FUTURE FORESTS RESEARCH

TECHNICAL NOTE

Number: RTN-016
Date: June 2013

Nearest Neighbour Imputation of Stand Attributes using LiDAR

Summary

This technical note describes the implementation of the k nearest neighbour (kNN) approach across a 4,000 ha swath of Kaingaroa forest. Model evaluation showed that the models produced performed well at predicting a range of important stand metrics including total recoverable volume, mean top height and grade mix. The success of this case study suggests that the kNN approach may provide a useful means of integrating aerial LiDAR scanning data into the current forest yield information systems of a forest management company.

Authors: J Dash, H Marshall, B Rawley

Introduction

Aerial Light Detection and Ranging (LiDAR) has long been the subject of forest research seeking to take advantage of patterns in the LiDAR point cloud to extract information about forest structure. Although the strength of the relationship between various LiDAR metrics and key forest parameters has been recognised many times the use of LiDAR technology for resource assessment purposes has remained in the research sphere with examples of application to operational forestry very limited. There are several reasons which have restricted the uptake and implementation of LiDAR for resource assessment: the cost of LiDAR acquisition has been prohibitively expensive, the computational power required to handle LiDAR datasets is large and statistical techniques which effectively incorporate LiDAR data into current forest information management systems in a robust manner have not been available. This objective of this project was to develop an inventory system which could take advantage of LiDAR data and incorporate this information into the current yield prediction framework of a forest management company. With this objective in mind a case study was initiated in a 4000 hectare contiguous swath of Kaingaroa forest in the central North Island of New Zealand.

Research to date has focussed around regression sampling and modelling approaches to describe the relationship between LiDAR and forest parameters. There are a number of reasons why regression techniques may not be optimal for this purpose and so a statistical technique known as k-Nearest Neighbour (kNN) imputation was investigated. Under a kNN approach the forest parameters of a given patch of forest are assigned based on its similarity, in statistical terms, to a set of reference observations for which there is both LiDAR and ground measurements. kNN imputation has the following properties that make it a favourable technique for resource assessment purposes:

- It offers favourable integration with the current yield prediction framework of the majority of New Zealand's forest management companies;
- kNN has the ability to extrapolate a small number of reference plots to deliver precise information about a large number of stands;
- The technique is non-parametric and free from distributional assumptions.

The purpose of this technical note was to report on a case study undertaken to examine the utility of the kNN imputation approach using LiDAR data for forest inventory purposes. Development of sampling error estimates is also required and this was a key technical challenge of this project.

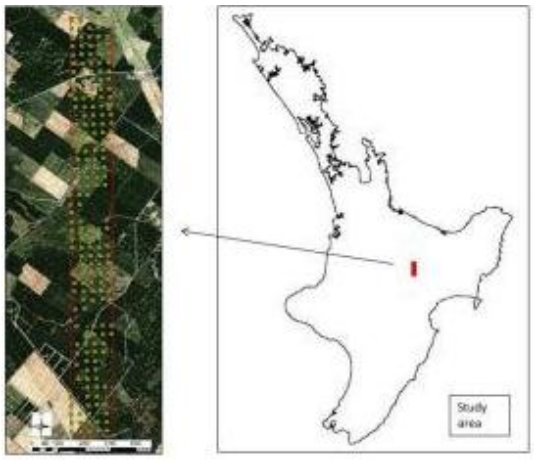


Figure 1. Trial study area. The left panel shows the 4000 ha study area (outlined in red) with installed plots shown as light yellow circles.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

Methodology

LiDAR data was acquired across the study area in winter 2012 using a fixed wing aircraft with a flying height of 950m above mean ground level. LiDAR data was acquired with a design pulse density per swath of minimum 4 pulses per square metre, and a swath overlap of 50%. This LiDAR data cloud was then characterised using the FUSION software product to produce 101 LiDAR metrics across the entire study area. Concurrently measurements were obtained from 213 circular, bounded field plots within the study area which allowed the calculation of forest parameters such as total recoverable volume (TRV), and stocking and also log product volumes as overlapping tree descriptions were recorded. At each ground plot a high grade survey GPS unit was used to fix the plot centre to sub 0.5m accuracy, these plot centres were used to produce LiDAR metrics for the part of the LiDAR point cloud that was exactly concurrent with the ground plot. This resulted in the production of two datasets one which contained ground plot measurements and LiDAR metrics, referred to as the reference dataset, and the other containing LiDAR metrics only at a 30m x 30m resolution, which is referred to as the target dataset.

The kNN technique uses patterns in the LiDAR data to identify which plot in the reference dataset is most similar to each pixel in the target dataset. The most similar reference plot is known as the nearest neighbour and the number of neighbours used as donors is referred to as k. Under a scenario of k=1 the single nearest neighbour from the reference dataset provides all the response variables (e.g. TRV, stocking, product mix) which had been recorded during ground plot measurement. The LiDAR metrics in both the target and the reference dataset are used to define the proximity of neighbours. 101 candidate predictor variables were produced as part of this study. An algorithm was developed to select only the most important ones and remove the unimportant ones. The algorithm used a technique called simulated annealing and resulted in the selection of 19 of the potential candidates variables for use in the modelling process. The selected variables were used to impute the desired response variables for every cell in the target dataset. In this manner the measurements recorded in the ground plots were extrapolated across the entire study area using the information derived from the LiDAR dataset.

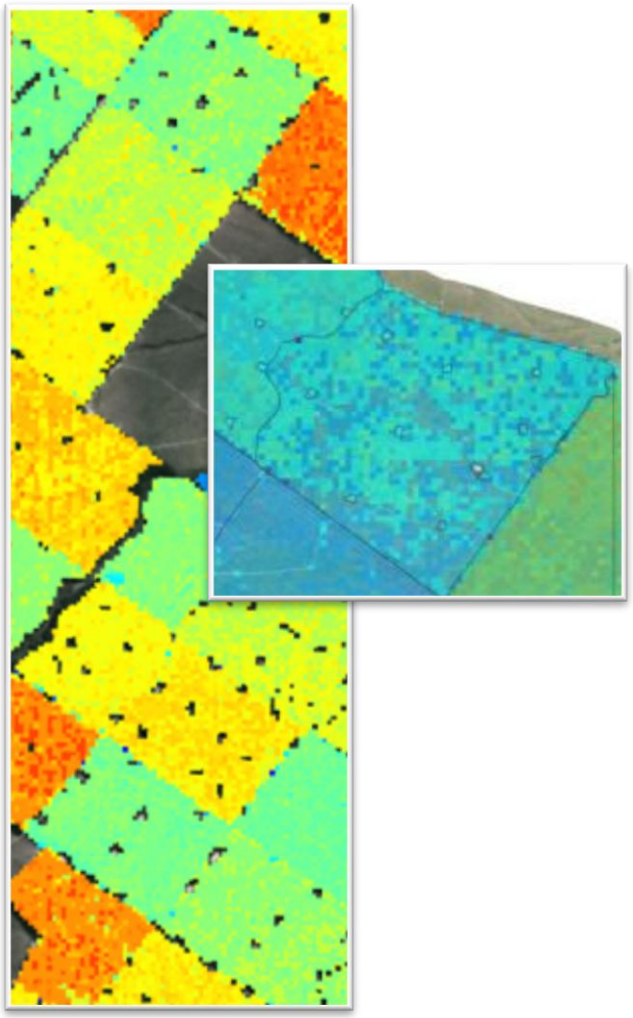


Figure 2. Raster of the target dataset with 30m resolution pixels with resulting response variable (in this case Top Height) for the study area and at a stand level.

Results

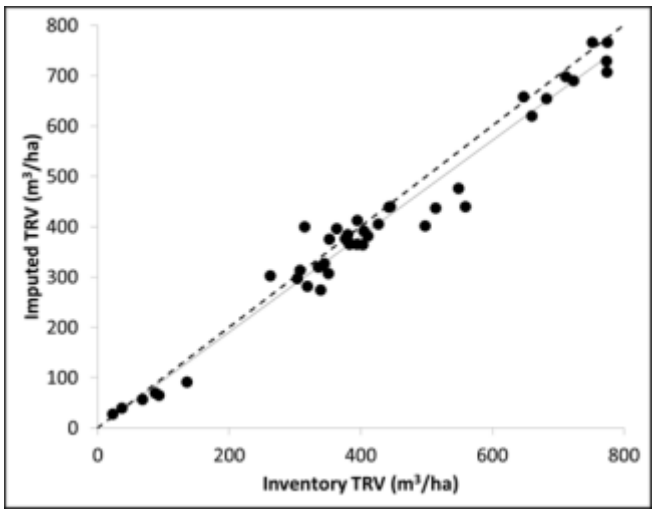
MODEL VALIDATION

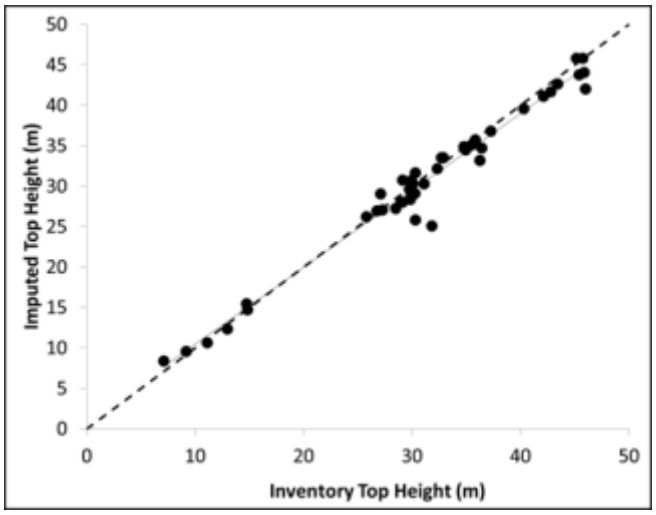
To provide a measure of the quality of the stand parameters produced pixels in the imputed surfaces were aggregated and averaged within the forest manager's stand boundaries in the study area. These were then compared with a validation dataset consisting of yield predictions from the forest manager's regular stand assessment and yield forecasting systems projected to LiDAR acquisition date. The results of the comparison with the validation dataset are shown in Figures 3 and 4. In Figure 3 each datum represents a stand in the study

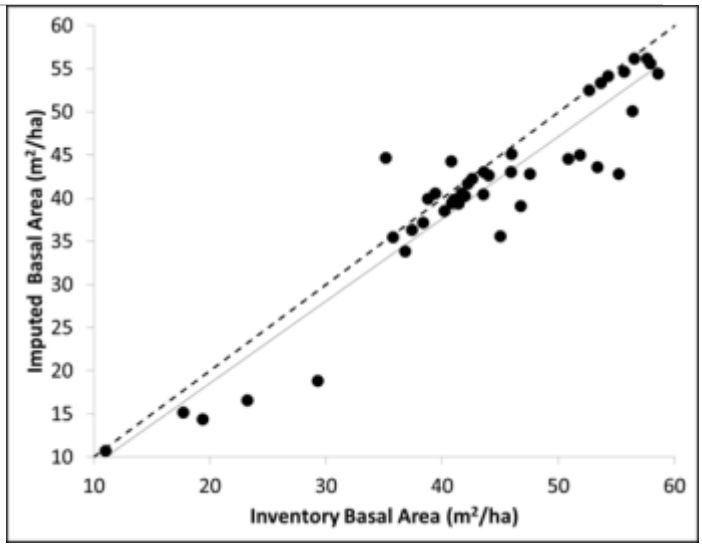
Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

area with a pre-existing stand inventory. The dashed line represents an unbiased correspondence between imputed and inventory values (1:1 line) and the solid line shows the linear relationship between imputed and inventory values.

Figure 3 indicates that there is a strong correlation between imputed and inventory values for TRV and top height for the majority of stands in the validation dataset. The correspondence between imputed and inventory values for basal area and stocking is somewhat weaker but acceptable for the intended use of this information in a forest management context.







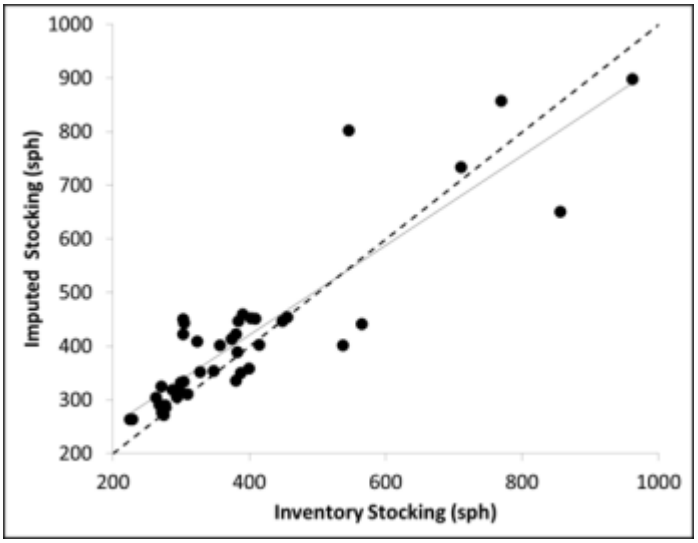


Figure 3. Relationship between the imputed and inventory values for several key forest parameters.

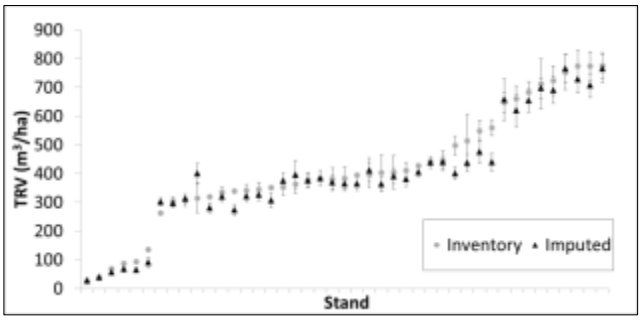


Figure 4. The imputed and traditional inventory TRV values and sampling error for validation stands in the study area.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

PRODUCT MIX

In the same manner the log product volume for any pixel in the target dataset can be imputed based on the log product volumes of the k nearest neighbours in the reference dataset. The imputed log product volumes for each pixel inside the forest manager's stand boundaries were aggregated and averaged to provide a comparison with the validation dataset. Figure 4 provides a summary of the average product mix as a proportion of TRV for all stands in the study area for which a conventional stand assessment was also available. This figure indicates that although there are some differences in the product mix produced the imputed product volumes are broadly consistent with those from the traditional stand assessments.

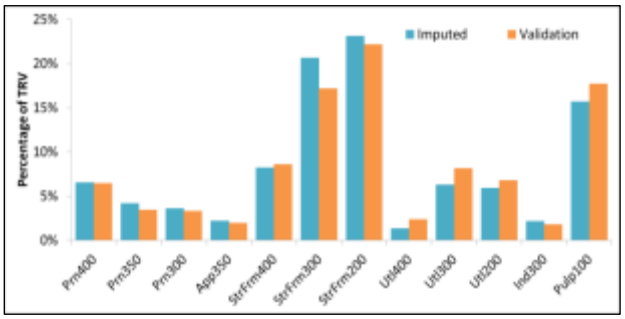


Figure 3. Comparison of imputed and grade mix and that derived from a traditional inventory for the validation stands. Values shown for each grade are averaged across all validation stands.

At a stand level the product mix was compared by multiplying log product volumes by a notional log price to get a easure of value ($/ha) as shown in Figure 6. The results of this analysis show that for the majority of stands there is good correspondence between the imputed and inventory (validation) product volumes. For some stands there are sizeable differences in product mix produced and this is due to a number of factors including unpruned stands acquiring product volumes from pruned reference plots. This is illogical and could be overcome in a number of ways in a production setting. Refitting the model to eliminate this was deemed beyond the scope of the current case study.

YEILD TABLE DEVELOPEMENT

A further objective of this case study was to integrate the kNN approach into the forest manager's current yield prediction framework and produce yield tables. Once a neighbour is selected under kNN the yield projections associated with the donor reference plot can be used to predict future forest conditions. The results of a comparison of the imputed and inventory projected TRV yield development (Figure 7) shows that there is a good correlation between the imputed and inventory values. The imputed yield predictions also show no bias when compared to the validation predictions. This exercise was designed as a proof of concept for the yield projection technique and some issues remain that will be addressed during a practical implementation of the technique.

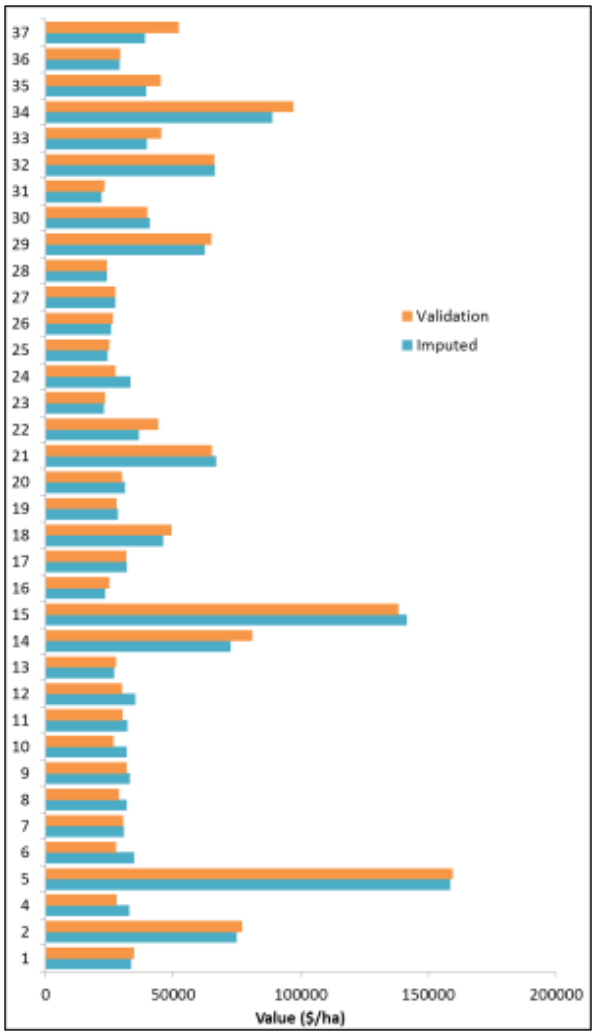


Figure 6. The notional value of stands in the validation dataset.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

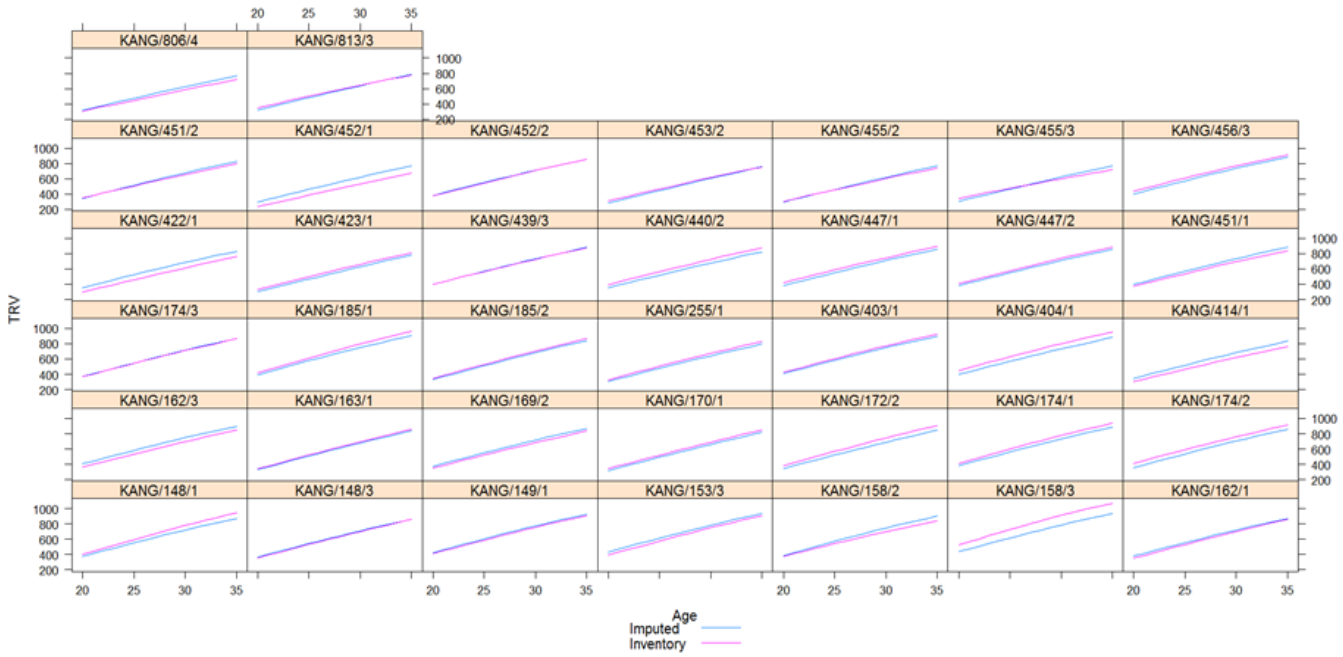


Figure 7. TRV development based on the imputed and the inventory datasets

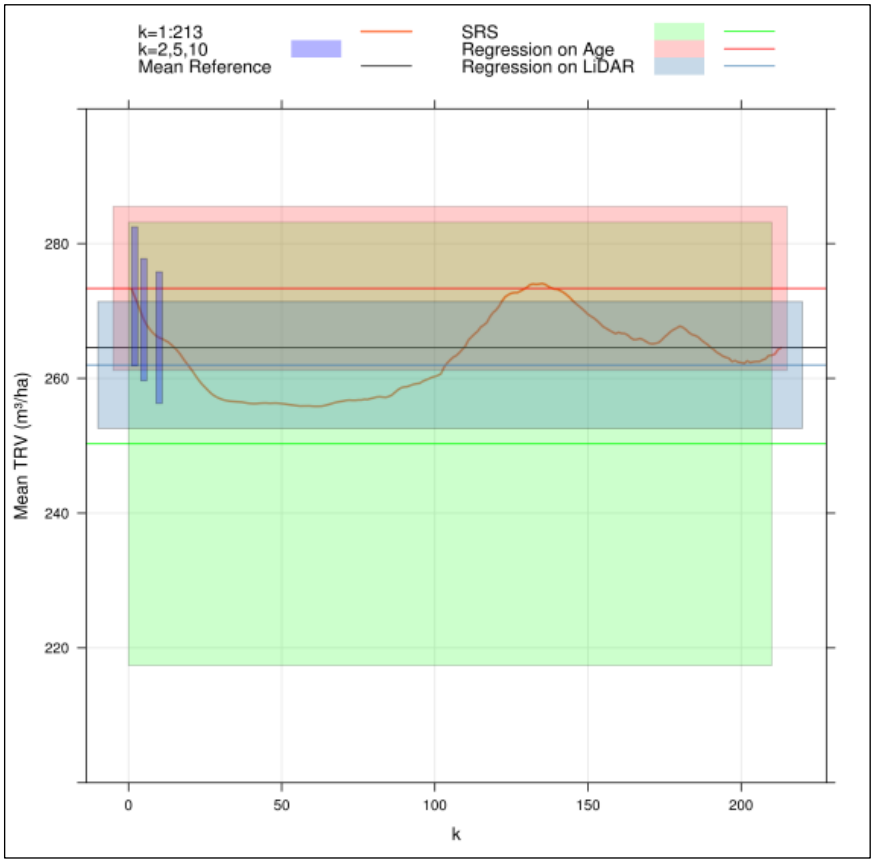


Figure 8. Estimates for TRV for the entire study area

SAMPLING ERROR

Calculating sampling error was one of the key technical challenges of this case study and the production of sampling error estimates of this type is the subject of on-going statistical research. Spatial correlation is the tendency for plots that are physically close together to be similar. If not adequately accounted for spatial correlation in the reference dataset can lead to inaccurate estimates of sampling error. The spatial correlation in the case study dataset has been explored and used in the calculation of sampling errors for all stands in the study area. This process is complex and computationally demanding. The estimates of TRV and sampling errors as produced by kNN, those reference plots established under a simple random sampling methodology, the mean of all reference plots and regression estimation approaches using Age and LiDAR as auxiliary variables are shown in Figure 8. This figure shows that there is no evidence for bias in kNN model predictions and provides confidence that the imputation model has been implemented correctly and is working well.

Critically the sampling error for any forest parameter can now be calculated for the kNN approach for any area of interest in the study area.

Figure 4 shows the imputed and traditional inventory TRV values and associated 95% confidence interval for the validation stands. There is a strong correspondence between kNN and inventory estimates of TRV. The kNN estimates of sampling error are smaller in most cases. The median kNN confidence interval for stands in the validation dataset is 27.9m3/ha compared to 37.89m3/ha from the traditional inventory. This result highlights the ability of the kNN approach to provide accurate and precise estimates of stand parameters for many stands from a small number (213) of reference plots.

CONCLUSION

This case study has implemented a new and innovative inventory technique for New Zealand that demonstrates great potential. The kNN technique can integrate LiDAR data and use it to provide accurate estimates of forest parameters at assessment date or at a desired point in time. This can be achieved within the limitations and scope of the current yield prediction framework of the majority of forest managers in New Zealand. Estimates of stand parameters are free from bias and appear to be working very well. Product mix can also be derived for any area of interest and although this appears to be broadly accurate additional work is required for a practical implementation of this approach. The sampling error for forest parameters in any area of interest within the study area can now be calculated. The sampling errors are comparable with a validation dataset derived from traditional, intensive, stand assessment procedures despite the small number of plots used to produce the kNN estimates.

The advantages of this technique in providing better resource assessment data for forest information systems and reducing costs through replacing some component of traditional forest inventory are significant. This case study has shown for the first time in New Zealand a complete stand level integration of LiDAR data into an industrial forest information system using the described yield prediction framework that can readily produce accurate and precise results.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz