and only using majority Spruce field plots (Spruce).

	AF				Spruce			
Parameter	RMSE	RMSE (%)	R^2	Bias	RMSE	RMSE (%)	R^2	Bias
QMDBH (cm)	3.95	19	0.70	0.0303	4.21	19	0.69	0.1179
Basal area (m^2 ha^{-1})	9.95	22	0.67	0.0778	10.49	21	0.61	0.0775
Stems (ha^{-1})	420	28	0.62	1	409	27	0.69	8
Volume (m^3 ha^{-1})	110	26	0.77	2	111	23	0.75	4

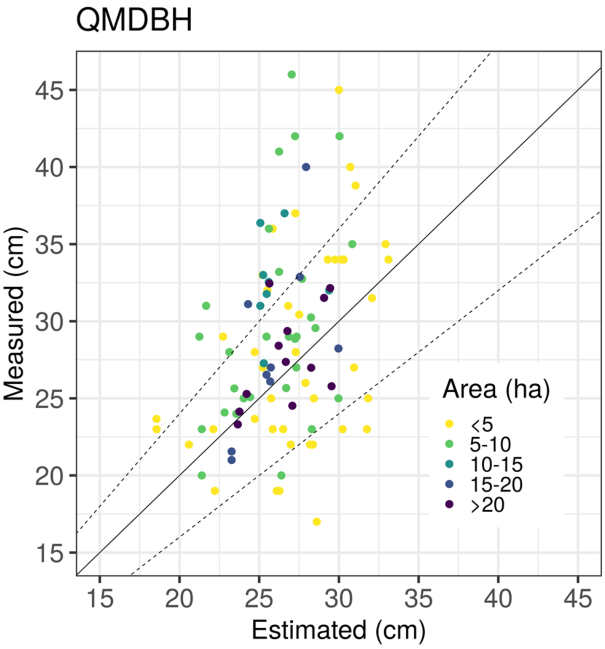
The modelling results show that all parameters except stem density have been accurately captured by the model. Stem density has been problematic for other studies due to the sub-dominant tree stems that do not reach the upper canopy likely being missed in the LiDAR returns (Maltamo et al., 2004, Woods et al., 2008).

The results illustrated in Figure 2, which are a comparison of pre-harvest forested areas and estimated values, show that regardless of the parameter, as the area increases, the measured and estimated values become more analogous. This is expected as the greater the area, the greater the number of field plots contained in the mean measured value and the greater the number of pixels in the mean estimated value and so both become more representative of the area as the size increases.

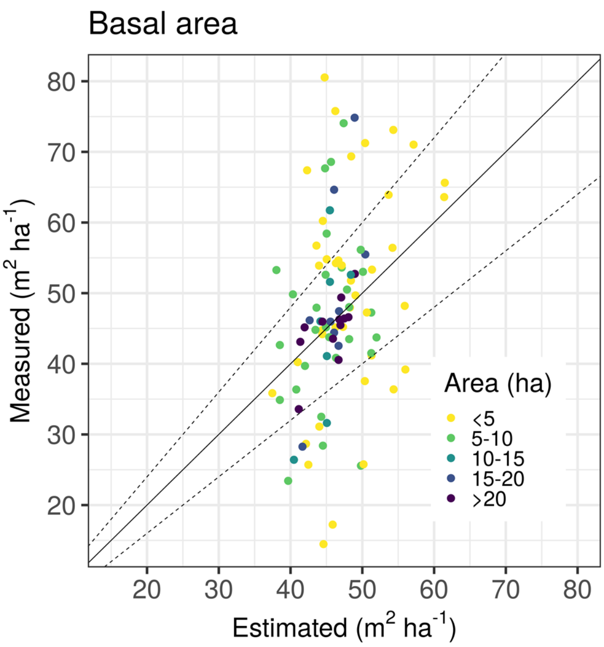
(a)



(b)


QMDBH
Measured (cm)
Estimated (cm)
15 20 25 30 35 40 45
Area (ha)
<5
5-10
10-15
15-20
>20

(c)


Basal area
Measured (m^2 ha^{-1})
Estimated (m^2 ha^{-1})
20 30 40 50 60 70 80
Area (ha)
<5
5-10
10-15
15-20
>20

(d)

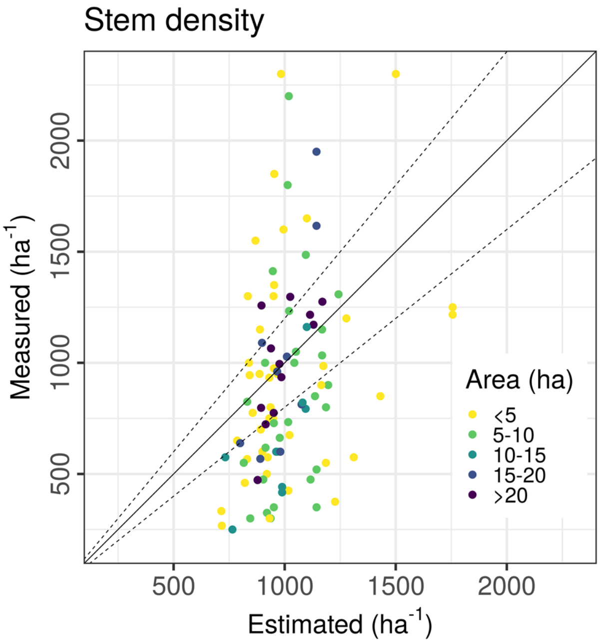


Figure 2: Estimated versus measured (a) volume, (b) QMDBH, (c) basal area, and (d) stem density for a range of areas, a 1:1 line (solid), and 20% error lines (dashed).

VARIANCE RESULTS AND DISCUSSION

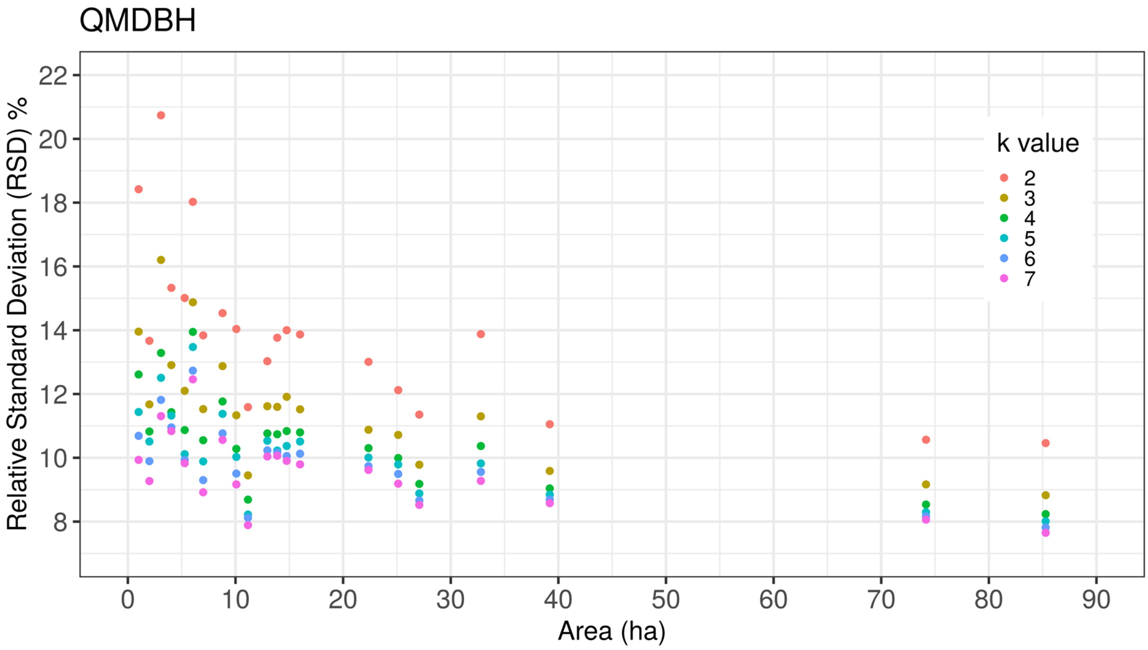
The results shown in Figure 1 and 3 show the relative standard deviation (RSD) for a range of k value and area sizes. For all parameters the results show:

(i) that for a given area, as the value for k increases the variance decreases. However, this is more influential for areas less than 10-15 ha in size.

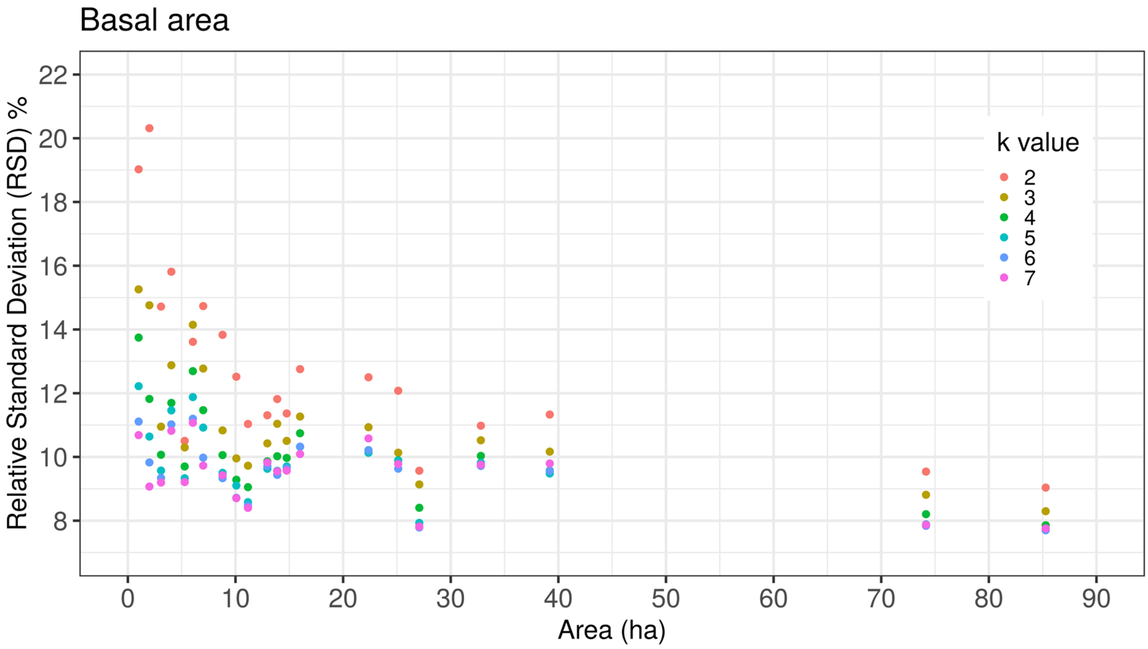
(ii) regardless of the value of k, as the size of the area increases, the RSD decreases However, this decreases does not continue indefinately and is more influential for low k values.

These results can be understood by exploring the equations defined in the methodology where the value of k and the number of pixels in an area (where a greater area has more pixels) are the denominators and so cause a greater affect that any increases associated with the numerator.

(a)



(b)



(c)

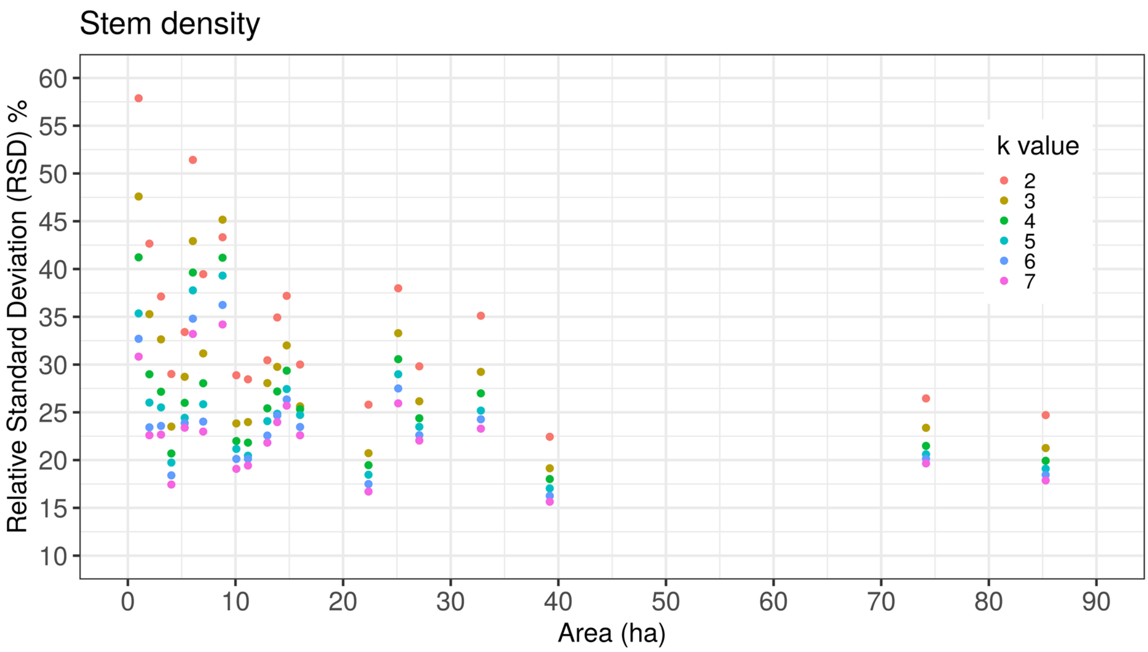


Figure 3: Relative Standard Deviation (RSD) of (a) QMDBH, (b) basal area, and (c) stem density for k values between 2 and 7 and for a range of area sizes.

Acknowledgements

The authors would like to thank the Irish Research Council and Coillte who have provided the funding for this research.

ABSTRACT

Combining auxiliary variables and field inventory data of forest parameters using the model-based approach is frequently used to produce synthetic estimates for small areas. These small areas arise when it may not be financially feasible to take ground measurements or when areas are inaccessible to field inventory crews. Until recently, these estimates have been calculated without providing a measure of the variance. This paper uses a Random Forest algorithm to produce estimates of quadratic mean diameter at breast height (QMDBH) (cm), basal area (m^2 ha), stem density (ha), and volume (m^3 ha) and subsequently estimates the variance using a *k*-NN technique. The area of interest (AOI) was the state owned commercial forests in the Slieve Bloom mountains in the Republic of Ireland, where the main species are Sitka spruce (*Picea sitchensis* (Bong.) Carr.) and Lodgepole pine (*Pinus contorta Dougl*.). Field plots were measured in summer 2018 during which a LiDAR campaign was flown and Sentinel 2 satellite imagery captured, which were both used as auxiliary variables. Root mean squared errors (RMSE) and R^2 values for the modelled estimates of QMDBH, basal area, stem density, and volume were 3.95 cm (0.70), 9.95 m^2 ha (0.67), 420 ha (0.62), and 110 m^3 ha (0.77) respectively. An independent dataset of pre-harvest forest stands was used to validate the modelled estimates. A comparison of measured values versus modelled estimates was carried out for a range of area sizes with results showing that estimated values in areas less than 10 - 15 ha in size exhibit greater uncertainty. However, as the size of the area increased, the estimated values became increasingly analogous to the measured values for all parameters. The results of the variance estimation highlighted (i) a greater value of *k* was needed for small areas compared to larger areas in order to obtain a similar relative standard deviation (RSD) and (ii) as the area increased in size, the RSD decreased, albeit not indefinitely. These results will allow forest managers to better understand how aspects of this variance estimation technique affect the accuracy of the uncertainty associated with parameter estimates. Utilising this information can provide forest managers with inventories of greater accuracy, therefore ensuring a more informed management decision. These results also add further weight to the applicability of the *k*-NN variance estimation technique in a range of forests landscapes.

REFERENCES

Maltamo, M., Eerik¨ainen, K., Pitk¨anen, J., Hyypp¨a, J., and Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning357laser altimetry and expected tree size distribution functions.Remote Sensing of Environment, 90(3):319–330.

McRoberts, R. E., Tomppo, E. O., Finley, A. O., and Heikkinen, J. (2007). Estimating areal means and variances of forest attributes using the375k-Nearest Neighbors technique and satellite imagery.Remote Sensing of Environment, 111(4):466–480

Morgan-Wall, T. (2020). rayshader: Create Maps and Visualize Data in 2D and 3D. R package version 0.19.2. https://CRAN.R-project.org/package=rayshader

Næsset, E. (1997a). Determination of mean tree height of forest stands using airborne laser scanner data.ISPRS Journal of Photogrammetry and377Remote Sensing, 52(2):49–56.

Rao, J. N. K. (2003).Small Area Estimation. Wiley, New Jersey, U.S.A, 1st edition

Woods, M., Lim, K., and Treitz, P. (2008). Predicting forest stand variables from LiDAR data in the Great Lakes - St. Lawrence forest of Ontario.398Forestry Chronicle, 84(6):827–839.