ISSN 1946-7664. MCFNS 2018
Vol. 10, Issue 1, pp. 6–8
Mathematical and Computational Forestry & Natural-Resource Sciences

AVAILABLE ONLINE AT HTTP://MCFNS.COM

Submitted: Mar. 5, 2018
Accepted: Mar. 14, 2018
Published: Mar. 30, 2018
Last Correction: Mar. 14, 2018

ARE MODELS THE ANSWER?

Kim Iles
Kim Iles & Associates Ltd.
412 Valley Place, Nanaimo, BC, V9R 6A6, Canada

Abstract. This note gives a short discussion about using models with or without correction by field measurements in forest inventory. While modelling with LiDAR may be tempting, the author would argue that these inherently limited models are best used for their ability to distribute overall totals after the fact, or to assist a more general unbiased inventory approach using models as an initial estimate.

Keywords: LiDAR; Forest Inventory; Growth&Yield modeling; models; sampling; growth measurements; virtual timber cruising;

The Discussion

I think not. This is particularly the case with LiDAR, and it will probably be the case with the next remote sensing method that comes along. I clearly remember a recent response from a friend when I asked how they measured growth in his country. His response was "we don't measure growth – we model it". Measurement, you see, is so messy. It involves not only field work, weather and variability - but time, expense, and pain. Models offer such precision and comfort in comparison to the messy processes of sampling and measurement. We already hear talk around the coffee pots at conferences about "virtual timber cruises" to be made up of odd bits of data combined in some way in order to avoid venturing into the forest, as if our forests were the dangerous void of space. The fact that these synthetic processes are almost indecipherable only seems to add to the anticipation of their potential performance.

There is, of course, no option to modeling when it comes to some situations. For combinations were we have no evidence, and no option for verification, models are an obvious choice, and a good one if their potential error is reasonably assessed. Some silviculture combinations are a clear example of this. Interpolation is a modeling strength, extrapolation is generally its weakness.

The recent rise of LiDAR is a modeling case of interest, and a method likely to fail for quite some time before it becomes thoughtful. New technology often does this. GIS technology is an example of many years being spent before someone came to their senses, stopped "doing GIS" and started "displaying and analyzing data". At present, the halls of upper management seem filled with people who want do "do LiDAR" rather than doing forest inventory. The fact that enormous numbers of data points are used, manipulated by algorithms (often unknown, indecipherable or proprietary), seems to indicate that the true experts in the process must be contractors who "do LiDAR" on the bankroll of the company. No amount of failure or budget overrun is considered an example of a mistake in this approach. Unpublished failures can certainly be found, many struggling to look like a success.

LiDAR, of course, is a wonderful opportunity, just as any of the remote sensing methods might be. The complete coverage it offers, along with automatic computer processing and engineering applications on the side, is very appealing and essentially a new opportunity. It cannot, however, measure forests. It takes no depth of thought to determine this. Regardless of this fact, reams of data are often collected, on large fixed plots, in questionable locations from a sampling point of view, in order to produce a "model" to apply to the remaining LiDAR data in the forest.

The problems of over-fitting and the inadequacy of this data to produce the wide range of results needed by a forest inventory are ignored. Often, so many of these plots are produced to "calibrate the LiDAR model" that they would provide a reasonably good overall answer by themselves. In such cases, the forest inventory total will change very little. What, then, is the advantage of LiDAR? Certainly it must be in solving the most pertinent issue of the last half a century – not "how much do you have", but "where is it". Better location of the total resource has been the practical problem for many

decades in forest inventory, but it is seldom indicated in sample statistics or addressed in the nearly unreadable papers produced by biometrics specialists. Getting total volume has not been the essential issue for many years. Distribution is.

Model calibration errors, as well as sampling errors, become biases when a model is applied directly. This is often forgotten by modelers, and unknown to many model users. The usual model applied to a set of data (LiDAR and otherwise) is filled with assumptions, and that can also be true with many of the measurements used to calibrate it. If these issues are brought out in court, the numbers of assumptions and faith-filled assurances that they "probably will not matter" might be enough to convince a court that people who rely on these models fall into the realm of true believers, and might not be thought of as having much depth.

There is a way to overcome this problem. There is a great deal of preparatory work in making a model where the outcome of its application *depends* upon it being unbiased. This can be replaced by a sampling process that is unbiased and corrects a flawed model which simply *assists* in the inventory process. I am avoiding the use of "model dependant" and "model assisted" as terms because they are so impenetrable, if not muddled, in the current literature.

The current sequence used by so many inventories [data/model/application] is perhaps not the best approach. Only decades of failure will prove this, but not everyone has missed the idea of [data?/model/application/*correction*] where the questionable model is applied over the landscape in the hope of a good relative distribution, and then corrected by sampling to eliminate the biases of the model. New data may not be needed to support an initial model. Why, after all, should the adjacent valley be essentially different and require a new set of data to calibrate the LiDAR data? Should the previous model not apply? Perhaps not, if it included serious overfitting.

Gathering a new set of data each time will disguise the overfitting, but not eliminate any biases it creates. It is a small change, but an important one, to shift the serious data gathering to the end of the process. That lets LiDAR have free reign to make correlated estimates with no necessity for precision or unbiasedness. LiDAR can then do what is needed most, and perhaps what it is best suited for – the distribution of volumes, species, and other items across the landscape. The idea of all stands in a strata having the same description is long outdated, and with the advantages of LiDAR it is not necessary to have the same density inside a stand.

In my opinion, the best role of LiDAR is not "what do we have", but "where is it located". In many cases, we already have a more than adequate inventory total – but LiDAR and other data could much improve the distribution of such an unbiased total. Accuracy is not necessary in LiDAR predictions to get the full benefit of the largest advantages of LiDAR – full coverage, automatic processing, and positioning on the landscape. These are practically new to the remote sensing world, and a great advantage to forest inventory. Sampling theory has lots of ways to measure and correct initial estimates. Good estimates from the model would greatly improve the sampling error, of course. Additional estimates can be provided by other means for many stand variables, rather than a LiDAR model produced from calibration plots.

I personally believe that LiDAR will be widely and cheaply available in the near future. For one reason, the engineering aspects are certainly more valuable than air photos, which are widely available at present. Use of the technology in self-driving cars and archeology applications will drive the cost down very rapidly. I believe that the vertical data alone is not the critical advantage, but rather the highly automated and very complete nature of the process. Forestry is a sideline for LiDAR, and we do not seem to have been thoughtful about the overall application of this powerful technology. If LiDAR "lumps" that might be trees (or clusters of them) can be accurately located in the field for measurement, something like Sorted List Sampling offers real potential, and the mapping aspect of the process for trees or clumps has great potential on its own.

There are, of course, similarities to other processes in forestry where we have failed to apply this opportunity. When we do check cruising, are the initial results adjusted by the check cruise results? Why not? Are the forest model predictions for large areas adjusted by the periodic measurement of the forests with temporary or permanent sample plots? How often do we hear about such comparisons?

British Columbia has built a correction system into their tree volume equations. Questionable sources and treatments of data were used to produce the "tree taper model" equations years ago. An actual sample of trees across the Provincial land base is gathered to check these volume equations, so corrections can be applied when necessary; including to sub-regions where the overall equations cannot be expected to apply directly. The basic tree form geometric model, as an estimate, does an excellent job overall and can be made unbiased by a few actual and careful measurements from a real sample.

Such an approach has only one consistent criticism. "Don't the small number of measurements for the correction process increase the variance?". The answer is simple. "Not if you don't choose to recognize it" is the baited reply. When the predictable reaction of "well you simply must consider that" is given, the response is "I

suppose so – how do you include the variability and bias from your old tree volume equation in your current sampling process?".

Models as estimates make great sense. If they are reasonable, they also provide better efficiency to a sampling process that will correct the biases that assumptions and other errors cause. Models as the assumed truth are often not necessary, seldom wise, and full of problems that are easily removed by a sample after the model is used as an estimate. LiDAR is a good current example of this issue. Distribution of inventory results over the land base is important, even if it produces no direct information about the total to be distributed. I believe that the progress of sampling methods over the last several decades has a clear central message –sample to correct an initial estimate – not to create a model. Those who are trying to do it that way need more company.