**PROJECT ID:** CAFS.20.83

**PROJECT TITLE:** Using predictive analytics to decompose site index

**INVESTIGATOR(S):** Jason Cross, University of Washington

Eric Turnblom, University of Washington

|  |
| --- |
| **PROJECT DESCRIPTION:**  Site Index (SI) is a heuristic device that describes the rate at which a location produces wood. It is the convolution of many and varied dimensions – elevation; topography; soil composition and depth; precipitation; daily/monthly/seasonal/annual temperature pattern; and seed source – into the expected height of dominant trees at a given base age. Site index likely differs by species at the same location and differs by location for the same species. It is an input in various growth and yield models, whose outputs support millions of dollars’ worth of silvicultural decisions that include planting, pre-commercial thinning, and fertilization.  Increased quantities of atmospheric carbon dioxide are contributing to changes in the observed ranges of factors once thought to be fixed when index values were conceived (e.g., theoretical maximum photosynthesis rate, precipitation, minimum and maximum temperatures). At a given location, these changes may combine to effectively alter it (by increasing or decreasing site index from its historic value), and threatens the accuracy of the index, even though it may continue to be measured with sufficient precision. Accounting for the effects of climate change requires investigation into the decomposition of site index into its additive subcomponents. |
| **HYPOTHESES or OBJECTIVES:**  Hypothesis: site quality measured as top height predicted by breast-height age may be defined by a specific parameterized non-linear function (modified Weibull cumulative distribution function). The asymptote, rate, and shape parameters can be modeled as a function of several biologic, geologic, geographic, physiographic, and climatologic variables.  The objective of this project is to verify and validate (elements of) growth & yield models, and improve their parameterization. Both have been identified as priorities in the most recent CAFS/IAB member survey. In particular, to build a direct model of site productivity that can be evaluated to provide the indirect measure of site index at a given base age. |
| **METHODS:**  Data was drawn from the Stand Management Cooperative’s database: 2350 plot measurements on 406 plots. Critical measurements include breast-height age and top height, along with geologic, topographic, physiographic, and climatic attributes associated with plot locations thought to affect site index. A large matrix of attributes by plot were standardized, and then its Principal Components were calculated. The proportion of total variance explained by each individual variable was computed and ranked. The top 10 variables were retained for modeling.  A customized implementation of the Nelder-Mead method for non-linear parameter estimation was coded in FORTRAN 90; this version allows for any number of variables (discrete, coded, continuous) to be added as predictors for each of the asymptote, rate, and shape parameters for the Modified Weibull cumulative distribution function. Parameters are estimated as a linear function of pooled variables.  Upon convergence, the dataset was bootstrapped (sampled with replacement) K times, and refit with a looser tolerance (0.01 vs 0.0001 difference in MSE between iterations). At K = 200, the distribution of parameter estimates was evaluated, and the least-significant predictor in one of the parameter predictor pools was removed, and the process was repeated until all parameter predictors had a p-value < 0.10. |
| **MAJOR FINDINGS:**  Top Height predicted by breast-height age using a Modified Weibull CDF with 95% R-squared, a prediction error of 4.3’, and a mean average percent error of 10%. Important predictors include:  Asymptote: Square of planting density  Rate: summer degree-days above 18°C; spring relative humidity; autumn Hogg’s climate moisture index; winter Hogg’s climate moisture index; elevation; spring precipitation; physiographic region; planting dominated by mix of TSHE and PSME; planting dominated by TSHE.  Shape: summer degree-days above 18°C; autumn Hogg’s climate moisture index; elevation; spring precipitation; physiographic region; planting dominated by mix of TSHE and PSME; planting dominated by TSHE, planting density. |
| **DELIVERABLES:**   1. Working paper to be delivered to SMC membership detailing data, methods, results. 2. Tool (Excel spreadsheet + FORTRAN.dll) for determining relative importance of variables within a dataset; can be used for other growth & yield modeling efforts. Tool (FORTRAN program) for fitting non-linear model where parameters are linear function of any number of variables. 3. Outward-facing web application to predict top-height using a point-and-click mapping program, requiring planting density as an input, alternately accepting age and trees per acre. Range is limited to western Oregon and Washington. Website to be hosted by SMC for its membership. |
| **MEMBER COMPANY BENEFITS:**  Quick site index lookup tool can be used region-wide, not limited to a specific ownership.  Methods for identifying variables and fitting models may be applied to most any growth & yield metric of interest. Methods are scalable, and easily replicated across regions. After initial models are fit, they are easily updated as new data and variables become available with much shorter turnaround times. |