Imputation Techniques for Modelling Natural Regeneration in Complex Stands

A Workshop

Organized by

Dr. Abdel Azim Zumrawi, Research Branch, BC Ministry of Forests

&

Dr. Valerie LeMay, Department of Forest Resources Management, UBC

Sponsored by The Research Branch of the BC Ministry of Forests, UBC

Funded by

Forestry Innovation Investment

Held at the

Faculty of Forestry, University Of British Columbia

December 17, 2002

Table of Contents

Table of Contents ii

Background 1

Overall FII Project Objectives 1

Workshop Objectives 2

Workshop Agenda 3

Workshop Participants 4

Presentations 5

Issues Raised in Presentations 6

Summary of Discussion Points 8

Background

The Prognosis^{BC} model was adapted from the USDA's Forest Vegetation Simulator (FVS), Northern Idaho (NI) variant, for use in the southeastern portion of British Columbia. FVS was selected because of its ability to model multi-age, multi species stands, similar to those commonly found in southeastern BC. The regeneration estimation component of the original model has been "turned off" because it was found to be an inadequate predictor of regeneration under BC conditions. Users of Prognosis^{BC} must presently supply their own estimates of regeneration.

Early work on calibration of the regeneration component in the ICH, Nelson Region, resulted in improvements over the pre-existing FVS NI version of the model, but because of the small sample size, lacked robustness. Imputation approaches have been tested to examine their ability to predict regeneration. Most similar neighbour techniques have been found to be more effective than tabular imputation approaches. However, the ability to accurately predict regeneration has remained problematic.

Overall FII Project Objectives

The primary objective of this project is to develop a natural regeneration model for Prognosis^{BC}. The model will be based on the imputation techniques developed, and data gathered by researchers at the University of British Columbia, Faculty of Forestry. The programs and databases will be sufficiently flexible such that growth-and-yield models other than Prognosis^{BC} may use them.

The specific objectives of this work are:

1. Define the specifications and standards for using imputation techniques to simulate natural regeneration in PrognosisBC.
2. Create the necessary databases (from existing collected data) to implement a Nearest Neighbour imputation method in predicting regeneration.
3. Develop the programming and linkages to the PrognosisBC core model and interface.

4. Create a prototype model for further testing and evaluation.
5. Document the standards, methodology, programming and databases used in 1-4 above.

Workshop Objectives

The Research Branch of the BC Ministry of Forests in cooperation with the Department of Forest Resources Management at UBC organized a 1-day workshop on modelling natural regeneration in complex stands.

The purpose of this meeting was to discuss the recent work on the regeneration component of the Prognosis^{BC} model at the University of British Columbia. A small focus group of selected participants was invited to discuss potential future directions of this research. Included in this mandate is the subject of how the regeneration prediction component should and does fit within the functioning of the overall Prognosis^{BC}. The objectives of the workshop were:

1. Initiate a discussion of the innovative use of imputation techniques in natural regeneration modelling.

2. Help define broader guidelines on how the Most Similar Neighbour (MSN) could be used to estimate natural regeneration in connection with the Prognosis^{BC} growth and yield model.

Workshop Agenda

9:00 – 9:10	Introduction. Valerie LeMay
9:10 – 9:30	Prognosis BC : Past and Current Developments. Abdel-Azim Zumrawi
9:30 – 10:00	Modelling Regeneration by Imputation: Some background. Albert Stage
10:00 – 10:30	Coffee Break
10:30 – 11:00	Regeneration Imputation Models for Interior Cedar Hemlock Stands. Valerie LeMay
11:00 – 11:30	Regeneration Prediction for PrognosisBC, IDFdm2 subzone variant, Invermere Forest District. Katrina Froese
11:30 – 12:00	Strategies on implementing an imputation based regeneration model. Nick Crookston
12:00 – 1:00	Lunch (provided)
1:00 – 2:45	Discussion: Round I
2:45 – 3:00	Coffee Break
3:00 – 4:00	Discussion: Round II
4:00 – 4:30	Wrap-up

Workshop Participants

Valerie LeMay, Dept. of For. Res. Mgmt., UBC Organizer/Facilitator	Deb Delong Forest Science, UBC Participant	Don Robinson ESSA Technologies Participant
Abdel-Azim Zumrawi MOF, Research Branch Organizer/Presenter	Bruce Larson, Dept. of Forest Science, UBC Participant	Steve Stern-Smith SIGY Participant
Katrina Froese UBC, For. Res. Mgmt. Presenter/Web site	Albert Stage, US Forest Service (Retired) Presenter	Robert Keefe Univ. of Idaho Participant
Cornel Lencar UBC, Wood Science Audio/Video	Nick Crookston US Forest Service Presenter	Roberta Parish MOF, Research Branch Participant
Rene Martin Consultant Recorder	Catherine Bealle-Statland, MOF, Research Branch Participant	
Jim Goudie MOF, Research Branch Participant	H. Temesgen For. Res. Mgmt., UBC Participant	

Presentations[1]

Prognosis^{BC}: Past and current development.
Abdel Azim Zumrawi

Regeneration Imputation Models for Interior Cedar hemlock Stands.
Valerie LeMay

Modelling regeneration by Imputation: Some background.
Albert Stage

Regeneration Prediction for Prognosis^{BC}, IDFdm2 subzone variant, Invermere Forest District.
Katrina Froese

Strategies on implementing an imputation based regeneration model.
Nick Crookston

[1] Presentations appear on the website http:// www.forestry.ubc.ca/prognosis/meeting.html

Issues Raised in Presentations

Presentation by Albert Stage

As a result of a discussion with the Prognosis[BC] team in 1997, a way in which a BC variant of Prognosis could be modelled was suggested:

- Develop program first
- Collect data subsequently
- Use lookup tables rather than regression
- Data can be augmented in future; preferable to fitted models.
- This approach is preferable based on ease of updating and the ability to retain variability.

Using a stratum-based random sample to impute regeneration was thought to be best. Using Nearest Neighbour (NN) or k-NN. It is suggested to use most similar neighbour (MSN) or other measure of similarity to choose reference plots. Problems with these methods include:

- Time index
  - Time since disturbance: how is it defined, what is considered a disturbance?
  - Changes in overstory?
- Choice of site and overstory characteristics?
- Degree of aggregation – point or stand?
- Statistical criteria:
  - Some measure of similarity
    - Form of a distance function (Euclidian, Mahalanobis?)
    - Transformation of variables
  - Weights for Euclidian distance (or other measure)
  - Weights for k-NN

Imputation errors:

- Minimised by wide distribution of target and reference points
- Depend on X's selected, transformations used, and density of reference data
- Depend on choice of similarity measures, Y selection, and Y transformation

- Sources of error include:
    - Distance
    - Measurement error – in earlier models was assumed to be zero. Now taken into account, with error from both target and reference data
- Nonlinearity – canonical correlation analysis assumes linearity between X's and Y's
- Sampling: need to spread sample data out to account for extremes so data are not biased to centre

<u>Presentation by Nick Crookston</u>

MSN version 2.0 is currently available, with new features being added. Currently, Forest Vegetation Simulator (FVS) inputs from databases or ASCII files; however, working on allowing it to read and add to a general database. The process for FVS is as follows:

- Do over stands
    - Input
        - Do over cycles (5-10 years)
        - Thinning/harvesting
        - Growth/mortality
        - Insects/disease/fire
        - Update state of each tree
        - Predict establishment*
        - Reporting of variables (summarise individual tree prediction, extrapolate to stand)
        - Next cycle
    - Next Stand

*The establishment model currently being used is that of Fergusen and Carlson (1993) of the Rocky Mountain Research Station, Moscow, Idaho. This could be replaced by bringing in establishment from some outside source i.e. imputation; however, we need to define what to read?

Summary of Discussion Points

Time Index

Participants agreed that using time since disturbance, as an index in imputation modelling is necessary. Also, the majority of disturbances can be clearly defined, including clear- and partial-cutting, fire, windthrow. It is essential for users to clearly state their definition of disturbance. For regeneration data gathered, a range of disturbances due to partial cutting was used in setting up the sampling grid.

Choice of Site and Overstory Attributes

It was mentioned that we do not know what changes will be made to data list in future, so it was thought that keeping track of all data was prudent – better to measure too much than too little. The drawback to this is that data collection is often limited by funding availability.

The availability of X variables may result in different measures of similarity. For example, if there are few Y variables compared to X, using Mahalanobis distance might be preferable over distance measurements that use canonical correlations between known variables from both the target and reference plots and those that are only known for the reference plots.

Temesgen mentioned that he would like to see patterns of retention (i.e. degree of clumpiness) and percent basal area (BA) retained be added as predictor variables. Percent BA information may come from pre-harvest silviculture plan data or from stump measures; however, we may not be able to discern which trees were dead at time of harvest. A clumpiness index may be derived from aerial data, or may be on a polygon label in the future. It was also suggested that for percent BA remaining, trees should be separated into height classes, as small trees may block more light on the ground than tall trees.

The knowledge of species composition and size of trees prior to disturbance may be a good indicator of advanced regeneration. This may also facilitate prediction of absence, which is as important as predicting presence. However, these data are challenging to obtain. We would need to know if the pre-harvest inventory excluded trees below a certain diameter at breast height (dbh) as this may affect predictions of advanced regeneration. Some issues surrounding early growth predictions include:

- The need for spatial data;
- The possible need for repeated measures over time;
- The need for data and interpretation that leads to guidelines for planting/spacing requirements;
- The need to determine green up/free to grow likely for different levels of partial cutting; and
- Need more confidence in natural regeneration success.

Knowledge of BA removal as well as BA retained, may be important, as it can indicate site carrying capacity (productivity).

Degree of Aggregation (Plot, stand, or Groups of Stands)

It is important for both imputation and projection to define what level the input data is at (e.g. point, stand) for every variable, as levels may vary depending on the variable. In general, it was thought that point-to-point data collection was best; the projections are then aggregated to allow for spatial heterogeneity. If the level of aggregation is too wide, extremes are lost due to averaging. The aggregation should be run at the level of inventory; for example, if 10 plots are measured, each plot should be projected, and the resulting projected summarized over the 10 plots. In the case of sub-plots, the sub-plots should be aggregated prior to projections, but the sub-plot identity should be retained as this gives information on local spatial variability.

Measure of Similarities and Weights

It was thought that users should have a choice of which measures and weights they wanted to use in nearest neighbour analyses including:

- Selecting randomly from a bin (stratum);
- Selecting an average from a bin;
- Using NN (one observation) vs. k-NN (average of k similar observations);

- Using MSN similarity measure vs. other measures – there may be no discernable difference between using absolute rather than squared differences; and
- For k-NN, instead of averaging the neighbors, appending them and just changing the area associated (trees per ha) with each tree in the list.

<u>Time Inconsistencies</u>

It was thought that after five years since disturbance, all regenerated trees should be established and this could be used as an input to a Prognosis projection. However, some data shows that additional trees come in later. This could be resolved by obtaining a chronosequence of data or using existing chronosequences. Challenges with this approach include:

- Data points are discrete times in current data and points in time don't necessarily connect. Could "patch" time sequences together, but would need to smooth them out over time for consistency; and
- Multiple-species may make smoothing data more challenging, as the best reference site for a single species target may include many species. Temesgen suggested stratifying by species might resolve this;
- Seed source can also affect predictions of establishment;
- Participants thought that keeping track of lagged variables (current and previous time measures) would help in deciding whether additional regenerated trees need to be added for the next time segment. If a high degree of resolution is not required, only the "best trees" (ones likely to survive based on some definition) could be brought forward and incorporated into the tree list. The remainder of the trees (i.e. not the best trees) could be clumped together as an additional variable affecting early establishment (additional competition in early years); and
- When generating tree lists for the current year, we would need to incorporate projections to current year based on regeneration input at year 5 and the current year regeneration from the sample data, in some consistent way.

Work by Ek *et al.* (1997) [2] showed that regeneration "stabilized" after 15 to 20 years. Rather than deal with projections from 5 years after disturbance, the amount of regeneration 20 + years after disturbance could be used as input for model projections. This would avoid the odd jumps and inconsistencies in the data, particularly in the early years where growth is very variable. Using MSN will incorporate and reflect this variability, whereas regression will be affected by this high degree of variability in the data.

It was suggested that to resolve time inconsistencies, an alternative to projecting from 20 years since disturbance is that we could start with the 5-year regeneration list (data 5 years following disturbance) to complete the tree list for the entire diameter distribution (stems/ha by diameter) and project the tree list forward in time (e.g., 20 years from disturbance). This projection could be compared to the regeneration list from the sample data for the 20 years since disturbance. These two estimates of the diameter distributions pictures could then be resolved by determining which trees in the 20- years since disturbance regeneration data came in over the 15 year projection period (from 5 to 20 years). This could be achieved by keeping track of age of regenerated trees in the model. However, in the regeneration sample datasets, age of regenerated trees is not measured as this would be too time consuming. Another way this could possibly be handled is by truncating the diameter distribution for the projected tree list and then adding in regeneration from the sampling data for 20-years from disturbance. This could result in species composition inconsistencies over time.

Spatial Measurements

Participants thought it might be useful to have overstory values in concordance with the sub-plots, rather than have them the same for all regeneration sub-plots. Bealle-Statland discussed how spatial measurements are incorporated into the Tree and Stand Simulator (TASS) model. For TASS, data for regeneration is based on large 1 ha stem mapped tree areas, with sub-sampling (400 – 900 m^2 plots) of

[2] Ek, A.R., A.P. Robinson, P.J. Radtke and D.K. Walters. 1997. Development and testing of regeneration imputation models for forests in Minnesota. For. Ecol. & Mgmt. 94: 129-140.

regeneration. The minimum tree height is 15 cm, and all regeneration is stem mapped in the sub-samples; this is based on the idea that seedlings are likely to survive if they reach 15 cm in height. It was thought that MSN could be used to get regeneration species per hectare by species and then arrange these spatially based on outside knowledge.

The spatial variables that are useful depend on the ecosystem studied. For example, in some areas, openings may be needed to promote regeneration; in others, such as dry sites, shelter may be needed. Soil compaction from harvesting machinery may prevent regeneration in open sites.

Bealle-Statland sited Melinda Moeur's PhD dissertation (University of Washington) as an example of using MSN to provide detail for sub-plots; however, regeneration was not sampled.

Actions and Indicators – Meeting Needs for Model Uses?

Currently, model use is motivated by need for information on partially cut stands. It may also be used for other disturbances such as fire and wind throw, and possibly root disease. A question was raised as to whether this would be used as input for other management practices such as brush control/clearing/vegetation mapping. If so, there are some design constraints that would need to be addressed.

Indicators that would trigger actions include:

- Assessing partial cutting outcomes. One of the procedures often used for partial cutting is to shift yield curve back in age;
- Assessing forest health issues outcomes;
- Assessing interactions between partial cutting and forest health issues;
- Needs for models for silviculture gaming;
- Needs to link models to inventory for timber supply. Used for timber supply in TFL 14 by Timberline. Difficulty was that it was not calibrated – hard to describe a polygon with different levels of partial cutting with one projection curve in timber supply analysis;

- Needs for information for habitat supply models that are based on tree attributes; and
- Need to track "snags" (non-spatially only) and fire changes

Truncated Dbh Distribution

One of the current problems with data is that the lower diameter (dbh) limit of the tree data is higher than the upper dbh limit of the regeneration data, leading to gaps in information on dbh distribution (not so with all data). If the tree data are projected forward without the regeneration data, an even wider gap between the lower limit of the tree data (projected to a future time) and the upper limit of the regeneration data comes in; therefore, the tree list must be "filled in" as soon as possible in the projection. Information about the minimum dbh tallied must be added into the FVS/Prognosis model, and the projection used must be as flexible as possible to account for the change in the lower dbh limits for the tree data.