

A model evaluation framework applied to the Forest Vegetation Simulator (FVS) in Colorado and Wyoming lodgepole pine forests



Benjamin A. Bagdon^{a,*}, Trung H. Nguyen^a, Anthony Vorster^a, Keith Paustian^{a,b}, John L. Field^a

^a Natural Resource Ecology Laboratory, Department 1499, Colorado State University, Fort Collins, CO 80523, United States

^b Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, United States

ARTICLE INFO

Keywords:

Forest Vegetation Simulator (FVS)
Model evaluation
Sensitivity analysis
Equivalence testing
Verification
Validation

ABSTRACT

The Forest Vegetation Simulator (FVS) growth and yield model is widely used throughout the United States, but recent studies have reported unexpectedly large bias for some regional model variants. Here we propose a general framework for model evaluation, designed to highlight model strengths and weaknesses and inform calibration efforts. We apply the framework to evaluate the Lodgepole Pine (LP) model of the FVS Central Rockies Variant (FVS-CR), which has rarely been evaluated in the literature despite its widespread use in the western US. We started with a qualitative verification of the structure and logic of the FVS-CR LP model against a modified Bakuzis matrix, determining that it adequately reproduces known patterns of stand dynamics. We then compared stand-level growth simulations to a chronosequence developed from 554 Forest Inventory and Analysis (FIA) plots measured in Colorado and Wyoming lodgepole pine forests. This quantitative validation exercise revealed that the default model settings substantially over-predict basal area and total stand carbon after 50 years for both pure lodgepole pine stands and mixed lodgepole pine stands containing a minor proportion of Engelmann spruce, subalpine fir, and quaking aspen. Using equivalence testing to validate the large-tree diameter increment model against a separate dataset of 3,790 remeasured trees on 124 lodgepole pine FIA plots, we found that the default FVS-CR LP model adequately predicts lodgepole diameter growth, though we observed large variation in model errors. Equivalence tests also revealed systematic under-estimation of spruce-fir and over-estimation of aspen productivity. Finally, we conducted a sensitivity analysis to identify the most important model parameters and data inputs driving simulated stand structure and carbon accumulation in both the short- (50 years) and long-term (200 years). The model shows the greatest sensitivity to initial species composition; a small proportion of aspen, spruce, or fir seedlings led to long-term stand re-structuring and greatly increased carbon accumulation. Other sensitive parameters included site index, the fixed large-tree diameter increment adjustment factor, and the parameter controlling the point at which density-dependent self-thinning begins (maximum stand density index). This evaluation leads us to conclude that the FVS-CR LP model tends to under-estimate tree mortality (particularly in young stands) and the default site index represents relatively high productivity stands, such that re-calibration may often be necessary to capture realistic long-term stand behaviors. Our sensitivity analysis provides guidance for future efforts to re-calibrate the FVS-CR LP model and highlights the importance of collecting site index and seedling species composition data wherever possible to produce the most realistic simulations.

1. 1. Introduction

Forest management planning has a rich history of applying growth and yield models to provide decision support. The Forest Vegetation Simulator (FVS) is perhaps the most widely used growth and yield modeling system in the United States due to its ability to simulate a large variety of forest management activities; provide the user with site-, stand-, and tree-level outputs; and its flexibility to link to other

models (Crookston and Dixon, 2005; Dixon, 2002; Keyser and Keyser, 2017). FVS is a distant-independent (i.e. non-spatially explicit), individual-tree model that uses a series of sub-models to simulate forest stand development by growing and assigning mortality to individual tree records (Dixon, 2002). FVS is regularly used for US National Environmental Protection Act (NEPA) analyses and for the development of Forest Management Plans throughout the US National Forest system (Bettinger et al., 2015; Hill, 1997; Martinez, 2008). Extensions to the

* Corresponding author.

E-mail address: benjamin.bagdon@bia.gov (B.A. Bagdon).

base model are available for simulating disturbance events such as fire effects and behavior (Fire and Fuels Extension, Rebain, 2010), climate change (Crookston, 2014), and insect and pathogen mortality (Dixon, 2002). FVS has been used in academic research and applied problems related to fire management (Finney et al., 2007; Fulé et al., 2004), wildlife habitat evaluation (Maffei and Tandy, 2001), climate change resilience and response assessment (Bagdon and Huang, 2014; Bagdon et al., 2017), and insect and pathogen mitigation (McMahan and Monahan, 2017; Rodrigue et al., 2017), as well as economic and timber supply analysis (Bettinger et al., 2015; Diaz et al., 2018).

The developers of FVS at the US Forest Service Forest Management Service Center (FMSC) continuously refine the model by embedding regionally-calibrated growth, mortality, and volume equations into the FVS framework, resulting in 22 different regional variants of FVS as of July 2020. The Central Rockies Variant (FVS-CR) covers a variety of forest types present in the central and southwestern regions of the US. FVS-CR includes a Lodgepole Pine (LP) model for lodgepole pine forest types that cover approximately 1.6 million hectares in this region ("EVALIDator Version 1.8.0.00," 2019). Species represented within the FVS-CR LP model include lodgepole pine (*Pinus contorta* var. *latifolia*), quaking aspen (*Populus tremuloides*), Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa* var. *lasiocarpa*), and Douglas-fir (*Pseudotsuga menziesii*), in addition to other less abundant species. FVS-CR LP has been used to predict stand development and fuel loads following mountain pine beetle (*Dendroctonus ponderosae*) outbreaks (Collins et al., 2012, 2011), resilience and adaptation to climate change (Buma and Wessman, 2013), the interaction between bark beetle outbreaks and fire behavior and effects (Caldwell et al., 2013; Collins et al., 2012; Klutsch et al., 2011), and the effect of management on water yield (Havis, 2017), as well as to inform forest management planning.

However, a growing body of literature has found that default settings in FVS do not always produce reasonable results and that a substantial model evaluation and calibration effort may be necessary for acceptable model performance in many applications (Canavan and Ramm, 2000; Diaz et al., 2018; Ex and Smith, 2014; Froese and Robinson, 2007; Leites et al., 2009; Pokharel and Froese, 2008; Russell et al., 2013). One study suggested that the Lake States Variant of FVS requires a major reengineering (Pokharel and Froese, 2008). FVS includes an auto-calibration feature that can apply species-specific multipliers to the various growth equations in the model in order to better match user-provided diameter and/or height increment data. While auto-calibration likely improves the accuracy and precision of individual tree records, it may overfit the model to a small non-representative data sample. More broadly, it remains unclear whether overall FVS-CR LP model accuracy is within acceptable limits. Formal evaluations of FVS variant performance are rare because 1) these activities are time-consuming, 2) they require data that are difficult to obtain or unavailable, and, 3) users may assume that the model's many variants, its many programmable options, and its long history of development by the USFS indicate that it is ready for use 'as is' in most situations. Examples of formal FVS model evaluation and calibration efforts are available in many regions of the US (e.g., Froese and Robinson, 2007; Lacerte et al., 2004; Leites et al., 2009; Pokharel and Froese, 2008; Russell et al., 2013), though such studies are relatively limited for the western U.S. (e.g. Ex et al., 2016; Ex and Smith, 2014). Because FVS has been widely adopted as a tool to inform research and forest planning decisions, the impacts of relying on biased model outputs have the potential to adversely affect forest health on a regional or even national scale.

The FMSC emphasizes the importance of evaluating and calibrating FVS for each unique application (personal communication Chad Keyser, 21 June 2018), and has created a protocol with suggested evaluation procedures to assess overall model behavior as well as the behavior of individual sub-models for FVS regional variants (Cawrse, 2010). We used suggestions from this protocol to develop a conceptual framework for model evaluation and calibration, presented in Fig. 1. Much debate

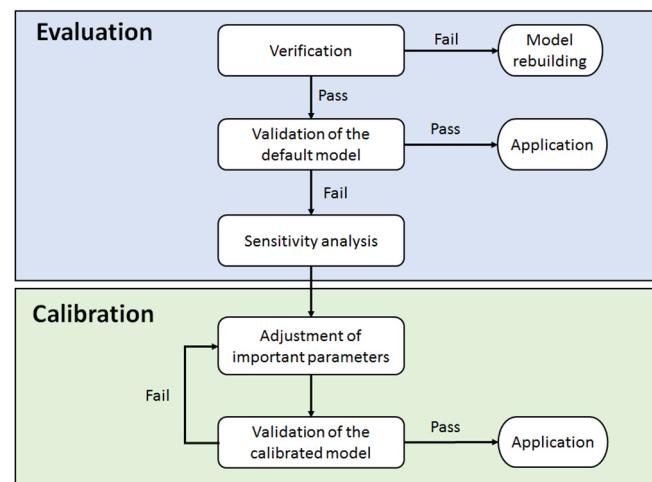


Fig. 1. Conceptual framework of model evaluation and calibration processes.

has centered on a standard vocabulary for model evaluation techniques and activities (Augusiak et al., 2014; Jørgensen, 2011; Rykiel, 1996; Vanclay and Skovsgaard, 1997). The term 'verification' generally describes activities to check that the conceptual structure of a model is sound and the computer coding is executing as intended (Augusiak et al., 2014), whereas model 'validation' connotes testing procedures that compare model predictions to independent observed values to evaluate model errors and judge accuracy and precision. Verification testing is often a qualitative exercise while validation is a quantitative one, though in both cases users should ideally determine *a priori* standard by which to judge whether the model performance is satisfactory. If structural problems become apparent during verification, the model should be redesigned to address the identified problems. If verification has shown that the structural aspects of the model are acceptable, validation tests can then determine whether the default settings result in reasonable estimates. If the default model settings are adequately accurate and precise, the model is ready for use in a particular application; if not, the model requires calibration. Sensitivity analysis can identify which parameters have the greatest effect on model behavior, helping to streamline calibration. Calibration efforts should examine how parameters were originally estimated and whether adjustments are scientifically justified.

This study evaluated the FVS-CR LP model against independent datasets of lodgepole pine forest structure and growth in Colorado and Wyoming, in support of ongoing efforts to better understand carbon dynamics in stands affected by the mountain pine beetle epidemic. The model was evaluated in four ways. First, we verified default model behavior against well-known relationships of forest growth in a Bakuzis matrix, as modified by Leary (1997). Second, we performed stand-level validation by comparing simulation results for stand live carbon accumulation and other key metrics of stand structure over time to a chronosequence derived from USFS Forest Inventory and Analysis (FIA) plot data. Third, we evaluated the large-tree diameter increment sub-model for four species against FIA tree re-measurements using equivalence testing (Robinson et al., 2005; Robinson and Froese, 2004). Fourth, we conducted a one-at-a-time local sensitivity analysis to identify and rank model parameters and inputs that have the greatest influence on key model outputs. We seek to answer three questions with these analyses:

- (1) Does the model comply with widely accepted theories of stand dynamics? (Verification)
- (2) Are the errors of the default model reasonable? (Validation)
- (3) What parameters and/or inputs most strongly drive model behavior? (Sensitivity Analysis)

This study does not present a FVS-CR LP model re-calibration, but rather uses the evaluation process shown in Fig. 1 to inform future calibration efforts.

2. Methods

2.1. FVS-CR model description

The FVS-CR LP model sequentially grows individual trees on single- or multi-species plots using four growth sub-models representing large-tree diameter increment, large-tree height increment, small-tree growth, and mortality. Simulation of a plot requires specification of a site identification number, the appropriate FVS regional variant, the inventory year, and information regarding the tree and seedling sampling plot designs. A corresponding initial tree list containing the species and diameter at breast height (DBH) of the trees sampled is also required input data for running the model. Other plot-level variables such as site index (SI), elevation, slope, and aspect, as well as tree-level variables such as the number of trees represented by a tree record, height, and crown ratio are optional model inputs that should be provided when possible but are not required to run the model. FVS-CR will fill in some of these missing variables with default values that are described in the “Central Rockies Variant Overview” (Keyser and Dixon, 2008). FVS-CR LP contains a partial establishment model which will, by default, simulate the re-sprouting of certain vegetatively-propagated species when they are present in the tree list. However, simulating stand regeneration requires specification of seedling counts for other species.

The FVS-CR LP model estimates diameter and height growth of small and large trees separately. At each time step of a simulation (typically 10 years), the model first estimates diameter growth of “large” trees (defined as ≥ 2.54 cm DBH for all species modeled in this study by default), which is then used to estimate large-tree height increment. Next, height growth is estimated for small trees, from which the small tree diameter increment is then derived. Finally, the mortality sub-model computes and distributes stand mortality across the tree list, and a crown sub-model adjusts characteristics of the tree crowns. Throughout this sequence, the tree list is scaled up from the sample plot level to the stand level, and the resulting stand-level metrics iteratively inform the nested set of equations in each submodel.

2.1.1. Diameter increment sub-models

Diameter increment is the primary driver of stand growth in FVS-CR. Every projection cycle begins by estimating diameter growth in larger trees, which is subsequently used to predict other processes such as height increment and mortality. The diameter of each large tree is modeled directly in the LP model using the GNGYM diameter growth model (Edminster et al., 1991):

$$DF = b_1 + b_2 DBH + b_3 BA + b_4 SI + b_5 \ln(DBH)^2 + b_6 (BAU/BAT) + b_7 \ln(BAT) + b_8 DBH^2 \quad (1)$$

where DF is tree diameter at breast height at the end of the cycle, DBH is current tree diameter at breast-height, BA is total stand basal area, SI is site index (default base age of 100 years), BAU is the total basal area of large trees (≥ 2.54 cm diameter class), BAT is the total stand basal area (restricted to a maximum value of $3.2 \text{ m}^2 \text{ ha}^{-1}$ for the purposes of this calculation), and b_1 thru b_8 are species-specific parameters.

Small-tree diameter increment is calculated as a function of tree height (see following section). The diameter increment for small trees over 4.5 feet tall is calculated as the difference of predicted diameters at the start and end of the projection period, adjusted for bark ratio, as estimated using Wykoff's et al. (1982) height-diameter logistic functional form:

$$HT = 4.5 + e^{\left(p_1 + \frac{p_2}{(DBH+1)}\right)} \quad (2)$$

where HT is tree height, and p_1 and p_2 are species-specific parameters. By default, diameter growth is set to zero for small trees below 4.5 feet tall.

2.1.2. Height increment sub-models

Large-tree height increment predictions are differentiated between even-aged and uneven-aged stands. A stand is considered even-aged when the difference between the oldest and youngest trees is less than or equal to 40 years. In both cases, large-tree height increment is calculated as the difference between estimated heights at the start and end of the projection period. Even-aged large-tree height is predicted from the site index curves developed by Alexander et al. (1967):

$$\begin{aligned} HTE = & 9.89331 + (-0.19177 \times AGETEM) + (0.00124 \times AGETEM^2) \\ & + (-0.00082 \times (CCF - 125) \times SI) + (0.01387 \times AGETEM \times SI) \\ & + (-4.55 \times 10^{-5} \times AGETEM^2 \times SI) \times AGEADJ \times FACTOR \end{aligned} \quad (3)$$

where HTE is the even-aged large-tree height estimate, $AGETEM$ is the tree's estimated age (constrained to range from 30 to 200 years for the purposes of this equation), CCF is stand crown competition factor (constrained below 125), SI is species site index, $AGEADJ$ is modifier when tree age is below 30 years, and $FACTOR$ is a reduction factor based on canopy position. GNGYM height equations are used instead for uneven-aged stands:

$$HTU = 4.5 + (a_1 + a_2 \times SI) \times [(1 - e^{(a_3 \times DBH)})^{(a_4 \times BAT)^{a_5}}] \quad (4)$$

where HTU is the uneven-aged large-tree height estimate, DBH is tree diameter at breast height, BAT is the total stand basal area (restricted to $3.2 \text{ m}^2 \text{ ha}^{-1}$ for the purposes of this equation), and a_1 – a_5 are species-specific coefficients. Small tree height increment is estimated directly as a function of site index, stand density, and the crown ratio:

$$HTG = \frac{SI}{(15 - (4 \times RELSI))} \times [(150 \times CR^3 \times e^{(-6 \times CR)}) + 0.3] \times PCTRED \quad (5)$$

where HTG is estimated small-tree height growth for the cycle, $RELSI$ is the relative site index of the site species, CR is a tree's live crown ratio, and $PCTRED$ is reduction in height growth due to stand density.

2.1.3. Mortality sub-models

FVS-CR uses a stand density index (SDI)-based mortality model (Dixon, 2002) to determine and disperse stand mortality to individual tree records. The model considers both background and density-related mortality rates based on the relationship between current SDI (Eq. (6)) and maximum SDI (Eq. (7)):

$$\begin{aligned} SDI = & \sum \left(TPA_j \times 10^{-1.605} \times \left(1 - \frac{1.605}{2}\right) \times QMD^{1.605} \right. \\ & \left. + 10^{-1.605} \times \left(\frac{1.605}{2}\right) \times QMD^{(1.605-2)} \times DBH_j^2 \times TPA_j \right) \end{aligned} \quad (6)$$

$$StndMaxSDI = \frac{\sum MaxSDI_i \times BA_i}{StndBA} \quad (7)$$

where TPA_j is trees per acre of tree record j , QMD is the quadratic mean diameter for all trees in the stand, DBH_j is diameter at breast height of tree record j , $StndMaxSDI$ is the maximum SDI for the entire stand, $MaxSDI_i$ is the maximum SDI for species i , BA_i is the total basal area of species i , and $StndBA$ is the total basal area for the stand. When a stand is below 55% of the maximum SDI threshold, a fixed annual background mortality rate is applied to each tree record and compounded throughout the projection cycle:

$$RIP = \sum 1 - \left[1 - \frac{1}{1 + e^{(b_0 + b_1 \times DBH_j)}} \right]^{YRS} \quad (8)$$

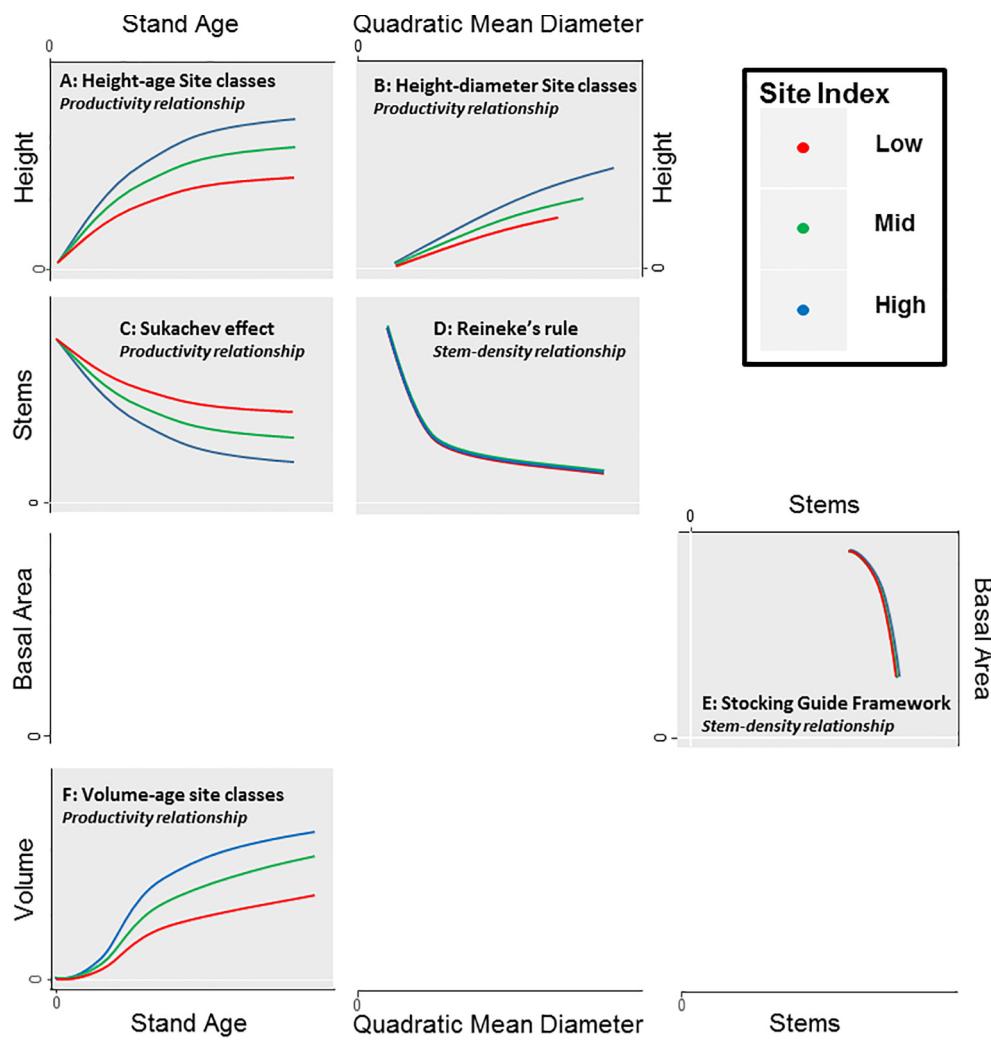


Fig. 2. Reformatted Bakuzis matrix with expected behaviors for six “law-like” relationships. Relationships are categorized as either ‘site productivity’ (A, B, C, and F) relationships that should exhibit differences between sites of different quality, or ‘stem density’ (D and E) relationships that should not show dependence on site quality.

where RIP is the background mortality rate for the current projection cycle, YRS is the number of years in the current projection cycle, b_0 and b_1 are species-specific coefficients ($b_0 = 5.9617$ and $b_1 = -0.03401$ for lodge pole), and Σ is summation over all trees in the stand. The total stand background mortality is the sum of the individual tree background rates. When stand density is above 55% of maximum SDI, density-related mortality is used. The mortality rate is adjusted to match the trend in trees per acre (TPA) predicted from the SDI relationship:

$$TPA = \frac{StndMaxSDI}{0.0248} \times QMD^{-1.605} \quad (9)$$

Once the amount of stand mortality has been determined, it is dispersed across the individual tree records by adjusting the number of trees per acre represented by the records. Each tree record receives a calculated mortality in a projection, until the target amount of tree mortality has been achieved. Mortality rates for individual tree records are determined as a function of the tree’s size relative to the other trees within the stand, the shade tolerance of the tree species, and the tree’s crown ratio (see section 7.3.2.2 in Dixon, 2002).

2.2. Independent dataset development and evaluation

We developed three independent datasets from inventory data of

lodgepole pine forest structure and growth in Colorado and Wyoming to evaluate the FVS-CR LP model. These three datasets are linked with this manuscript and are available for download. Two of the datasets—used for validating multiple stand-level metrics and large tree diameter growth rates, as detailed in the following sections—were derived from FIA, the national-scale forest inventory program designed to highlight forest status and trends. These datasets are described in detail in subsequent sections. The third dataset was developed to initialize FVS using regeneration inventory data collected in 2015 at the Colorado State Forest (CSF).

Since FVS-CR only contains a partial-regeneration model, it was necessary to identify a realistic number and species composition of seedlings to use in initializing the LP model for simulations of lodgepole forest stand regeneration in northern Colorado. Because post-disturbance regeneration data is difficult to accurately glean from FIA plot data in our study area, a separate regeneration dataset was derived from another source. In 2015, we surveyed 27 plots within the CSF in lodgepole pine stands that were regenerating after previous clearcutting. We randomly placed 30 m × 30 m plots in stands that were clearcut 8–40 years prior to sampling (mean 22 years; Woodward et al., 2017), and counted all trees above 2.54 cm DBH. Seedlings and saplings were then counted in a 20 m × 5 m subplot at the center of the larger plot. Eight of the 27 plots had been pre-commercially thinned after the initial clearcut but before our sampling, and in those cases we included

the stumps of thinned trees within our tree count (but did not count any older stumps that pre-dated the thinning). The counts of mature trees, thinned stumps, seedlings, and saplings were summed for each plot, and the mean of those total regeneration counts across all 27 sampled plots—8,246 trees per hectare (TPH)—was used to set the initial seedling density for the FVS simulations in this study. We also calculated the species composition of the 19 plots that were not thinned (because the species of the thinned stumps could not be reliably determined) for use in initializing simulations of mixed-species lodgepole pine stands.

2.3. Verification through a modified Bakuzis matrix

The first step in our model evaluation procedure was to qualitatively verify model behaviors by comparing the model predictions with well-observed relationships about stand dynamics. This test does not inform about model accuracy or precision, but rather about potential structural deficiencies. Failure in capturing well-recognized stand patterns might require major model re-engineering to address problems instead of a simple calibration. We used the modified Bakuzis matrix (Fig. 2) described in Leary (1997) to perform a verification analysis on the FVS-CR LP model. Leary (1997) simplified the 64-frame Bakuzis' matrix (Bakuzis, 1969) into a 4×4 matrix that describes relationships of stand behavior that have been found to occur repeatedly in even-aged, single-species stands, with what he termed as “law-like regularities”.

The six relationships chosen by Leary fit into two categories: site productivity relationships and stem density relationships. The height-age, height-diameter, density-age, and volume-age relationships are highly influenced by site quality. The height-age (Fig. 2A) and height-diameter (Fig. 2B) relationships are expected to show diverging curves depending on site quality. The stem density versus age relationship, or Sukachev effect (Fig. 2C), posits that higher quality sites self-thin faster than poor quality sites, and the volume-age relationship (Fig. 2F) is often used as a basis for assessing site quality. In contrast, density-diameter and density-basal area relationships are more dependent on the number of trees per unit area. Reineke's rule (Fig. 2D) is a mathematical relationship stating that the number of stems in a stand has a negative exponential relationship (with a coefficient of approximately -1.6 for many species) with the QMD, regardless of site quality, and that stem density and QMD should be linear on log-log scales. The stocking guide framework (Fig. 2E) is widely used by foresters (e.g., stocking charts, density management diagrams), and is based on the observation that stands are limited in the number of trees and basal area that can be supported, i.e., full stocking (see Ginrich, 1967).

Some of these relationships are represented more-or-less directly in the internal equations underlying the various FVS sub-models (Section 2.1). Site index is a direct input used in the calculation of large-tree diameter increment (Eq. (1)) and height increment for both small and large trees (Eqs. (3)–(5)), facilitating the expected stand height-age and volume-age behaviors (Fig. 2A&F). Similarly, density-dependent mortality is implemented such that stand density is adjusted in response to current values of QMD and BA (Eqs. (7) & (9)), which implies a direct representation of Reineke's rule and the stocking guide framework (Fig. 2D&E) in stands that have grown above 55% of the maximum stand density index. In contrast, the height-diameter and density-age relationships (Fig. 2B&C) are the product of dynamic interactions between the height increment in even aged stands, diameter increment, and density-dependent mortality models, and thus the expected dependencies on site productivity could be considered an emergent property of the model. In all of these cases, independent verification of model behaviors is required to check for unexpected interactions between sub-models or other structural issues.

Evaluating the FVS-CR LP model against a modified Bakuzis matrix (Fig. 2) required simulating the growth of lodgepole pine stands from bare ground across sites of different quality (i.e., different SI). We conducted separate simulations for low, medium, and high productivity

sites by setting SI values of 12.2, 15.5, and 20.8 m height (base age 100 years), respectively, corresponding to the 10th, 50th, and 90th percentile values from the FIA-derived chronosequence dataset described in the next section. We then initialized each simulation with 8,246 TPH of lodgepole pine seedlings (Section 2.2), simulated stand development over a 200-year projection, and plotted the appropriate stand output variables within the modified Bakuzis matrix framework. While a quantitative assessment of predicted versus observed behavior within the Bakuzis matrix framework would be useful, it was outside the scope of the current study.

2.4. Validation of stand-level growth through a chronosequence approach

We quantitatively evaluated long-term model growth and development behavior by comparing FVS-CR LP model predictions to chronosequences constructed from FIA plot observations at physically-similar sites and sorted by stand age (i.e., space-for-time substitution). First, we identified all FIA plots across Colorado and Wyoming, measured from 2002 to 2017 that were composed of $\geq 65\%$ lodgepole pine by basal area. We filtered out any plots that were identified as having wildfire disturbance (0.9% of our plots). However, we elected to retain stands that were coded as having disturbance from weather events and insects (34% of our plots), since periodic outbreaks of the endemic mountain pine beetle are normal in this region. An indication of management was not present on any of our plots in the FIA database. We then grouped the resulting 554 plots of this FIA-derived chronosequence dataset (Section 2.2) into ten-year age classes. For each age class, we calculated the mean and 95% confidence interval for stand density (trees per hectare including seedlings), BA ($m^2 ha^{-1}$), height of dominant and codominant trees (m), QMD (cm), live carbon ($Mg C ha^{-1}$), and carbon in dead wood ($Mg C ha^{-1}$), resulting in a 200-year chronosequence of lodgepole pine forest development (Fig. 5; see Rebain 2010 and supplementary material for definitions of carbon pools). We tested whether splitting the FIA chronosequences according to lower and higher site index would provide additional insight into site-productivity dependent model behavior. However, our exploratory analysis found a high degree of chronosequence overlap between the lower and higher site qualities for all metrics other than height. This exploratory analysis can be found in the supplementary materials.

Because the FIA plots also contained fractions of Engelmann spruce, subalpine fir, and quaking aspen (as did the CSF plots), we constructed two separate scenarios in FVS simulating the development of both ‘pure’ and ‘mixed’ lodgepole pine stands for comparison against the chronosequence. The pure lodgepole scenario was initialized with 8,246 TPH of lodgepole pine seedlings, and the mixed scenario with the same total seedling density but with the species composition adjusted to reflect the average of our CSF dataset (6,087 TPH of lodgepole pine, 1,183 subalpine fir, 759 quaking aspen, and 216 Engelmann spruce seedlings). We conducted 200-year FVS simulations with default settings for each scenario, and separately calculated the root mean squared error (RMSE; Eqn. (10)), normalized RMSE (nRMSE; Eqn. (11)), and bias (Bias; Eqn. (12)) of the FVS-simulated stand characteristics against the corresponding mean observed values in the chronosequence. A benefit of using RMSE and bias is that these metrics are reported in the same units as the measured and predicted values, increasing the ease of interpretability. Normalized RMSE (nRMSE) is the RMSE expressed as a percentage of the observed mean.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(observed_i - predicted_i)^2}{n}} \quad (10)$$

$$nRMSE = \frac{RMSE}{mean(observed)} \quad (11)$$

$$Bias = \sum_{i=1}^n \frac{(predicted_i - observed_i)}{n} \quad (12)$$

2.5. Validation of tree-level growth through equivalence testing

RMSE, Bias, and nRMSE, and other commonly reported metrics are generally used as a first order measure of a model's ability to reasonably predict reality. Equivalence testing is a more formal method of validation that seeks to determine whether model predictions are statistically equivalent to observed data. Rather than testing for similarity between the predicted and observed values, the null hypothesis in equivalence testing is that the model *does not* produce statistically similar estimates to the measured data (i.e., dissimilarity). Only when a model produces unbiased estimates can the null hypothesis be rejected.

To enable a more detailed validation of the large-tree diameter growth increment sub-model of FVS-CR LP, we compiled an additional dataset based on individual tree records within the FIA database (Section 2.2) that have been re-measured at least once. We started with the same set of lodgepole pine FIA plots used for the chronosequence. In order to evaluate the growth of both lodgepole pine and other species that frequently co-occur in lodgepole pine forests, we further filtered to keep only plots that contained re-measured tree data for lodgepole pine, quaking aspen, Engelmann spruce, and subalpine fir. The resulting filtered dataset contained increment data for 3,790 individual trees across 124 FIA plots. The time elapsed between revisits varied between plots, so we followed the method used by Pokharel and Froese (2008) to standardize diameter increments to a 10-year interval.

We then performed a “two one-sided equivalence test” (TOST), described by Robinson et al. (2005), on the resulting dataset. In this method, predicted values are shifted by their mean and then linearly regressed against the observed values in order to make separate independent evaluations of whether predicted and observed population means match (as indicated by comparing the regression intercept to the prediction mean), and whether the model predicts accurately across the full range of the observations (as indicated by comparing the regression slope to a 1:1 line). A non-parametric bootstrapping method is used to construct confidence intervals (CI) around both the slope and intercept of the regression. The TOST method then checks whether these slope and intercept CIs fall within previously defined equivalence intervals (EI) representing a ‘tolerance’ level for an acceptable fit (e.g., $\pm 10\%$). The *a priori* selection of EI tolerance values is arbitrary, similar to the selection of ‘alpha’ prior to conducting a *t*-test. When the CIs are completely contained within the EIs, the null hypothesis of dissimilarity can be rejected (Fig. 3B). It is also possible for the intercept to pass the equivalence test while the test of the slope fails, indicating that the average model error is acceptable but the model struggles to accurately

predict larger and/or smaller values. For a more detailed discussion on the philosophical and statistical underpinnings of equivalence testing in a forest modeling context see (Robinson et al., 2005; Robinson and Froese, 2004). For our TOST, we used 10% and 25% of the observed mean decadal diameter increment for each species as the EI tolerance for the intercept and slope, respectively. These tolerance values were similar to those used in Robinson et al. (2005). Statistical computing was performed using the “equivalence” package in R (R Core Team, 2018; Robinson, 2016).

2.6. Local sensitivity analysis

Model sensitivity refers to the change in model predictions relative to a change in input data or internal parameter values (Campolongo et al., 2007; Jørgensen and Fath, 2011; McKenzie et al., 2019). Local sensitivity analysis (LSA) involves systematically quantifying the relative influence of individual data inputs and parameters of interest by adjusting each one at a time, holding constant all others, and recording the change in the model output variable(s) of interest. This analysis can identify which parameters have the greatest effect on model behavior, and thus inform subsequent calibration efforts.

We conducted an LSA to evaluate the influence of 20 different FVS model inputs and parameters on eight important model outputs: total stand carbon (TotC), forest surface carbon (SurfC), live tree basal area (LiveBA), standing dead tree basal area (MortBA), trees per hectare (TPH), stand density index (SDI), quadratic mean diameter (QMD), and mean height of dominant and co-dominant trees (Ht). The different inputs and parameters tested were selected from among all those used in FVS to represent the different growth sub-models described in Section 2.1, as well as crown growth, regeneration, geophysical setting, and the Fire and Fuels Extension snag and surface fuel models (Table 1). We initialized the simulation with the CSF seedling count (8,246 TPH) and perturbed each parameter value by 1%, 10%, and 25% above and below its default setting, holding all other parameters constant at their default value and evaluating the response in the eight output variables after 50 years (short-term) and 200 years (long-term) of simulation. We ranked the relative sensitivity of each output to the different inputs/parameters by separately computing the absolute value of the response at each variation level ($\pm 1\%$, $\pm 10\%$, and $\pm 25\%$) for each time-frame (50 and 200 years), and then calculating the grand average across variation levels.

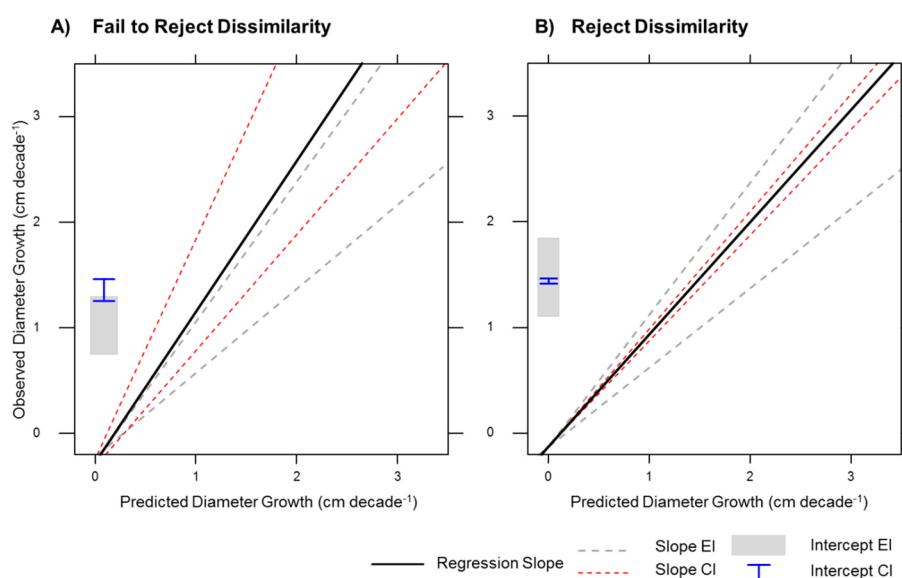


Fig. 3. Graphical representation of TOST equivalence testing. Panel A illustrates an example of a failure to reject the null hypothesis of dissimilarity for both the slope and intercept, as neither confidence interval (CI) is completely contained within the corresponding equivalence interval (EI; shown in grey). This indicates that both the predicted population mean and the modeled-vs-measured slope fall outside of pre-determined tolerance range. In contrast, Panel B shows the null hypothesis of dissimilarity being rejected because both confidence intervals are completely contained within the corresponding equivalence intervals.

Table 1

Parameters and inputs examined in the local sensitivity analysis.

FVS-CR Component Models / Parameters		Parameter Abbreviation	Default Values (LP model)
	<i>Diameter Increment Models (large and small tree)</i>		
1	Pre-calibration multiplier for large tree DBH growth (P)	LrgTrPreCal	1
2	Large-tree fixed diameter growth multiplier (P)	LrgTrFixDG	1
3	Large-tree basal area increment multiplier (P)	LrgTrBAI	1
4	Small-tree diameter increment multiplier (P)	SmTrDG	1
	<i>Mortality Models (Density-related and Background)</i>		
5	Maximum Stand Density Index (P)	MaxSDI	675 lodgepole pine
6	Mortality Multiplier- Distribute Uniformly (P)	MortUni	1
7	Mortality Multiplier- Distribute small to large (P)	MortSmLrg	1
8	Mortality Multiplier- Distribute large to small (P)	MortLrgSm	1
	<i>Height Increment Models</i>		
9	Large-tree height increment multiplier (P)	LrgTrHt	1
10	Small-tree height increment multiplier (P)	SmTrHt	1
11	Site Index (I*)	SI	19.8 m, base age 100
	<i>Crown Model</i>		
12	Crown Ratio estimate multiplier (P)	CR	1
	<i>Regeneration Input (Partial establishment)**</i>		
13	Lodgepole pine planted in period 1 (I)	LpReg	8,246 TPH
14	Proportion of spruce-fir planted to lodgepole pine (I)	SpFirReg	
15	Proportion of aspen planted <i>with sprouting</i> to lodgepole pine planted (I)	AsSprout	Sprouting on
16	Proportion of aspen planted <i>without sprouting</i> to lodgepole pine planted (I)	AsNoSprout	
	<i>Geophysical Settings</i>		
17	Elevation (I)	Elev	2,743 m
18	Aspect (I)	Asp	0
	<i>Fire and Fuels Extension</i>		
19	Snag fall rate multiplier (P)	SngFall	1
20	Surface fuel decay rate multiplier (P)	SurfFuelDec	1

(P) denotes fixed parameter; (I) denotes a user input

* Site index is an optional input parameter in FVS-CR; a default value of 19.8 m is applied if none is input.

** The partial establishment model requires users to specify initial trees per hectare.

3. Results

3.1. Modified Bakuzis matrix

The verity of the site productivity relationships predicted by the FVS-CR LP model (Fig. 4) were qualitatively evaluated by judging whether cells A, B, C, and F follow the expected trajectories illustrated in Fig. 2. The simulated height-age relationship (Fig. 4A) appeared consistent with expectations (Fig. 2A), with the modeled dominant and codominant trees growing in height faster at higher quality sites than lower quality sites. The expected height-diameter relationship (higher quality sites having taller trees within a given QMD range) and volume-age relationship (faster volume accumulation at higher-quality sites) were also verified (Fig. 4B & F). FVS-simulated stands at higher quality sites also showed somewhat faster self-thinning as compared to lower quality sites (Fig. 4C), consistent with the Sukachev effect.

The modeled results only partially followed the expected relationships concerning Reineke's rule. As expected, the density-diameter relationship exhibited no site quality dependency, i.e., the curves of Fig. 4D overlap. However, the modeled relationship between TPH and QMD only exhibited the expected linear relationship with slope of approximately -1.6 on a log-log scale (Supplemental materials) after approximately 90 years into the simulation. This suggests that the background mortality rate in young stands is much lower than that of the density related mortality model, as evidenced by the relatively flat initial curve and then discontinuity around QMD = 7 cm.

3.2. Stand-level validation against FIA-derived chronosequence

Performance of the pure lodgepole and mixed stand simulations relative to the FIA-derived chronosequence is shown in Fig. 5. The FIA data in early periods (i.e., 0 to 30-year-old stands) appears to contain

uneven-aged stands with a significant fraction of older trees, resulting in relatively higher initial values for height and QMD (Fig. 5C&D) than a true regenerating even-aged stand might be expected to actually contain. Similarly, the relatively low initial stand density (Fig. 5A) shows minimal evidence of an initial high density of seedlings as is often encountered during stand regeneration after a stand-replacing disturbance. However, the rate of increase and end values for stand density, BA, height, QMD, and carbon density still provide a useful comparison point for our FVS simulations.

The CSF-derived initial seedling density of our simulations is quite high compared to the stand densities observed in the FIA chronosequence data, though the simulated stands begin to self-thin after ~ 30 years. The density of the simulated pure lodgepole and mixed stands drop down to within the confidence interval of the chronosequence after approximately 100 and 60 years of simulation, respectively, and after year 100 the simulated density of the mixed stand is consistently lower than that of the chronosequence. Simulated stand basal area (Fig. 5B) and live carbon density (Fig. 5E) both show aggressive growth after 10–20 years of simulation, exceeding the chronosequence CI after 30–50 years, and reaching values more than double the chronosequence mean by year 200. Evaluated across the full 200-year length of the simulation, these metrics have nRMSE values nearing or exceeding 100%. The simulated mixed stand reaches a live carbon density of $\sim 250 \text{ Mg C ha}^{-1}$ by the end of the simulation, approximately three times the mean value in the chronosequence. Simulated stand height (Fig. 5C), QMD (Fig. 5D), and dead wood carbon density (Fig. 5F) all show some over-estimation by the end of the simulation as well, though to a much lesser extent.

3.3. Tree-level validation

When evaluating FVS-predicted large-tree diameter increment

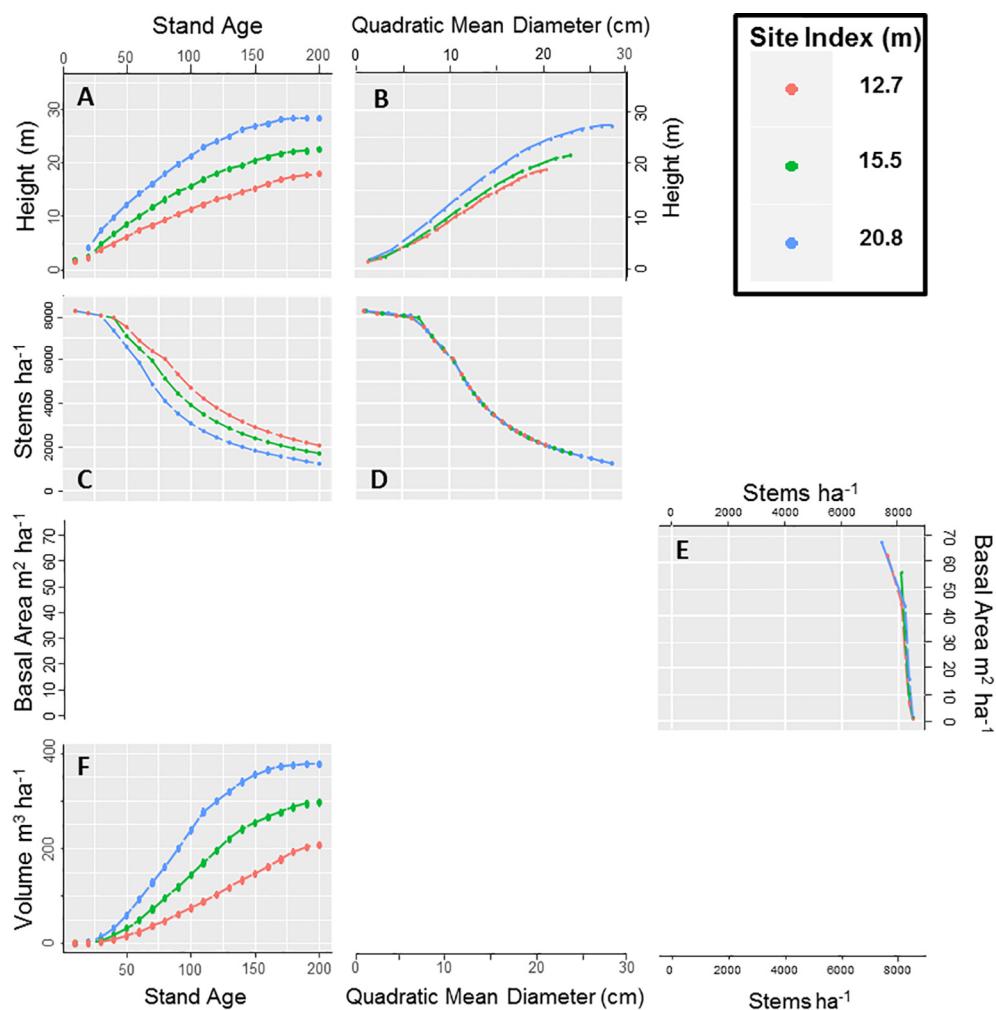


Fig. 4. Modified Bakuzis matrix, as adopted from Leary (1997), for the FVS Central Variant Lodgepole Pine Model. Figure legend shows the site index (base age of 100 years) used to represent lower-quality sites (red), average-productivity sites (green), and higher productivity sites (blue). Site–productivity relationships are characterized by cells A, B, C, and F, while stem–density relationships are shown in D and E. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

against FIA data, only one of the species tested—lodgepole pine—passed both the intercept and slope TOST equivalence tests (i.e., confidence interval contained completely within the equivalence interval; Fig. 6, Table 2). However, regression revealed that the variance in observed versus predicted values was not well explained for lodgepole pine ($R^2 = 0.11$). Together these results suggest that there are many factors affecting individual tree growth that are not well-represented in the FVS-CR LP model, but the GENGYM model form and default parameter values nonetheless do succeed in producing unbiased predictions of lodgepole pine diameter increment across the full range of simulated values. In contrast, the intercept equivalence tests for subalpine fir and Engelmann spruce showed 95% and 72% of the bootstrapped CI estimates of the intercept falling above the EI, respectively, indicating an under-estimation bias for the default models of those species in FVS-CR LP. The quaking aspen intercept CI fell completely below the EI, indicating substantial over-estimation of diameter growth for that species.

Equivalence tests of the slope reveal how much prediction bias varies across the full range of prediction values, i.e., whether increment estimates are systematically less reliable at the low or high end of the prediction range. Model robustness increases as the regression slope approaches one. The confidence interval for the slope of the quaking aspen diameter growth increment regression ranged from -0.24 to 0.53 , indicating minimal relationship between simulated and observed

increment values. Traditional measures of fit also indicated poor model performance for quaking aspen, $R^2 = 0.003$, p-value = 0.55. The CI for subalpine fir was only partially contained by the EI and 75% of the bootstrapped estimates fell outside and above the EI. This indicates that the model increasingly underestimates diameter increment as trees get larger. The Engelmann spruce CI was not contained by the EI and tended to underestimate diameter growth of larger trees.

3.4. Local sensitivity analysis

Table 3 identifies, by color, the five most sensitive variables for each of the eight outputs examined (for full results see [supplemental materials](#)). For instance, adjusting the amount of aspen planted while leaving the default sprouting mechanism on (AsSprout input), resulted in an average absolute change of 24% to the dead tree basal area (MortBA) output. In addition to the MortBA output, the total stand carbon (TotC) and forest surface carbon (SurfC) were most sensitive to adjustments in the AsSprout input. The elevation, aspect, and rate of snag fall did not exhibit any output variance relative to input changes and are not shown in Table 3 (see [Supplementary Material](#) for full results), suggesting that FVS-CR LP relies heavily on optional site index data to represent site-level variability in forest productivity.

The key findings of the sensitivity analysis were that 1) the partial establishment model was the most influential component of the FVS-CR

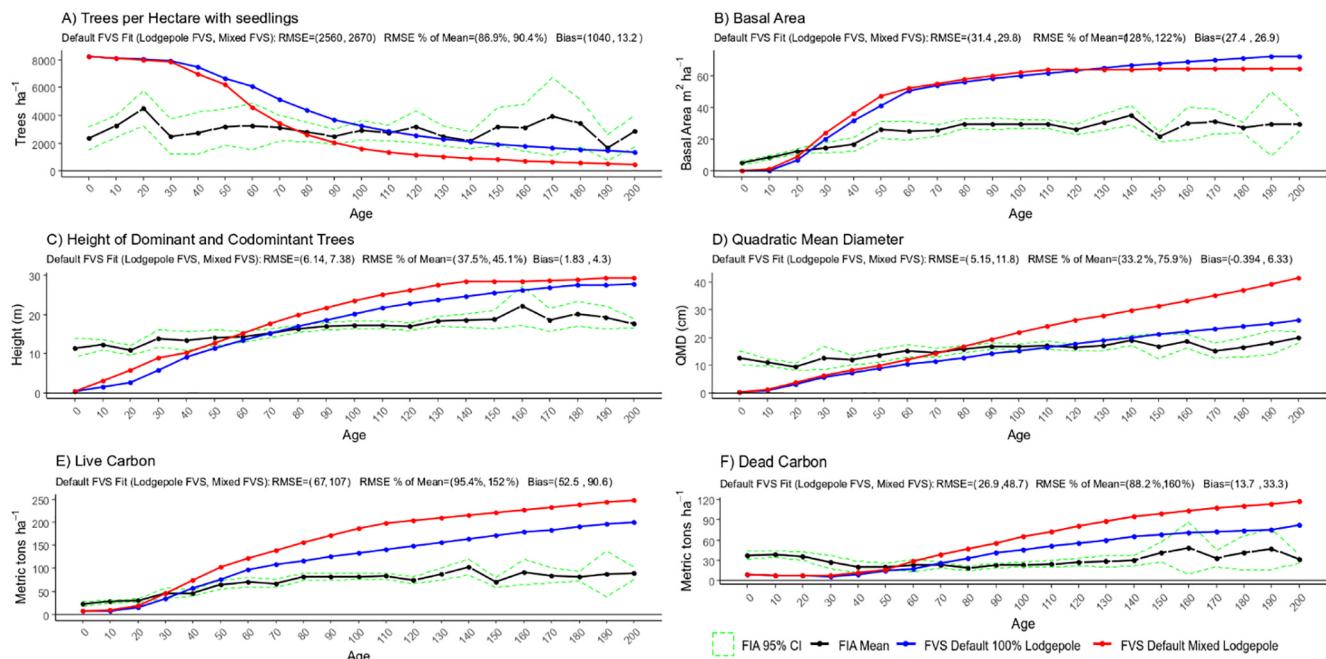


Fig. 5. Stand level validation test of the FVS Central Rockies Variant Lodgepole Pine model using a chronosequence developed with 554 lodgepole pine dominant FIA plots sampled in Colorado and Wyoming from 2002 to 2017.

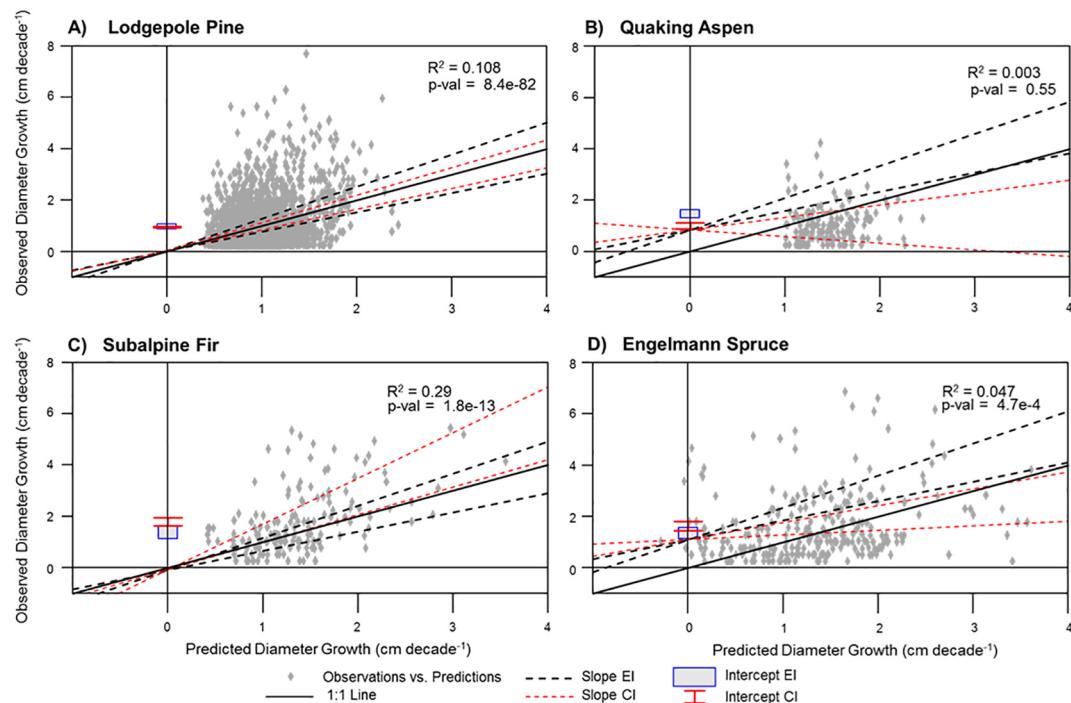


Fig. 6. TOST equivalence test results comparing the FVS-predicted large tree diameter growth increment of lodgepole pine (A), quaking aspen (B), subalpine fir (C), and Engelmann spruce (D) to FIA-derived observational data. Pairwise values falling below the 1:1 line indicates FVS model over-estimation, and vice versa.

LP model and contained 6 of the 8 most sensitive response variables evaluated; 2) any amount of aspen present, with and without the sprouting mechanism, tended to result in large structural changes at the stand level; and 3) output sensitivity is dependent on the time period considered (short-term versus long-term). The partial establishment model was most sensitive to species other than lodgepole pine and had the highest influence of all the parameters tested on the MortBA, TPH, QMD, TotC, SurfC, and SDI outputs. Aspen particularly influenced these outputs, both with and without sprouting, and had an average absolute percent change of 14.8% (without sprouting) and 14.3% (with

sprouting) on all output variables tested. We also observed that site index, the fixed large-tree diameter-increment multiplier, the maximum stand density index, and the presence of spruce and/or fir are parameters that highly influence stand development.

We found that sensitivity of some outputs depended on the time horizon at which the outputs were examined. Fig. 7 plots the sensitivity results for four relatively sensitive parameters, evaluated at both 50 and 200 years. For example, stand mortality (MortBA) was very sensitive to LrgTrFixDG parameter for a 50-year-old stand, but not for 200-year-old stands (for which aspen sprouting and maximum stand density index

Table 2

Summary of equivalence-based regression results for four species in the FVS-CR Lodgepole Pine Model. The upper table reports results of the equivalence test of the intercept while the lower table reports the results for the slope.

Summary of regression-based equivalence results for the intercept (cm decade^{-1}) when evaluating diameter increment predictions													
Species	n	\bar{y}	\hat{y}	Mean of observed		Mean of predicted		Confidence Interval		Equivalence Interval		Proportion of Bootstrapped Estimates falling...	Reject dissimilarity?
				Lower CI	Upper CI	Lower EI	Upper EI	Below Interval	Within Interval	Above Interval			
Lodgepole Pine	3,226	0.94	0.97	0.92	0.97	0.88	1.06	0%	100%	0%	Reject		
Quaking Aspen	147	1.02	1.48	0.91	1.14	1.32	1.62	100%	0%	0%	Not Reject		
Subalpine Fir	163	1.77	1.37	1.58	1.90	1.13	1.61	0%	5%	95%	Not Reject		
Engelmann Spruce	254	1.62	1.35	1.45	1.75	1.13	1.57	0%	28%	72%	Not Reject		

Summary of regression-based equivalence results for the slope (cm decade^{-1}) when evaluating diameter increment predictions													
Species	n	Slope of regression	Confidence Interval		Equivalence Region		Proportion of Bootstrapped Estimates falling...			Below Interval	Within Interval	Above Interval	Reject dissimilarity?
			Lower CI	Upper CI	Lower EI	Upper EI	Below Interval	Within Interval	Above Interval				
Lodgepole Pine	3,226	0.94	0.81	1.07	0.75	1.25	0%	100%	0%	Reject			
Quaking Aspen	147	0.12	-0.24	0.53	0.75	1.25	100%	0%	0%	Not Reject			
Subalpine Fir	163	1.35	0.98	1.71	0.75	1.25	0%	25%	75%	Not Reject			
Engelmann Spruce	254	0.38	0.18	0.56	0.75	1.25	100%	0%	0%	Not Reject			

were much more influential). Likewise, MortBA, TPH, and QMD were very sensitive at age 200 to 25% increases in the AsNoSprout input, but were relatively insensitive to adjustments at the 50 year time period.

4. Discussion

The evaluation procedure illustrated in this study allowed us to identify components of the FVS-CR LP model that produced accurate results and the areas where the model struggled. For instance, the model seemed to capture key site quality relationships, including

emergent stand behaviors such as the Sukachev effect and height–age relationships across varying site qualities. However, multiple lines of evidence seemed to indicate that FVS-CR LP does not always accurately simulate stand mortality. The model did not follow Reineke's rule in early stages of stand development, highlighting potential underestimation of background mortality. This was likely exacerbated by the relatively high seedling densities we used to initialize the model. The net result was relatively high stand densities for the first 50–90 years of simulation, a period when simulated BA and live carbon density accelerated past the mean values observed in the chronosequence. In

Table 3

Sensitivity analysis results, showing the mean absolute percent change in output variable for systematic changes to parameters and inputs (see Table 1 for abbreviations). Color shading shows the five most influential parameters for each output variable (column), with red indicating the most sensitive parameter, orange the 2nd most important, yellow 3rd, mint 4th, and green 5th.

Model Component / Parameter	MortBA	TPH	QMD	TotC	SurfC	SDI	LiveBA	Ht	Mean (Parameters)
Partial Establishment									
AsNoSprout	21	18	27	14	19	9	5	5	14.8
AsSprout	24	15	22	19	20	5	4	5	14.3
SpFirReg	2	11	12	12	6	4	5	9	7.6
LpReg	3	5	1	3	3	4	3	0	2.8
Diameter Increment									
LrgTrFixDG	17	11	9	10	7	4	6	1	8.1
LrgTrPreCal	11	10	7	8	5	4	5	0	6.3
LrgTrBAI	11	10	7	8	5	4	5	0	6.3
SmTrDG	2	1	1	2	1	1	1	1	1.3
Height Increment									
SI	12	10	7	11	12	4	5	13	9.3
LrgTrHt	0	0	0	2	5	0	0	9	2.0
SmTrHt	2	1	1	2	2	1	1	3	1.6
Mortality Model									
MaxSDI	6	12	2	5	3	9	8	0	5.6
MortLrgSm	13	5	4	1	2	3	3	2	4.1
MortSmLrg	8	2	2	1	2	2	2	0	2.4
MortUni	6	2	0	1	1	2	2	0	1.8
Other									
SurfFuelDecay	0	0	0	1	5	0	0	0	0.8
CR	1	0	0	0	0	0	0	1	0.3
Mean (output)	7.7	6.3	5.7	5.6	5.4	3.1	3.1	2.7	4.9

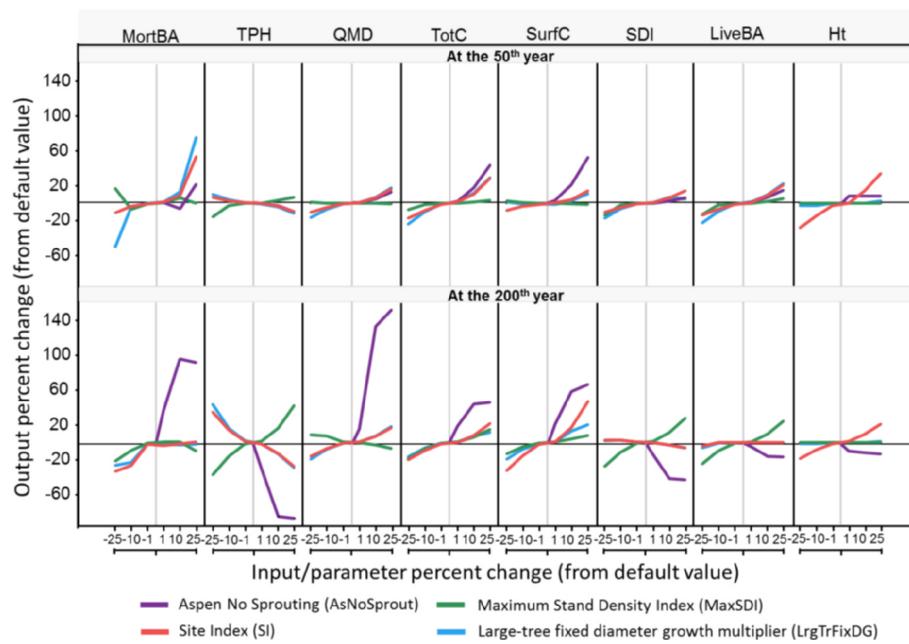


Fig. 7. Relative change in eight key outputs at 50 and 200 years as four sensitive parameters/inputs were varied by $\pm 1\%$, $\pm 10\%$, and $\pm 25\%$ of their default value. We show the most sensitive parameter/input from each model component.

addition, the unrealistically high level of biomass accumulated in the early stages of stand development resulted in a rapid accrual of dead carbon around age 50 once Reineke's rule, i.e. the SDI-based mortality model, was triggered by passing the 55% of maximum SDI threshold.

Despite this significant over-prediction of stand growth, our equivalence testing found that individual tree productivity was only systematically over-estimated for one species that composed a minor portion of simulated stands (aspen). If individual tree productivity can be ruled out as a factor for over-prediction of stand BA accumulation, then this indirectly suggests mortality may be too low. This is also consistent with the sensitivity analysis results that demonstrated the maximum SDI setting, a determinant of density-dependent mortality, was the most sensitive parameter for two of the eight outputs considered, including live BA.

We also observed that FVS-CR LP model outputs are often more sensitive to site-level data inputs (e.g., site index and seedling counts) than fixed parameters. In particular, the initial seedling density and species composition exerts an overwhelming influence on the long-term development of pure and mixed lodgepole pine stands. The proportion of aspen seedlings initially present, both with and without the default FVS sprouting mechanism, is the strongest control on six of the eight response variables that showed significant sensitivity to any factor, and in many cases the proportion of spruce and fir seedlings was the second most important factor. The extreme sensitivity of long-term stand growth to initial aspen presence is troubling, as aspen often exists as an early successional species in this region, giving way to slower-growing but more shade-tolerant conifers after 60–80 years (2005 Report on the Health of Colorado's Forests – Special Issue: Aspen Forests). In the default FVS-CR LP, a small initial proportion of aspen seedlings will eventually out-compete other species and comprise the dominant share of basal area at these sites under the default FVS-CR LP model calibration, an effect that manifested in our sensitivity analysis results. Unfortunately, such species composition shifts cannot be extracted from FIA data, and thus any future adjustments of aspen growth or mortality for more realistic behavior must rely on other sources of ecological knowledge.

Our sensitivity analysis showed zero response of stand growth to slope, aspect, and elevation, suggesting that the optional site index input is essential for representing site-level variation in productivity. In

addition, the FVS-CR LP model default value for SI (19.8 m at a base age of 100 years) is at the high end (~ 85 th percentile) of the range observed in our FIA dataset (which had a mean value of 16.0 m). This relatively high default SI value may have contributed to the over-estimation of stand growth. However, the over-estimation of tree height (Fig. 5C) was modest compared to the over-estimation of BA and live carbon density, suggesting that much of this problem is due to other factors (e.g., under-estimation of mortality in early stages). Splitting the FIA chronosequence data into low- and high-SI groupings also suggested that SI has very limited influence on carbon accumulation rate and most other stand characteristics besides height (see [supplementary materials](#)).

Because stand dynamics in FVS-CR are largely driven by individual tree growth, we believe the effort of validating the large-tree diameter increment model with equivalence testing was necessary. Equivalence testing showed statistically valid simulation of lodgepole pine diameter increment but revealed underestimation of subalpine fir and Engelmann spruce growth. Prediction errors for quaking aspen exhibited large variability over the range of diameter increments predicted, but systematically overestimated the increment growth. Pairwise observed-versus-predicted values for the four species were widely variable when plotted and compared against a 1:1 line. This was somewhat surprising considering that lodgepole pine passed the equivalence test (i.e., the test of dissimilarity was rejected) for both the intercept and slope, yet the relatively small model R^2 (0.11) indicated the model had little explanatory power for the factors influencing the diameter growth of individual trees. However, [Robinson et al. \(2005\)](#) noted that increasing the sample size will always have the effect of strengthening evidence of similarity with equivalence testing. Our findings suggest that a recalibration of the large-tree diameter increment model is necessary for the other species simulated in the FVS-CR LP model.

While FIA data is widely available for performing model validation in different forest types across the US, constructing a regional chronosequence from FIA plot data was somewhat database-intensive and highlighted a number of methodological challenges. For instance, the fact that so many of the stands coded as age zero had significant large tree presence suggests that many of the lodgepole stands were not evenaged and that the stand ages coded have limited value for young stands.

This finding led us to rely on a localized inventory of post-harvest regeneration on the CSF to determine realistic initial seedling densities. Additionally, such space-for-time approaches are not readily able to elucidate stand successional dynamics, e.g., stands that may have started out as lodgepole-dominated but eventually transitioned to spruce/fir could have been classified in FIA as spruce/fir and excluded from our analysis. Furthermore, the western US experienced an unprecedented beetle outbreak from 1997 to 2012 resulting in 5.2 million hectares (12.8 million acres) of insect-caused tree mortality (Hicke et al., 2016). We hypothesize that the effect of the outbreak is reflected in FIA chronosequence data by greater-than-average mortality, thereby requiring mortality rates in FVS-CR LP to be increased to better fit the new realities of forest condition in this region.

While these evaluation procedures were highly informative and worthwhile, there were aspects of applying them that we found challenging, and we have gained insights we hope will benefit others who apply this evaluation methodology. The Bakuzis matrix was straightforward to implement, as it consisted only of running different bare-ground stand regrowth simulations for sites of different SI (and an assumption of initial seedling density as required by the partial establishment model in this FVS variant) and plotting various model outputs on different axes. We recommend more widespread use of this model performance check when verifying other species models and variants, as it is well within the capabilities of most model users, having minimal computational requirements other than a reasonable estimate of initial seedlings.

While the chronosequence and equivalence testing exercises were instructive and serve as a good basis for future model recalibration efforts, others may find it more practical to work with existing forest growth curves derived from the synthesis of FIA data and simple growth models (Smith et al., 2006), rather than working with the raw FIA data itself. Neither the simulation of 124 individual stands in FVS, nor the processing of individual tree-level simulation results in a statistical computing environment (R) was particularly computationally-intensive. However, the FIA system is not yet well-integrated with FVS, and we spent considerable time formatting the raw FIA data to be suitable for batch simulation in FVS. For instance, model users should be aware that the lodgepole pine site index data reported by FIA uses a different base age (50 years) than the FVS-CR LP model (100 years), and a non-trivial effort is required to rectify this discrepancy. While we consider quantitative stand- and tree-level validation against FIA data as a valuable tool in the context of academic research projects and periodic high-level model evaluation exercises, the difficulty of properly formatting FIA data for use in FVS is an obstacle that only advanced users are likely to overcome. Similarly, sensitivity analysis is a tedious process that involves executing the model many times with small modifications. These sensitivity analysis results are a resource that anyone interested in better understanding and calibrating the FVS-CR LP can utilize, though those using different FVS-CR forest type models or different FVS variants would probably need to do their own sensitivity analysis that reflects the different equations and behaviors of those different models.

Thoughtful prioritization of data collection for both evaluating and parameterizing the FVS-CR LP model, and likely all FVS variants, is necessary. Post-harvest and disturbance regeneration data is vital for initializing the model. We found that the FIA data contains many young plots (aged 0–30 years) in lodgepole pine forests with high basal area, suggesting that these were in fact multi-aged stands. Therefore, we relied on additional inventory data collected at the CSF to initialize bare ground simulations, and we recommend that others carefully collect regeneration data including species tallies. Site index was another sensitive variable that should be collected wherever possible.

5. Conclusion

We proposed a framework for evaluating model behavior and

identifying components needing calibration, and applied it to the Lodgepole Pine model of the FVS Central Rockies variant—a model widely used for forest carbon accounting and other applications but that has not been previously evaluated in the literature. Qualitative verification using a Bakuzis matrix confirmed that the internal logic of the model was structurally sound and captured key expected relationships. However, validation of stand level growth against a chronosequence indicated problematic performance of the default model, which greatly over-estimated biomass accumulation, likely due to under-estimation of early stage mortality. Formal equivalence testing showed that the default lodgepole pine large tree diameter increment model was statistically significant but had weak explanatory power, but that the increments for other key species are under- or over-estimated. Sensitivity analysis provided insight into which parameters and inputs most strongly influenced model output, and highlighted how aspen presence drives long-term stand structure and model behavior. Other important factors include the large-tree fixed diameter increment multiplier, the maximum SDI setting of the mortality model, and the site index input. Our identification of sensitive parameters and inputs will hopefully aid land managers in their data collection efforts, and in future efforts to calibrate the FVS-CR LP model. This model evaluation framework identifies the limitations and strengths of both the individual model components and the outputs of the integrated system, better-informing future calibration efforts.

CRediT authorship contribution statement

Benjamin A. Bagdon: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Trung H. Nguyen:** Conceptualization, Investigation, Writing - original draft, Writing - review & editing. **Anthony Vorster:** Conceptualization, Investigation, Resources, Writing - original draft, Writing - review & editing. **Keith Paustian:** Funding acquisition, Project administration, Writing - review & editing. **John L. Field:** Conceptualization, Investigation, Project administration, Supervision, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dan Tinker, Chuck Rhoades, and Paul Hood of the Bioenergy Alliance Network of the Rockies for sharing their insights on lodgepole stand regeneration, aspen mortality, and FVS model calibration. We are grateful to Russ Gross and John Twitchell of the Colorado State Forest Service for the information they shared and for helpful discussions. We thank the two reviewers and the editor for their helpful comments which greatly strengthened this work. We also thank Aaron Gagnon, Chad Keyser, and Michael VanDyke of the US Forest Service Forest Management Service Center for their assistance and insights into our work. This work was supported by the National Institute of Food and Agriculture (NIFA) of the United States Department of Agriculture [grant numbers 2013-68005-21298, 2017-67019-26327, 2019-67019-29469]. These funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foreco.2020.118619>.

References

- Alexander, R.R., Tackle D., and W.G. Dahms. 1967. "Site Indicies for Lodgepole Pine with Corrections for Stand Density Methodology." Research Paper RM-29. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Augusiaik, J., Van den Brink, P.J., Grimm, V., 2014. Merging validation and evaluation of ecological models to 'Evaluation': A review of terminology and a practical approach. *Ecol. Model.* 280 (May), 117–128. <https://doi.org/10.1016/j.ecolmodel.2013.11.009>.
- Bagdon, B.A., Huang, C., Dewhurst, S., Sánchez-Medor, A., 2017. Climate change constrains the efficiency frontier when managing forests to reduce fire severity and maximize carbon storage. *Ecol. Econ.* 140 (October), 201–214. <https://doi.org/10.1016/j.ecolecon.2017.05.016>.
- Bagdon, B.A., Huang, C.-H., 2014. Carbon stocks and climate change: management implications in Northern Arizona Ponderosa Pine Forests. *Forests* 5 (4), 620–642. <https://doi.org/10.3390/f5040620>.
- Bakuzis, E.V., 1969. Forestry viewed in an ecosystem perspective. In: *The Ecosystem Concept in Natural Resource Management*. Elsevier, pp. 189–258.
- Bettinger, P., Merry, K., Mavity, E., Rightmyer, D., and R. Stevens. 2015. Chattahoochee-Oconee National Forest, Georgia, United States of America. In: Forest Plans of North America, 277–284. Elsevier. Doi: 10.1016/B978-0-12-799936-4.00031-X.
- Buma, B., Wessman, C.A., 2013. Forest resilience, climate change, and opportunities for adaptation: a specific case of a general problem. *For. Ecol. Manage.* 306 (October), 216–225. <https://doi.org/10.1016/j.foreco.2013.06.044>.
- Caldwell, M.K., Hawbaker, T.J., Briggs, J.S., Cigan, P.W., Stitt, S., 2013. Simulated impacts of mountain pine beetle and wildfire disturbances on forest vegetation composition and carbon stocks in the southern rocky mountains. *Biogeosciences* 10, 8202–8222. <https://doi.org/10.5194/bg-10-8202-2013>.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environ. Modell. Software* 22 (10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>.
- Canavan, S.J., Ramm, C.W., 2000. Accuracy and precision of 10 year predictions for forest vegetation simulator-lake states. *North. J. Appl. For.* 17 (2), 9. <https://doi.org/10.1093/njaf/17.2.62>.
- Cawse, D., Keyser, C., Keyser, T., Sanchez-Medor, A., Smith-Mateja, E., Van Dyck, M., 2010. Forest vegetation simulator model validation protocols. United States Department of Agriculture, Forest Service.
- Collins, B.J., Rhoades, C.C., Battaglia, M.A., Hubbard, R.M., 2012. The effects of bark beetle outbreaks on forest development, fuel loads and potential fire behavior in salvage logged and untreated lodgepole pine forests. *For. Ecol. Manage.* 284 (November), 260–268. <https://doi.org/10.1016/j.foreco.2012.07.027>.
- Collins, B.J., Rhoades, C.C., Hubbard, R.M., Battaglia, M.A., 2011. Tree regeneration and future stand development after bark beetle infestation and harvesting in Colorado Lodgepole Pine Stands. *For. Ecol. Manage.* 261 (11), 2168–2175. <https://doi.org/10.1016/j.foreco.2011.03.016>.
- Crookston, N.L., 2014. Climate-FVS Version 2: content, users guide, applications, and behavior. RMRS-GTR-319. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Doi: 10.2737/RMRS-GTR-319.
- Crookston, N.L., Dixon, G.E., 2005. The forest vegetation simulator: a review of its structure, content, and applications. *Comput. Electron. Agric.* 49 (1), 60–80. <https://doi.org/10.1016/j.compag.2005.02.003>.
- Diaz, D., Loreno, S., Ettl, G., Davies, B., 2018. Tradeoffs in timber, carbon, and cash flow under alternative management systems for douglas-fir in the Pacific Northwest. *Forests* 9 (8), 447. <https://doi.org/10.3390/f9080447>.
- Dixon, G.E., 2002. Essential FVS: A User's Guide to the Forest Vegetation Simulator, U.S. Department of Agriculture, Forest Service, Internal Rept. Revised September, 2018: 244 pp.
- Edminster, C.B., Mowrer, H.T., Mathiasen, R.L., Olsen, W.K., Hawksworth, F.G., 1991. GENGYM: a variable density stand table projection system calibrated for mixed conifer and ponderosa pine stands in the southwest. Research Paper. RM-297. Fort Collins, CO: U.S. Department of Agriculture, Forest Service.
- EVALIDator Version 1.8.0.00. 2019. Forest Inventory EVALIDator Web-Application Version 1.8.0.00. April 19, 2019. <http://apps.fs.fed.us/Evaluator/evaluator.jsp>.
- Ex, S.A., Smith, F.W., 2014. Evaluating forest vegetation simulator performance for trees in multiaged ponderosa pine stands, Black Hills, USA. *Forest Sci.* 60 (2), 214–221. <https://doi.org/10.5849/forsci.12-054>.
- Ex, S.A., Smith, F.W., Keyser, T.L., Rebain, S.A., 2016. Estimating canopy bulk density and canopy base height for interior Western US conifer stands. *Forest Sci.* 62 (6), 690–697. <https://doi.org/10.5849/forsci.15-118>.
- Finney, M.A., Selig, R.C., McHugh, C.W., Ager, A.A., Bahro, B., Agee, J.K., 2007. Simulation of long-term landscape-level fuel treatment effects on large wildfires. *Int. J. Wildland Fire* 16 (6), 712. <https://doi.org/10.1071/WF06064>.
- Froese, R.E., Robinson, A.P., 2007. A validation and evaluation of the prognosis individual-tree basal area increment model. *Can. J. For. Res.* <https://doi.org/10.1139/X07-002>.
- Fulé, P.Z., Cocke, A.E., Heinlein, T.A., Covington, W.W., 2004. Effects of an intense prescribed forest fire: is it ecological restoration? *Restor. Ecol.* 12 (2), 220–230. <https://doi.org/10.1111/j.1061-2971.2004.00283.x>.
- Ginrich, S.F., 1967. Measuring and evaluating stocking and stand density in upland hardwood forests in the Central States. *Forest Sci.* 13 (1), 38–53. <https://doi.org/10.1093/forestscience/13.1.38>.
- Havis, R.N., 2017. The FVS-WRENSS water yield post-processor: validation of snow-dominated procedures. In: Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. GTR-SRS-224. United States Department of Agriculture, Forest Service, Southern Research Station, 24–33.
- Hicke, J.A., Meddens, A.J.H., Kolden, C.A., 2016. Recent tree mortality in the Western United States from bark beetles and forest fires. *Forest Sci.* 62 (2), 141–153. <https://doi.org/10.5849/forsci.15-086>.
- Hill, W., 1997. Using FVS to prepare the custer state park resource management plan. In: *Proceedings: Forest Vegetation Simulator Conference*. INT-GTR-373. Fort Collins, CO: United States Department of Agriculture, Forest Service, Intermountain Research Station, 79–85.
- Jørgensen, S.E., 2011. Fundamentals of Ecological Modelling: Applications in Environmental Management and Research. 4th ed. Vol. 21. *Developments in Environmental Modelling*. Elsevier.
- Jørgensen, S.E., Fath, B.D., 2011. Concepts of modelling. In: *Developments in Environmental Modelling*, 23, 19–93. Elsevier. Doi: 10.1016/B978-0-444-53567-2.00002-8.
- Keyser, C., Dixon, G.E., 2008. Central rockies (CR) variant overview. United States Department of Agriculture, Forest Service, Internal Rept. Fort Collins, CO. Revised: 2014. 70 pp.
- Keyser, C., Keyser, T., 2017. Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. E-Gen. Tech. Rep. SRS-224. United States Department of Agriculture, Forest Service, Southern Research Station, Asheville, N.C.
- Klutsch, J.G., Battaglia, M.A., West, D.R., Costello, S.L., Negron, J.F., 2011. Evaluating potential fire behavior in lodgepole pine-dominated forests after a mountain pine beetle epidemic in North-Central Colorado. *West. J. Appl. For.* 9. <https://doi.org/10.1093/wjaf/26.3.101>.
- Lacerte, V., Larocque, G.R., Woods, M., Parton, W.J., Penner, M., 2004. Testing the Lake States Variant of FVS (Forest Vegetation Simulator) for the main forest types of Northern Ontario. *Forest Chronicle* 80 (4), 495–506. <https://doi.org/10.5558/tfc80495-4>.
- Leary, R.A., 1997. Testing models of unthinned red pine plantation dynamics using a modified Bakuzis matrix of stand properties. *Ecol. Model.* 98 (1), 35–46. [https://doi.org/10.1016/S0304-3800\(96\)01935-7](https://doi.org/10.1016/S0304-3800(96)01935-7).
- Leites, L.P., Robinson, A.P., Crookston, N.L., 2009. Accuracy and equivalence testing of crown ratio models and assessment of their impact on diameter growth and basal area increment predictions of two variants of the forest vegetation simulator. *Can. J. For. Res.* 39 (3), 655–665. <https://doi.org/10.1139/X08-205>.
- Maffei, H., Tandy, B., 2001. Methodology for modeling the spatial and temporal effects of vegetation management alternatives on late successional Habitat in the Pacific Northwest. In: Second Forest Vegetation Simulator Conference; 2002 February 12–14; Fort Collins, CO, 9.
- Martinez, David S., 2008. United States Forest Service (USFS) NRIS-INFORMS-Fire/Fuels Analysis Tool. United States Department of Agriculture, Forest Service, Pacific Southwest Research Station. 181–184. In: *Proceedings of the second international symposium on fire economics, planning, and policy: a global view*.
- McKenzie, Patrick F., Duveneck, Matthew J., Morreale, Luca L., Thompson, Jonathan R., 2019. Local and global parameter sensitivity within an ecophysiological based forest landscape model. *Environ. Modell. Software* 117 (July), 1–13. <https://doi.org/10.1016/j.envsoft.2019.03.002>.
- McMahan, A.J., Monahan, W.B., 2017. Estimating changes to forest structure as a result of forest pests: using FVS to simulate potential effects of emerald ash borer across a broad landscape. In: *Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference*. United States Department of Agriculture, Forest Service, Southern Research Station, 149–151.
- Pokharel, B., Froese, R.E., 2008. Evaluating alternative implementations of the lake states FVS diameter increment model. *Forest Ecol. Manage.* 255 (5–6), 1759–1771. <https://doi.org/10.1016/j.foreco.2007.11.035>.
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rebain, Stephanie, 2010. The fire and fuels extension to the forest vegetation simulator: updated model documentation. United States Department of Agriculture, Forest Service, Internal Rept. Fort Collins, CO. Revised: 2014, 409 pp.
- Robinson, A.P., Duursma, R.A., Marshall, J.D., 2005. A regression-based equivalence test for model validation: shifting the burden of proof. *Tree Physiol.* 25 (7), 903–913. <https://doi.org/10.1093/treephys/25.7.903>.
- Robinson, A.P., 2016. Equivalence: Provides Tests and Graphics for Assessing Tests of Equivalence. (version 0.7.2). <https://CRAN.R-project.org/package=equivalence>.
- Robinson, A.P., Froese, R.E., 2004. Model validation using equivalence tests. *Ecol. Model.* 176 (3), 349–358. <https://doi.org/10.1016/j.ecolmodel.2004.01.013>.
- Rodrigue, J.A., Keyser, C.E., Nowak, J.T., 2017. "Use of the Forest Vegetation Simulator and the Southern Pine Beetle Event Monitor to Identify Silvicultural Treatments for the Reduction of Southern Pine Beetle Hazard and Enhancement of Restoration on the North Carolina Piedmont." In: *Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference*. GTR-SRS-224. United States Department of Agriculture, Forest Service, Southern Research Station, 140–148.
- Russell, M.B., Weiskittel, A.R., Kershaw, J.A., 2013. Benchmarking and calibration of forest vegetation simulator individual tree attribute predictions across the Northeastern United States. *North. J. Appl. For.* 30 (2), 75–84. <https://doi.org/10.5849/njaf.12-034>.
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecol. Model.* 90 (3), 229–244. [https://doi.org/10.1016/0304-3800\(95\)00152-2](https://doi.org/10.1016/0304-3800(95)00152-2).
- Smith, J.E., Heath, L.S., Skog, K.E., Birdsey, R.A., 2006. "Methods for Calculating Forest Ecosystem and Harvested Carbon with Standard Estimates for Forest Types of the United States." NE-GTR-343. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. Doi: 10.2737/NE-GTR-343.
- Vanclay, J.K., Skovsgaard, J.P., 1997. Evaluating forest growth models. *Ecol. Model.* 98 (1), 1–12. [https://doi.org/10.1016/S0304-3800\(96\)01932-1](https://doi.org/10.1016/S0304-3800(96)01932-1).
- Woodward, B., Engelstad, P., Vorster, A., Beddow, C., Krail, S., Vashishth, A., Evangelista, P., 2017. Forest harvest dataset for northern Colorado Rocky Mountains (1984–2015) generated from a Landsat time series and existing forest harvest records. *Data in Brief* 15, 724–727. <https://doi.org/10.1016/j.dib.2017.10.030>.
- Wykoff, W.R., Crookston, N.L., Stage, A.R., 1982. User's Guide to the Stand Prognosis Model. INT-GTR-122. Ogden, UT: U.S. Department of Agriculture.