



# Assessing forest vegetation and fire simulation model performance after the Cold Springs wildfire, Washington USA

Susan Hummel<sup>a,\*</sup>, Maureen Kennedy<sup>b</sup>, E. Ashley Steel<sup>c</sup>

<sup>a</sup> USDA Forest Service, PNW Research Station, Portland, United States

<sup>b</sup> University of Washington, Seattle, United States

<sup>c</sup> USDA Forest Service, PNW Research Station, Seattle, United States

## ARTICLE INFO

### Article history:

Received 1 March 2012

Received in revised form 15 August 2012

Accepted 18 August 2012

Available online 22 October 2012

### Keywords:

Forest structure

Forest Vegetation Simulator (FVS) and FFE-FVS

Multi-criteria assessment

Pareto optimality

Fire behavior and effects

## ABSTRACT

Given that resource managers rely on computer simulation models when it is difficult or expensive to obtain vital information directly, it is important to evaluate how well a particular model satisfies applications for which it is designed. The Forest Vegetation Simulator (FVS) is used widely for forest management in the US, and its scope and complexity continue to increase. This paper focuses on the accuracy of estimates made by the Fire and Fuels Extension (FFE-FVS) predictions through comparisons between model outputs and measured post-fire conditions for the Cold Springs wildfire and on the sensitivity of model outputs to weather, disease, and fuel inputs. For each set of projected, pre-fire stand conditions, a fire was simulated that approximated the actual conditions of the Cold Springs wildfire as recorded by local portable weather stations. We also simulated a fire using model default values. From the simulated post-fire conditions, values of tree mortality and fuel loads were obtained for comparison to post-fire, observed values. We designed eight scenarios to evaluate how model output changed with varying input values for three parameter sets of interest: fire weather, disease, and fuels. All of the tested model outputs displayed some sensitivity to alternative model inputs. Our results indicate that tree mortality and fuels were most sensitive to whether actual or default weather was used and least sensitive to whether or not disease data were included as model inputs. The performance of FFE-FVS for estimating total surface fuels was better for the scenarios using actual weather data than for the scenarios using default weather data. It was rare that the model could predict fine fuels or litter. Our results suggest that using site-specific information over model default values could significantly improve the accuracy of simulated values.

Published by Elsevier B.V.

## 1. Introduction

Resource managers rely on computer simulation models when it is difficult or expensive to obtain vital information directly. In the United States, for example, a suite of simulation tools, including fire behavior models, are routinely used in support of wildfire management activities. Firefighter safety and incident containment objectives require that pertinent information on potential fire behavior be synthesized and made available immediately. Simulation models are also used to evaluate a variety of ecological phenomena and the effects of managing them, including the effects of disturbance on forest dynamics at varying spatial scales and in different forest types (Barreiro and Tomé, 2011); interaction of fire and insects on forest structure (Hummel and Agee, 2003; James et al., 2011); capacity of stream habitats to support salmonids

(Lichatowich et al., 1995), and tradeoffs in fire threat and late-seral forest structure (Calkin et al. 2005).

Given the widespread use of model output, it is responsible to assess model performance (Regan et al., 2002; Pielke and Conant, 2003). In other words, it is important to evaluate how well a particular model satisfies applications for which it is designed. We are mindful of the aphorism ‘all models are wrong, but some are useful’ (Box and Draper, 1987). Describing when a model falls short and when it performs satisfactorily is informative. Nonetheless, it is rarely done. The development of tools for model assessment has lagged behind model development (McElhany et al., 2010). One reason is a lack of data and another is a lack of resources. After a tool like a forest growth model or a fire simulation model is created, available resources are often invested in using the tool rather than in evaluating it. This is not surprising, perhaps, because the need to evaluate a large number of parameters and parameter combinations can make assessing model performance difficult and time-consuming. A single formula for analyzing model sensitivity does not exist. The same model, for example, might be evaluated in multiple ways to reflect multiple management needs (e.g., Steel

\* Corresponding author. Address: Portland Forestry Sciences Laboratory, USDA Forest Service, PNW Research Station, 620 SW Main Street, Suite 400, Portland, OR 97205, United States. Tel.: +1 503 808 2084.

E-mail address: [shummel@fs.fed.us](mailto:shummel@fs.fed.us) (S. Hummel).

et al., 2009). Complex simulation models can have linked components and a large number of internal parameters, even within one variant of the model. There may be interactions between parameters or inputs that are unknown. While an evaluation of the full potential model domain may be impossible, an evaluation of sub-domains is often possible. Of particular importance for evaluation are the model applications and outputs used in making decisions about forest management or the use of public funds.

One simulation model used widely in the US is the Forest Vegetation Simulator (FVS) (Dixon, 2010). The FVS model, which originated in part from an earlier forest growth model called “Prognosis” (Stage, 1973), is publicly available and maintained by employees of the USDA Forest Service (Crookston and Dixon, 2005; Forest Management Service Center (FMSC), 2011a). There are protocols for evaluating FVS that recommend sensitivity analysis prior to validating model performance (Cawrse et al., 2010). Evidence is scarce, however, that these protocols are routinely followed to evaluate the full scope of the model. Instead, analyses that use FVS for other purposes have yielded insight into model performance (e.g. Ager et al., 2007, Johnson et al., 2011) or model performance has been evaluated for a limited domain (Hood et al., 2007). Training on FVS is given annually around the country to audiences that include the staff of state, tribal, and federal resource management agencies (FMSC, 2011b). Such training can be important, because regional variants of the model exist and calibration for local conditions is often necessary. The demand for training also suggests the ongoing addition of new FVS users. Furthermore, the scope and complexity of the FVS model continues to expand. In the decades since the original base model was created, different extensions have been developed, including western root disease (RD) (Frankel tech co-ord, 1998), fire (FFE) (Rebain comp, 2010), western spruce budworm (Crookston et al., 1990), dwarf mistletoe (DM) (David, 2005). Outputs of the FVS model and its extensions are used in planning, management decisions, and research and as input to other analysis tools (e.g. Dixon, 2010). Despite its widespread use, model performance is not well understood and is infrequently documented.

Methods such as sensitivity analysis and multi-criteria analysis can aid in describing the domain over which a model is adequate (e.g., Reynolds and Ford, 1999). To be most helpful a model assessment should reflect how a model is actually used. Both the FVS base model and FFE-FVS are used in an array of management decisions. Modeled estimates of tree basal area and mortality derived from FVS have been used to guide silvicultural activities or anticipate their effects, for example, and model output has been used to select and document management choices among different alternatives (e.g. USDA, 2003, 2010). When combined with information about tree size and species killed or injured by fire, knowledge about mortality levels can affect decisions about reforestation (e.g. how many seedlings and of which species to order), salvage (e.g. how much merchantable volume deteriorating how quickly), and dead tree management (e.g. retain for habitat or fell for safety). Modeled estimates of forest fuel by size class are also used in both pre- and post-fire decisions because they affect fire behavior which, in turn influences both firefighter safety and post-fire forest conditions. If multiple model outputs are routinely considered in making decisions about forest management, then a model performance assessment should include them. Indeed, in complex biological systems like forests we think it can be more informative to compare modeled outputs of key variables simultaneously rather than individually (Kennedy and Ford, 2011).

In our study area FVS, together with its disease and fire extensions, has been used in previous management and research applications because the structure of the models and the input requirements matched project needs (Hummel and Barbour, 2007). Specifically, FVS can recognize the contribution of individual

trees to forest structure at both within-stand and among-stand (landscape) scales and can track residual stand structure following silvicultural treatments or disturbance (e.g., Crookston and Stage, 1991). The model accounts for harvested trees by size and species and can be linked to a spatial database. As well, it is publicly available and nationally supported by the USDA Forest Service (Hummel and Barbour, 2007). Data that form the starting point of this analysis were originally used by the local Ranger District to evaluate alternative management scenarios and to select a preferred action for forest density management and fuels reduction activities (USDA, 2003).

Our analysis contributes to the literature on performance of the FFE-FVS model. We identified a unique opportunity to assess the performance of one variant of the model when a site for which vegetation and fuels data were collected in 2001 (Hummel and Calkin, 2005) was subsequently burned by the Cold Springs wildfire in 2008. The fire originated from a lightning strike near the eastern boundary of the Gifford Pinchot National Forest (GPNF) in Washington State. Due to its size (>3300 ha) and because it burned across multiple ownerships, extensive information was generated during the fire, including daily weather records and observations, that enabled assessments of model performance with actual weather conditions.

We sought to understand how stand-scale predictions of post-fire tree mortality generated by FFE-FVS compared to observed post-fire tree mortality within the Cold Springs fire perimeter. We focused on the stand-scale because it is the spatial scale at which FVS is commonly run and its outputs used to inform management planning on public land. We had two goals in our analysis, namely to evaluate model accuracy across suites of model outputs and to evaluate the sensitivity of the model to weather, disease, and fuel inputs. In particular, we were interested in the accuracy of tree mortality and fuel load estimates. The innovations of the analysis are (1) the comparison of post-fire field measurements to model simulations based on pre-fire data for the same site and (2) the simultaneous evaluation of multiple model outputs.

## 2. Methods

### 2.1. Site description

The study area is located in the Cascade Range of Washington (WA) and covers the southeastern flank of Mt. Adams (T.7 N., R. 10 E) at the eastern border of the GPNF (Fig. 1). Elevation ranges from about 1000–2000 m and annual precipitation from 90–165 cm (Topik, 1989). Forests are predominantly within the grand fir (*Abies grandis*) series (Franklin and Dyrness, 1988), with higher elevations in mountain hemlock (*Tsuga mertensiana*) (Diaz et al., 1997). An administrative unit, the Gotchen late-successional reserve (Gotchen LSR), was established in the area as part of the Northwest Forest Plan (USDA and USDI 1994) (Fig. 1). In 2001 aerial photos were used to delineate and classify the ca. 6000 ha forest reserve into 159 polygons, or stands. Approximately seventy percent of the area was typed as Douglas-fir/grand fir (*Pseudotsuga menziesii*/*A. grandis*) (Hummel and Calkin, 2005). In contrast, at elevations above about 1375 m subalpine fir predominates.

In June 2008, a fire was ignited by lightning at approximately 1300 m in the northwestern area of the reserve. It smoldered until it was located by district patrols on the morning of July 13. Within 12 h, the locally named “Cold Springs” fire, which was moving due east/southeast, had increased in size to over 1600 ha and, within 24 h, the perimeter had nearly doubled. By the time containment was declared in September 2010, approximately 3600 ha had burned. The fire originated on federal land, which also accounted for the largest area burned among all the affected land ownership

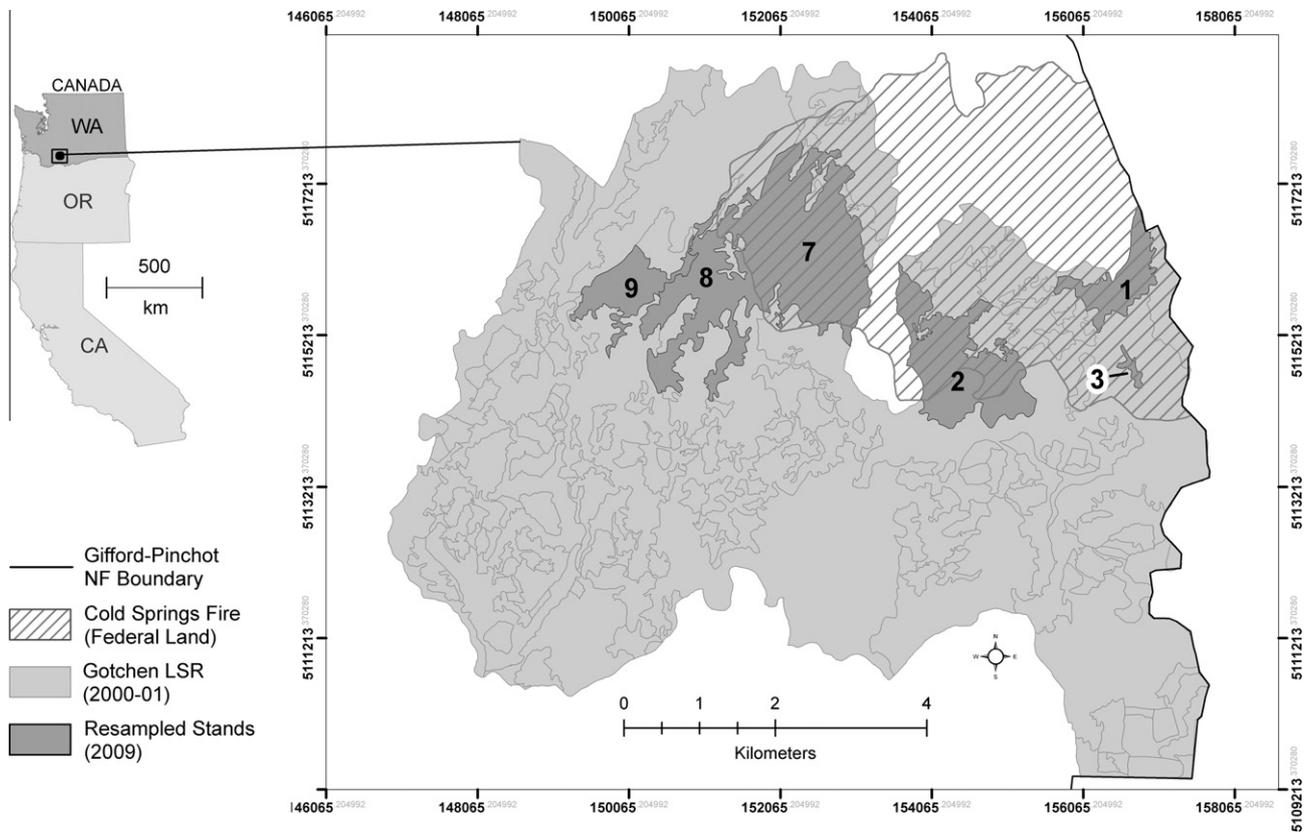


Fig. 1. Map of the study area. The numbered polygons represent the six stands in which pre-fire (2001) and post-fire (2009) vegetation and fuels data were collected.

classes (Fig. 1). The data used in this analysis are from federal land within the Cold Springs fire.

The burned area is a swath oriented from west-to-east within an elevation band of ca 1170–1816 m. The lower portion of this band, in the eastern Cascades grand-fir zone (1100–1500 m) (Franklin and Dyrness, 1988), has a fire regime that is intermediate between those of upper and lower elevation forests. A variable severity regime helps maintain the dominance of grand fir associates like lodgepole pine (*Pinus contorta*), Douglas-fir (*Pseudotsuga menziesii*), or ponderosa pine (*Pinus ponderosa*) (Agee, 1993). Root pathogens and insects contribute to the patchy structure of forests in the grand-fir zone. Above 1500 m, in the subalpine forest, fire is the primary large-scale disturbance agent. Fires are often weather-driven, which complicates estimating fire behavior (Agee, 1993). Many fires are stand replacement events, however, because major tree species (subalpine fir: *Abies lasiocarpa* and lodgepole pine) lack resistance to high heat.

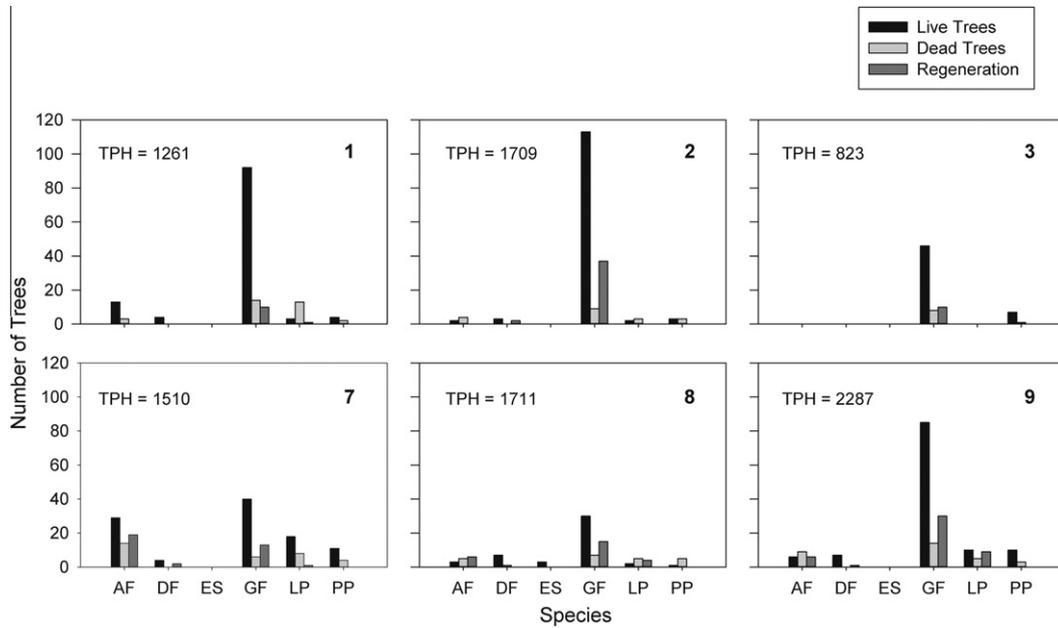
## 2.2. Pre-fire field sampling (2001)

In 2001, measurements of forest stand structure and composition were made in the Gotchen LSR (details in Hummel and Calkin (2005)). Twelve stands were selected for field sampling by using probability proportional to size; the selected polygons were visited between June–August 2001. A systematic sample of vegetation was made by following stand exam protocols from Region 6 of the Forest Service and incorporating a random starting location within each sampled stand. The allowable error in measurements ranged from 0% to 10%, depending on the variable. In addition, estimates of downed wood were made on each plot at the request of the wildlife biologist and the fire management officer on the Mt. Adams Ranger District of the Gifford Pinchot National Forest.

Two methods were used to estimate downed wood on each plot. First, a photo series code was recorded that corresponded best to the visual condition of the plot and then the code was converted directly to 1, 10, 100, and 1000 h fuel classes (Maxwell and Ward, 1980). Additionally, measurements were recorded for all logs >12.7 cm diameter that intersected a 22.8 m down wood transect established from plot center using a random compass azimuth (after Bull et al. (1997)). These additional “tally” data on 1000-h coarse woody debris were converted to metric tons/hectare by using equations in Waddell (2002) together with species-specific wood density values in FFE-FVS and decay-class adjustment factors for softwoods (after Bull et al. (1997)). The original sample plot locations were mapped, but not geo-referenced.

## 2.3. Post-fire field sampling (2009)

The six 2001 stands whose boundaries intersected with, or were contained by, the Cold Springs fire perimeter were visited again a year after the fire origin (July 2009) (Fig. 1). In all six stands the sample of the pre-fire forest was dominated by grand fir (Fig. 2, Table 1); average tree diameter (Fig. 3) and quadratic mean diameter (QMD) were approximately 55 cm (20”) (Table 1). The QMD of ponderosa pine and Douglas-fir trees in the sample was generally larger than this average, which is consistent with forest stand dynamics in a mixed-severity fire regime. The vegetation and fuels were re-sampled by using the same measurement protocols and field personnel. Five of the stands had 20 sample plots and one stand had 10 sample plots. The location of specific plots differed from the 2001 measurement, but the same number of plots were measured in each stand so their density was the same in each sample. We refer to the 2009 measured field data as the “post-fire” observed data. Most of the re-measured stands lie within the *Abies*

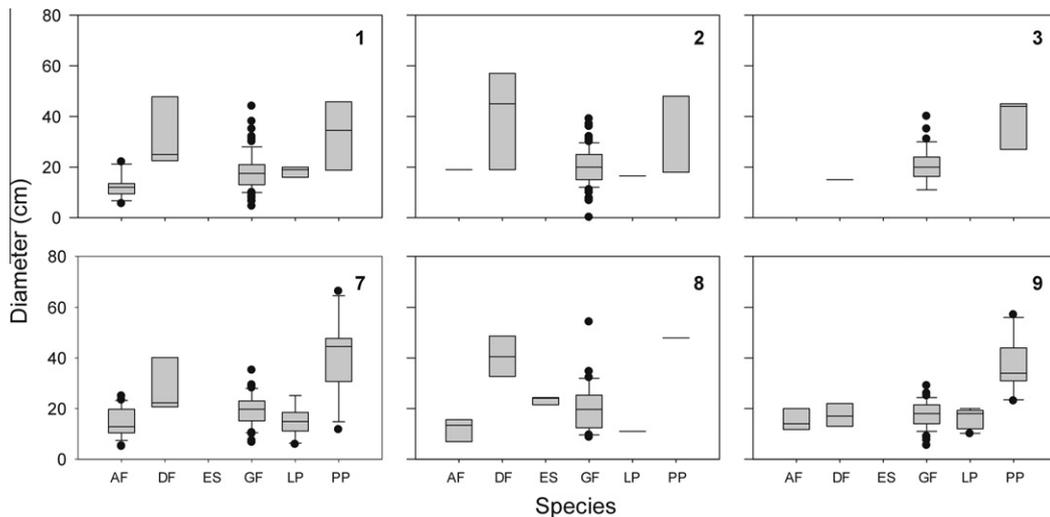


**Fig. 2.** The number of trees (live, dead, and advance regeneration <12 cm) by species in the six polygons shown in Fig. 1. Data represent pre-fire (2001) forest conditions. The trees per hectare (TPH) value was estimated by FVS using the sample data.

**Table 1**

Pre-fire conditions in each of the six polygons. Stands 1–9 correspond to the six polygons shown in Fig. 1. BA = basal area; n = measured trees; QMD = stand QMD (cm). AF = subalpine fir; DF = Douglas-fir; ES = Englemann spruce; GF = grand fir; LP = lodgepole pine; and PP = ponderosa pine.

Stand	BA (m <sup>2</sup> /ha)	n	QMD (cm)	AF QMD (cm)	% Of stand	DF QMD (cm)	% Of stand	ES QMD (cm)	% Of stand	GF QMD (cm)	% Of stand	LP QMD (cm)	% Of stand	PP QMD (cm)	% Of stand
1	24	116	51	33	11	88	3	–	–	49	79	47	3	90	3
2	33	123	58	48	2	110	2	–	–	55	92	43	2	103	2
3	17	56	62	–	–	38	2	–	–	55	86	–	–	101	13
7	26	102	57	39	28	75	4	–	–	52	39	41	18	110	11
8	16	46	67	32	7	106	15	59	7	57	65	29	4	122	2
9	26	118	53	44	5	49	6	–	–	47	72	43	9	97	9



**Fig. 3.** The diameter distribution of all trees measured (excluding advanced regeneration) in the six sample polygons shown in Fig. 1. Data represent pre-fire (2001) forest conditions.

*grandis*/*Carex geyeri* (grand fir/elk sedge) plant association (five of six stands). The remaining stand is in the *Tsuga mertensiana* (mt. hemlock/big huckleberry/beargrass) plant association.

#### 2.4. Simulating fire effects using FFE-FVS

The pre-fire data were formatted for use in the East Cascades (EC) variant of the Forest Vegetation Simulator (FVS) (Stage, 1973). The FVS model is frequently used deterministically (Hamilton, 1991), which is the default setting. In the deterministic setting multiple runs with the same tree list will generate the same results. However a single FVS projection is only one of several possible outcomes for the growth and corresponding structural and compositional development of a forest stand. To explore the effects of variation within FVS on the results it simulates, the model can instead be run in a stochastic mode. This is done by seeding a random number into each simulation run by using the “RANNSEED” keyword. Random effects are then incorporated in the FVS model via the distribution of errors associated with the prediction of the logarithm of basal area increment (Hoover and Rebain, 2010). In the EC variant, including random effects in a simulation alters the model equations for diameter growth and crown ratio. The model equation for height growth is not directly affected although indirect effects on height growth are possible. By using RANNSEED and analyzing the same simulation under different random numbers, it is possible to report output distributions for any metric estimated by FVS. For this analysis we ran the model in a stochastic mode.

The distribution of our simulated data tended to be skewed and non-normally distributed; no data transformations remedied the distributions. In such an instance, statistical measures such as the mean and confidence interval are problematic. We therefore chose non-parametric measures based on ordered statistics for our summary and analysis. The median simulated value is a measure of central tendency that is less sensitive to skewed data and outliers than the mean value. The interquartile range (IQR) is a non-parametric measure of variability in a distribution. For example, McGarvey et al. (2010) used the interquartile range of simulated values for a fish density model to compare simulated distributions to observed fish densities. By using ordered statistics we no longer relied on distributional assumptions that were violated by the simulated data. This less powerful test makes our comparisons between simulated and observed data more conservative than if distributional statistics were used.

To use the FVS model, information about the initial condition of a stand must first be formatted into a tree list. A tree list is created from a sample of stand conditions and is the basis for projecting the development of vegetation in FVS. At a minimum, the diameter and species of each tree on a sample plot are required. If more detailed information such as tree height is available it can also be

used (Dixon, 2010). Model users typically create tree lists from information collected on many sample plots within a stand during forest management planning; these “stand exam” data are then combined to create a file that represents a composite condition with respect to tree size and species (Dixon, 2010). Hence, at least two sources of error exist in FVS projections: variation in the equations and parameters that comprise the model and variation in the sample data used to build the tree lists (Gregg and Hummel, 2002). Additional sources of variation are added with each model extension, as a result of both new input data and additional internal parameters.

We used the 2001 pre-fire data to estimate tree mortality and fuel loads in the six re-measured polygons by using FVS together with disease and fire extensions. The pre-fire tree lists were first projected for seven years (2001–2008) by using the East Cascades variant of FVS and different sources of fuel and disease information (Fig. 4). After FVS predicts values for forest growth and mortality and before it applies the values to tree records, an extension like the western root disease model simulates the dynamics of disease and then adjusts the predicted growth and mortality values accordingly (Frankel tech co-ord, 1998, David, 2005).

For each set of projected, pre-fire stand conditions, a fire was simulated in 2008 by using FFE-FVS. The simulated fire approximated the actual conditions of the Cold Springs fire as recorded by the local portable weather station (RAWS) on July 14. This was the date on which the fire expanded significantly via an eastward run across the federal land which comprised the sample stands. We used the maximum temperature value in late afternoon (14.00–16.00 h) and the maximum wind speed value, adjusted for the height of the weather station. As described below, we also simulated a fire using model default values. The FFE-FVS extension uses the FVS tree list and information on stand history and cover or habitat type to select one (or more) fuel models that represent fuel conditions (Rebain comp, 2010). The east Cascades variant uses the Anderson (1982) fire behavior fuel models. In this analysis, fuel model selections were made by allowing FFE-FVS to determine a weighted set, which included number 8 (closed timber litter) and 10 (timber litter/understory) (Rebain comp, 2010). The selected fuel model was then used to predict fire behavior. From the simulated post-fire conditions, values of tree mortality and fuel loads were obtained for comparison to the post-fire, observed values by using the same methods (Fig. 4).

We designed eight scenarios to evaluate how model output changed with varying input values for three parameter sets of interest: fire weather, disease, and fuels (Table 2). The scenarios included initial conditions that represented combinations of actual weather versus default weather, actual disease versus no disease, and fuel loads estimated either from both quick (photo series) and measured (transect “tally”) sources or from the tally alone. A three-digit code for each scenario identified which weather,

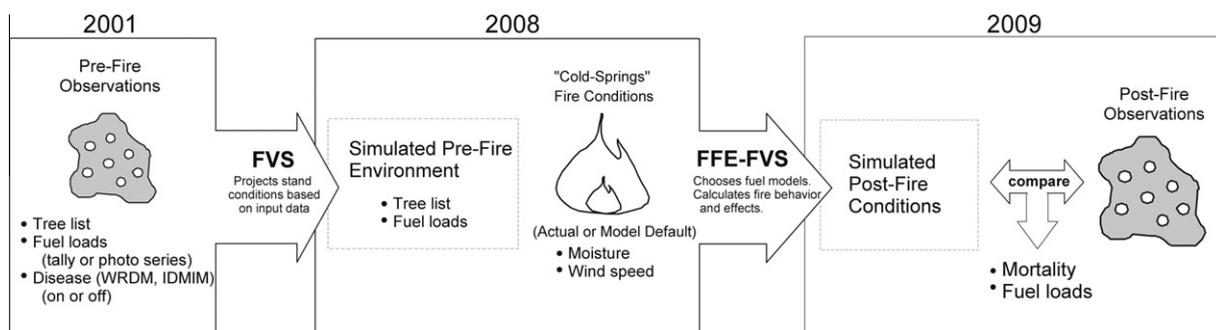


Fig. 4. Flow chart of the process for using 2001 sample data, together with FVS and its extensions, to generate estimates of tree mortality and fuel loads for comparison with post-fire field observations.

**Table 2**

Scenario design and model evaluation results for model outputs describing tree mortality and fuels. Scenarios designed to evaluate how model output changed with three types of inputs: fire weather (W), disease (D), and fuels (F). The first four scenarios, names begin with “W”, used weather data recorded during the Cold Springs fire. The other four scenarios, names begin with a 0, used the default settings for weather in the East Cascades variant of the FFE-FVS model. DM = Dwarf mistletoe; RD = Root disease. Photo indicates that a photo series code was recorded to describe fuel conditions. Tally indicates that estimates of fuel from individual logs along a transect were also used as model inputs. Results for model outputs describing tree mortality are the number of stands which met the binary selection criteria for mortality outputs, summarized as either basal area or as trees per acre (tpa). Results for model outputs describing fuels are the number of stands for which met the binary selection criteria for fuels of different size classes and for total fuels.

Scenario design													
Scenario name	Weather		Disease		Fuels		Tree mortality			Fuels			
	Moisture	Wind speed	Mistletoe (DM)	Root disease (RD)		Tally	Mortality (basal area)	Mortality (tpa)	Pareto rank	10-h 0.25–1"	100-h 1–3"	Litter	Total
WDF	Dry	11 mph	On	On	Photo	Tally	1	1	2	0	3	0	3
WD0	Dry	11 mph	On	On	Photo		2	0	2	0	2	1	2
W0F	Dry	11 mph	Off	Off	Photo	Tally	2	1	1	0	2	0	3
W00	Dry	11 mph	Off	Off	Photo		2	0	2	0	2	1	2
0DF	Very dry	20 mph	On	On	Photo	Tally	1	1	2	0	0	0	0
0D0	Very dry	20 mph	On	On	Photo		1	1	2	0	1	0	0
00F	Very dry	20 mph	Off	Off	Photo	Tally	1	1	2	0	1	0	0
000	Very dry	20 mph	Off	Off	Photo		1	1	2	0	1	0	0

disease, or fuel information was included; a “W” for actual weather, otherwise “0”; a “D” for disease included, otherwise “0”; and a “F” for fuels from both photo series and tally estimates, otherwise “0” for photo alone. Hence, the “WDF” code was actual weather, disease included, and fuel estimates derived from the photo series and augmented with the tally data, the “0D0” code was default weather, disease included, and fuel estimates from photo only, the “W00” code was actual weather, disease omitted, and fuel estimates from photo only. Table 2 depicts all combinations made by following this established coding system.

Four of the scenarios were simulated with wind speed (20 mph) and fuel moisture values that would have been selected as the default values for fire and vegetation conditions in the east Cascades variant of FFE-FVS. The remaining four scenarios were simulated using wind speed (11 mph) and fuel moisture values derived from the actual conditions recorded by the RAWs station located near the fire on July 14, 2008. Within each group of actual weather and default weather scenarios we also varied both the disease levels and source data of fuels. Infection levels of root disease (RD) (*Armillaria spp.*, *Phellinus spp.*) and dwarf mistletoe (DM) (*Arceuthobium spp.*) were calibrated by including information recorded for each sampled polygon in 2001 (Hummel and Calkin, 2005) or by excluding RD and DM from the FVS simulations. Sampled RD was severe (level 4), but DM infection was light (<25% trees per acre (TPA)). Site calibration values of fuel loads in FVS were done in two ways. For one set of scenarios, the tons/ha estimates of all fuels by size class from the photo series code (Maxwell and Ward, 1980) were augmented with the tons/ha estimates of 1000-h fuels from the “tally” method. In the other, the photo series conversion alone was used for estimating fuel load in the pre-fire environment (Fig. 4).

Once each of the eight scenarios was developed by using customized keyword files and output tables, the FFE-FVS model was run for each of the six stands. We used the RANNSEED keyword and ran the model 500 times for each of the six stands and eight scenarios to generate the median and IQR of all model outputs of interest: tree mortality, expressed both as basal area (BA) and TPA, dead TPA by size class and species, surface fuels by size class, crown characteristics, torching and crowning indices, canopy cover, merchantable cubic foot volume, and stand density index. A subset of these variables became our model assessment criteria. To create comparable variables for the 2009 observed data, we used FFE-FVS and the same customized keyword files and output tables. We assumed that these 2009 data represent true post-fire conditions.

## 2.5. Analyzing and ranking multiple model runs

### 2.5.1. Model sensitivity

We examined the sensitivity of multiple model outputs to alternative types of input data. Specifically, we made three comparisons of model performance: (1) default weather versus actual weather; (2) fuel transects + photo series (“tally”) versus photo series alone (“photo”); and (3) disease (DM and RD) presence or absence (Table 2). Including all possible combinations of weather, disease severity, and fuel was outside the scope of this analysis.

Our evaluation included two types of model output, which we termed either “measured” or “derived.” The former term refers to model output that was the same or similar to variables measured in the field. Examples of measured variables included the number of live grand fir trees 10–20" dbh or dead ponderosa pine trees 20–30" dbh. In contrast, derived variables refer to model output not directly measured, but instead calculated by the model based on measured input. Examples of derived criteria included merchantable volume of wood in cubic feet and torching index (the wind speed required for a surface fire to ignite the crown layer).

FFE-FVS model output was generated for the same set of six measured and seven derived variables for each of the eight scenarios. For each variable of interest we then compared all 28 possible pairs of scenarios. If, for a given pair of scenarios, the IQR of a specific variable did not overlap, we considered the model sensitive to the differences between the inputs used for the two scenarios. For example, we may have compared scenarios that differed only in the use of actual versus default weather conditions. If the IQR for ‘simulated torching index’ did not overlap across these two scenarios, we concluded that ‘simulated torching index’ was sensitive to weather parameters. The total sensitivity of each variable was calculated as the proportion of scenario pairs for which their IQR did not overlap. A value of 1 means that the model output was sensitive for all scenario pairs; whereas a value of 0 means that the model output was not sensitive in any of the scenario pairs.

### 2.5.2. Model accuracy

We were interested not just in how the model responded to different inputs (sensitivity), but also in which inputs produced output that adequately matched our observed data (accuracy). For the accuracy assessment we evaluated how stand-scale model predictions of post-fire tree mortality and fuel loads compared to tree mortality and fuel loads sampled after the fire in the same burned stands. Binary criteria have been recommended for such

an evaluation (Hornberger and Cosby, 1985, Reynolds and Ford, 1999) rather than minimization of error between the model output and some target value. To create binary criterion a target range must be chosen in advance for a given variable. If a corresponding model output value falls within the target range the model is assumed to satisfy that criterion (it adequately replicates that variable; it is given a 1). If the model output value falls outside of the target range then the model is assumed not to satisfy that criterion (the model is deficient with respect to that variable; it is given a zero). We elected to use binary selection to compare model output for multiple variables simultaneously. As our target range we chose the IQR, a non-parametric estimate of variability across plots (for normally distributed data with a sample size of 10 this would be equivalent to a 95% confidence interval). To assess model adequacy, we checked whether the median simulated value for the stand fell within the inter-quartile range of plot-scale values (Fig. 5). There is no existing standard for evaluation; our criterion answered whether or not the simulated value fell within what we considered to be a “reasonable” range of observed values for

each stand. We defined a performance level based on the number of stands that satisfied the evaluation criteria for a particular model output: <3 stands = poor; 3–4 stands = fair; 5–6 stands = good.

2.5.3. Simultaneously evaluating multiple output metrics

Many model evaluation techniques use a single value (or criterion), such as a goodness-of-fit statistic, to measure model performance. These single-variable techniques have been adapted to accommodate multiple model outputs by combining the multiple measures of model fit into a single value, usually a weighted sum (i.e. select the model scenario that minimizes the sum of goodness-of-fit statistics). Weighting across measures can be arbitrary, however, and when multiple measures of model fit are combined the information each measure contributes individually to model performance is obscured.

We chose to evaluate FFE-FVS with respect to Pareto optimality. By using the concept of Pareto improvement we could undertake a multivariate assessment of model output without pre-assigning weights to specific criteria of interest, such as tree mortality and

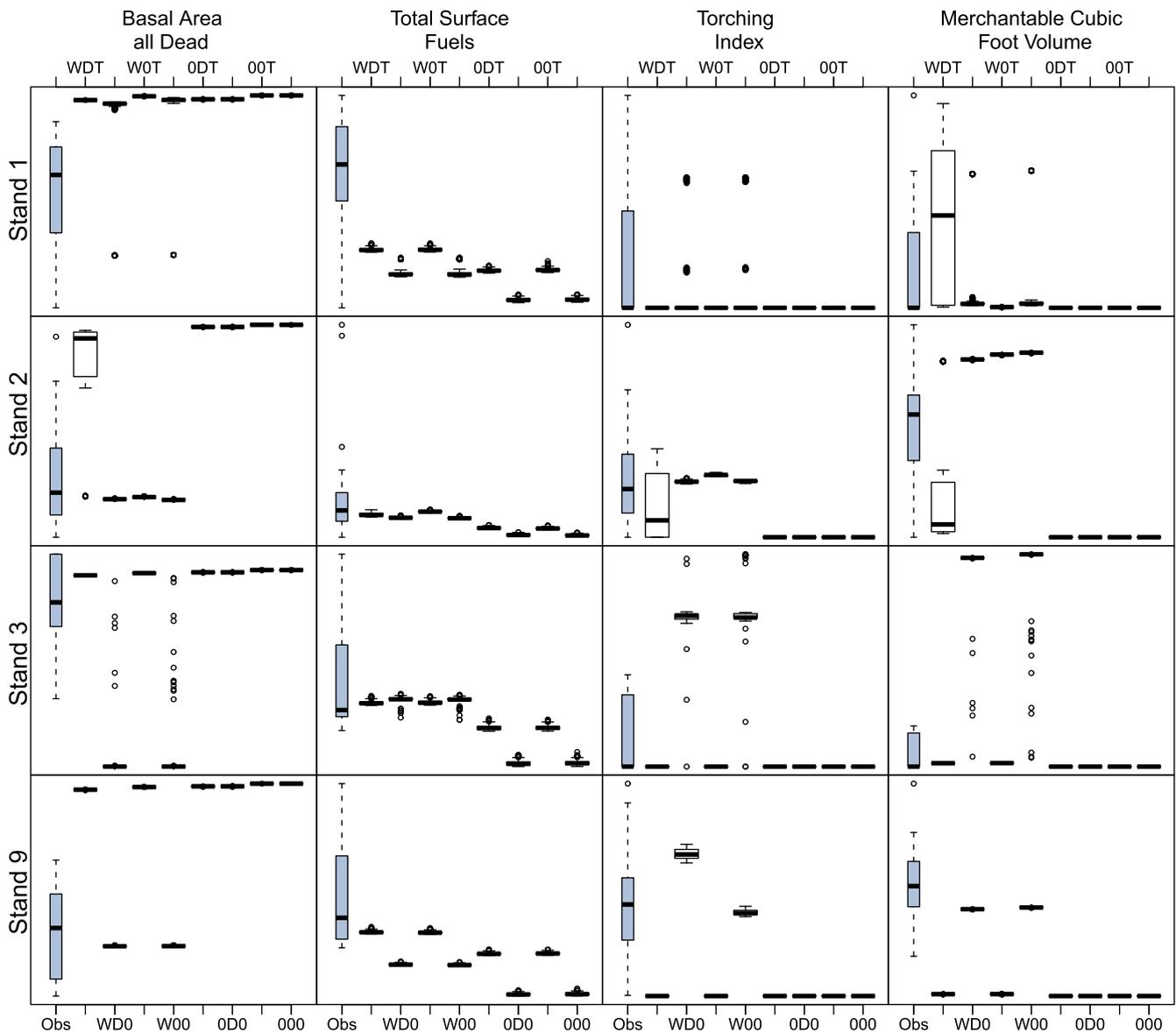


Fig. 5. Boxplots for two measured variables: the basal area of all dead trees and total surface fuels; and two derived variables: torching index and merchantable cubic foot volume and four stands (1, 2, 3, and 9). Each panel shows the relation of the plot data (far left of the panel) to the simulated values for that variable in each scenario.

fuel loads. Pareto improvement is directly related to the economic theory of Pareto efficiency, and has been applied to ecological model evaluation (Reynolds and Ford, 1999; Komuro et al., 2006; Turley and Ford, 2009; Ford and Kennedy, 2011), environmental management and decision-making (Kennedy et al., 2008; Thompson and Spies, 2010), and optimality studies (Rothley et al., 1997; Schmitz et al., 1998; Vrugt and Robinson, 2007; Kennedy et al., 2010).

We compared model output for our criteria of interest across the 28 total scenario pairs that our 8 scenarios made possible. For example, we compared the model outputs of WDF and WDO for four variables (10-h fuels, 100-h fuels, litter, and total surface fuels). If WDF performed better than WDO for all of the outputs, given our selection criteria, then it represented an improvement in accuracy. If, however, WDF performed better for two of the outputs and WDO performed better for the other two, then it was not possible to judge which scenario performed better for all of the criteria. We used this process of comparison to identify the set of scenarios for which no further improvement was possible and termed these “rank 1” scenarios. Hence, the best performing scenarios are rank 1, followed by 2, 3, etc. We continued in the same manner for all our output variables of interest and for every scenario pair until all of the scenarios were ranked.

### 3. Results

#### 3.1. Overall results

The early stages of our analysis provided insight into the data and a general overview of the sensitivity of model output to changes in input values. For example, the boxplots for total surface fuels in stand 2 and stand 9 reveal sensitivity between the scenarios using actual weather data (WDF, WDO, WOF, W00) and those using the defaults for weather (ODF, OD0, OOF, O00) (Fig. 5). Results differed by stand. Stand 2, for example, showed little sensitivity to disease or fuel inputs across the four scenarios using actual weather data. In contrast, stand 9 displayed sensitivity to the inclusion of the more detailed fuel data.

We also observed that variability in plot-scale measurements was at times greater than variability in the simulated outputs

and that the derived values appeared less variable than measured ones (Fig. 5).

#### 3.2. Evaluating model performance for a suite of measured and derived variables

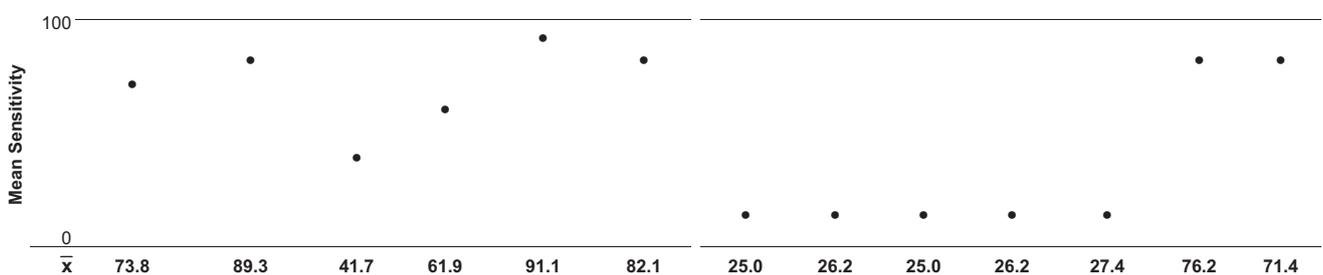
All tested model outputs displayed some sensitivity to alternative model inputs across scenarios. The degree of sensitivity depended on the variable. The proportion of scenario pairs whose inter-quartile ranges did not overlap was calculated for each stand; mean sensitivity was calculated as the mean of those six proportions (Table 3). A higher sensitivity value indicates a variable whose value was more sensitive when changing from one scenario to another. The measured variables, as shown in Table 3 had mean sensitivity values ranging from 41.7 to 91.1. The measured variable dead Douglas-fir trees (20–30" dbh or DTDF20) was less sensitive than the other measured variables. In contrast, derived variables had mean sensitivities that ranged from 25.0 to 76.2 (Table 3). These derived variables tended to be less sensitive than the measured variables; four of the six derived variables showed no sensitivity at all (0.0) in some stands (Table 3). Exceptions were merchantable cubic foot volume (MCuFt) and stand density index (SDI), both of which exhibited sensitivity commensurate with the suite of measured variables (>70).

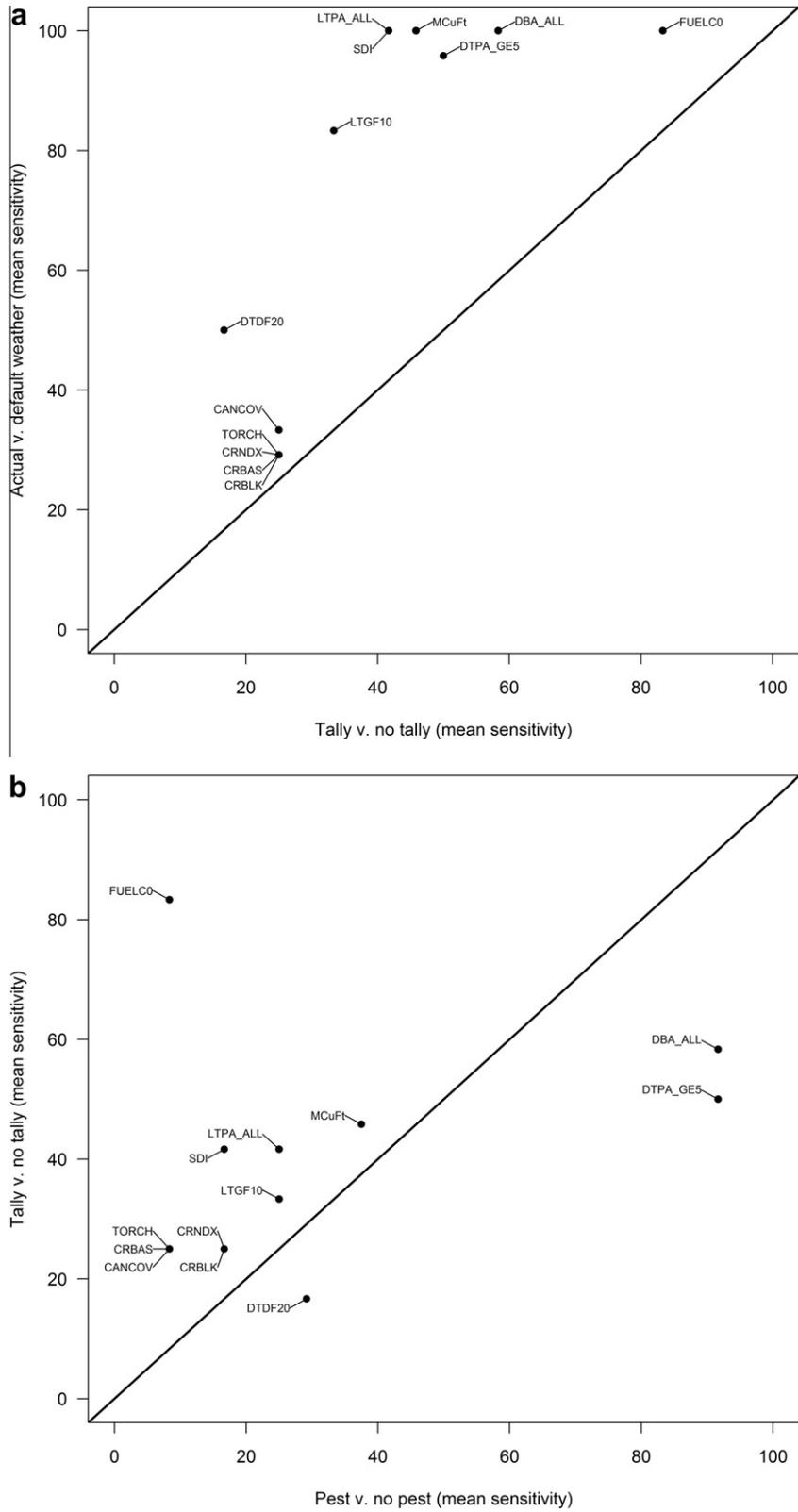
Our results suggest that the suite of model outputs we examined was most sensitive to whether actual or default weather was used and least sensitive to whether or not disease data were included as model inputs. In Fig. 6 we illustrate a comparison between the mean sensitivity of model outputs to fuels tally versus no fuels tally and the mean sensitivity of outputs to disease presence versus absence. There are thirteen points on the plot, one for the mean sensitivity for each variable (six measured plus seven derived). The line overlain on the plot is of a 1–1 relationship. If the sensitivity to fuels tally were equal to disease data, then the mean sensitivity would fall on the line. If the sensitivity to fuels tally were greater than the sensitivity to disease data, the mean sensitivity would fall above the line. We found that the sensitivity to whether fuel tallies were included tended to be greater than the sensitivities to whether disease data were included (Fig. 6a). We also found that sensitivities to using actual versus default weather

**Table 3**

Mean sensitivity of each output variable by stand (measured variables are shown to the left of the broken line and derived variables are shown to the right of it). (LTPA\_all = all live trees per acres; DTPA\_GE5 = dead trees per acre greater than or equal to 5"; DTDF\_20 = dead Douglas-fir trees 20–30" dbh; LTGF10 = live grand fir trees 5–10" dbh; DBA\_all = basal area of all dead trees; FUELCO = total surface fuels; Torch = torching index; crndx = crowning index; CrBAs = crown baseheight; CRBlk = crown bulk density; cancov = canopy cover; MCuFt = merchantable cubic foot volume; SDI = stand density index). Higher mean sensitivity values suggest the variable was more sensitive when changing from one scenario to another. The dots under each column illustrate the mean value of that column relative to the others. It is evident that those to the right of the broken line are comparatively insensitive.

StandID	Measured						Derived						
	LTPA_ALL	DTPA_GE5	DTDF20	LTGF10	DBA_ALL	FUELCO	TORCH	CRNDX	CRBAS	CRBLK	CANCOV	MCuFt	SDI
1	75.0	78.6	0.0	75.0	85.7	85.7	0.0	0.0	0.0	0.0	0.0	75.0	75.0
2	78.6	92.9	0.0	78.6	96.4	89.3	60.7	64.3	64.3	64.3	78.6	78.6	75.0
3	71.4	92.9	78.6	0.0	89.3	89.3	42.9	46.4	42.9	46.4	42.9	75.0	75.0
7	71.4	92.9	0.0	75.0	92.9	85.7	0.0	0.0	0.0	0.0	0.0	75.0	71.4
8	71.4	85.7	85.7	71.4	89.3	57.1	0.0	0.0	0.0	0.0	0.0	75.0	57.1
9	75.0	92.9	85.7	71.4	92.9	85.7	46.4	46.4	42.9	46.4	42.9	78.6	75.0
$\bar{x}$	73.8	89.3	41.7	61.9	91.1	82.1	25.0	26.2	25.0	26.2	27.4	76.2	71.4





**Fig. 6.** A comparison of the mean sensitivity of model outputs to fuel tally versus no tally and disease versus no disease (“pest”). There are thirteen points on the plot, one for the mean sensitivity for each variable (six measured plus seven derived). The line overlain on the plot is of a 1–1 relationship. If the sensitivity to fuels tally were equal to disease data, then the mean sensitivity would fall on the line. If the sensitivity to fuels tally were greater than the sensitivity to disease data, the mean sensitivity would fall above the line. See Table 3 for abbreviations.

were greater than the sensitivities to whether fuel tallies were included (Fig 6b).

### 3.3. Evaluating model performance for tree mortality output

#### 3.3.1. Overall mortality

When accuracy of modeled tree mortality was considered, the third scenario (WOF: actual weather /fuel tally) was the only scenario that achieved a Pareto rank 1. It was the only scenario evaluated able to achieve the best level of performance for tree mortality expressed as both basal area and by TPA (Table 2). When mortality was expressed as basal area, the best performers among the scenarios satisfied two out of the six stands (WDO, WOF and W00 all achieved two stands). When mortality was expressed as TPA all scenarios except WDO and W00 were able to achieve one stand. Scenarios 2 (WDO) and 4 (W00) were able to satisfy two of the stands for basal area mortality, but satisfied none of the stands when mortality was expressed as TPA. All of the default scenarios were able to satisfy one stand for mortality expressed as TPA; however, none of them were able to satisfy the basal area mortality criterion for two stands.

#### 3.3.2. Basal area mortality by tree species and diameter class

We evaluated model performance in predicting mortality across 35 species \* diameter size class combinations. The model performed consistently across the scenarios for the simulated values of mortality by species and size class (Table 4). When the scenarios were ranked across all species and size classes, six of the eight scenarios were assigned rank 1; only 00F and 000 scenarios assigned rank 2. For the criteria in which there were differences among the scenarios, the scenarios with actual weather data generally performed better than the scenarios with default weather data (Table 4). To further summarize the results we calculated for each species the proportion of diameter classes for which the model achieved a fair or good classification (>3 stands satisfied), and for each diameter class the proportion of species for which the model achieved a fair or good classification (>3 stands satisfied).

We evaluated model performance for predicting tree mortality across species (Table 4). The dominant tree species in the six stands we evaluated in our analysis was grand fir (Fig. 2), and for grand fir, across all of the size classes, WOF was the only scenario assigned a Pareto rank 1, implying that this scenario (actual weather, no disease, fuel tally) performed best when simulating grand fir mortality across all diameter size classes (Table 4). Overall, results for model performance in predicting grand fir by size class were

mixed, with the model able to meet the binary selection criterion for basal area mortality for as few as one stand (the 30–40" dbh diameter class) to as many as five stands (over 50" dbh diameter class) (Table 4). For grand fir in the 10–20" dbh diameter class, which represents the average diameter in the sample, half of the stands (3) met the selection criterion. Overall for grand fir, the model achieved a fair or good classification (>3 stands satisfied) for 4/6 of the diameter classes (a proportion of 0.6). The number of times the model achieved a fair or good performance for basal area mortality in the 10–20" dbh size class was 4/5 species (including grand fir, a proportion of 0.8).

We also evaluated model performance for predicting tree mortality across size classes (Table 4). In some instances where few or no dead trees existed in a particular diameter class (e.g. ponderosa pine 5–10" or subalpine fir >50"), all of the stands met the selection criteria. For Douglas-fir, the model did not replicate the observed number of dead trees for all of the stands in any of the diameter classes. For a given stand, this means that no scenario estimated the basal area mortality of Douglas-fir in each individual diameter class to a level that met our selection criteria. Furthermore, no scenario estimated the basal area mortality of Douglas-fir in all diameter classes to a level that met our criteria, despite several scenarios meeting the criteria for individual diameter classes (5–10, and all larger than 30"). The model performed best for the smallest and largest diameter size classes and for lodgepole pine.

With respect to Pareto ranks, all scenarios were assigned rank 1 for subalpine fir and for the all diameter, 5–10" and 30–40" size classes, implying no differences in performance among the scenarios. When there were differences in performance among the scenarios, 00F and 000 were never assigned rank 1, and 0DF and 0D0 are assigned rank 1 only for lodgepole pine. The remaining rank 1 assignment was given to various combinations of the scenarios that used actual weather data.

### 3.4. Evaluating model performance for fuel outputs

The performance of the model for estimating total surface fuels was better for the scenarios using actual weather data than for the scenarios using default weather data (WDF, WOF and W00 were assigned Pareto rank 1), although the model only achieved a fair or good classification for two of the four fuels classes (100-h and total dead surface fuels; Table 2). For fine fuels and litter it was rare that any stand met the selection criterion. The best scenarios were scenarios with actual weather and disease data (regardless of whether the fuel tally was included; WDF, WDO), as well as the actual weath-

**Table 4**

Mortality table: species x diameter class. For cases when there was a different performance among scenarios, those that performed better than the rest are identified in the table. Otherwise, all scenarios performed identically. For the criteria in which there were differences among the scenarios, the scenarios using actual weather data generally performed better than the scenarios using default weather data. To further summarize the results we calculated for each species the proportion of diameter classes for which the model achieved a fair or good classification (>3 stands satisfied), and for each diameter class the proportion of species for which the model achieved a fair or good classification (>3 stands satisfied).

Diameter class (")	Ponderosa pine	Lodgepole pine	Douglas-fir	Subalpine fir	Grand fir	Proportion	Best performing scenarios (rank 1)
All	2 (WDO, W00)	<b>3</b>	0	2	2 (WDF, WOF, 0DF, 0D0, 00F, 000)	0.2	All scenarios
5–10	<b>6</b>	<b>4</b>	<b>5</b>	<b>3</b>	<b>4</b>	1.0	All scenarios
10–20	<b>3</b>	<b>3</b>	<b>3</b>	2	<b>3</b> (WDO, WOF, W00)	0.8	WDO, WOF, W00
20–30	1	2 (WDF, WDO, 0DF, 0D0)	1	<b>3</b>	2 (WDO, WOF, W00)	0.2	WDO
30–40	2	<b>6</b>	<b>4</b>	<b>6</b>	1	0.6	All scenarios
40–50	2 (WDO, W00)	<b>6</b>	<b>3</b>	<b>6</b>	<b>4</b>	0.8	WDO, W00
50+	<b>5</b> (W00)	<b>6</b>	<b>4</b> (WDO, WOF, W00)	<b>6</b>	<b>5</b>	1.0	W00
<b>Proportion Rank 1 scenario(s)</b>	0.4 W00	0.9 WDF, WDO, 0DF, 0D0	0.7 WDO, WOF, W00	0.7 All scenarios	0.6 WOF		

er scenario without disease data and without the fuel tally (W00). This is not unexpected, as 1000-h fuel data from the tally method do not drive fire behavior in FFE-FVS. In our analysis, the model performed fairly for total surface fuels and poorly for fine fuels and litter (Table 2). The model performed best in predicting fuels when actual weather data were included, but whether including the fuels tallies improved performance depended on the fuel class. Including the fuel tallies improved performance for total surface fuels and 100-h fuels, but not for litter and 10-h fuels (Table 2). There was no difference in model performance between including the disease data and not including it (disease and no tally performed the same as no disease and no tally). Mortality from DM was estimated by FVS to be < 1% over the 7 year simulation period prior to the fire. The contribution of DM brooms to the actual fire cannot be determined and is not considered in FFE-FVS simulations.

#### 4. Discussion

Our combination of sensitivity and accuracy assessments provided insight into performance of the east Cascades variant of the FVS-FFE model at the site of the Cold Springs wildfire. While a sensitivity analysis shows whether model outputs are sensitive to changes in an input variable, model accuracy describes the relation of model output to observed, post-fire sample data. The results for tree mortality indicate that none of the scenarios we designed satisfied the selection criteria for a majority of stands. Indeed, for most of the stands the median simulated value for mortality fell outside of the interquartile range of plot data. Depending on the intended use of output, a model user may be more concerned about ways to improve the accuracy of predictions than about the sensitivity of output to changes in input values. If accuracy is a key consideration, our results suggest that using site-specific information over model default values could significantly improve the accuracy of simulated values. If understanding model sensitivities is most important, then this analysis offers a framework upon which to build. In particular, useful future steps for analysis would be to explore variation in model sensitivity across model variants, and for other fuelbed and weather conditions.

The peer-reviewed literature on performance assessments for FVS and its suite of extensions is sparse. Our study, while limited by the use of one variant and a small data set for a limited domain, therefore offers a contribution. Namely, that model users need to be mindful of the sufficiency of default values for their individual applications and to consider the potential effects of model sensitivities and insensitivities on the output values they use or report. In this section, we consider these points in turn.

##### 4.1. Model sensitivity and accuracy

The East Cascades variant of the FFE-FVS model was sensitive to input values for weather and fuels, but not for disease (Fig. 6). The model performed better when actual data on weather and fuels were used as input values than when default values were used. The sensitivity of model outputs to model inputs differed by model outputs and by stand (Table 3), implying that model performance may vary by stand.

The East Cascades variant of the FFE-FVS model was more accurate when estimates for tree mortality were expressed on a basal area basis versus TPA. None of the scenarios, however, satisfied our binary accuracy criterion for tree mortality or fuels for a majority of our six stands (Table 2). The accuracy of predicted basal mortality of grand fir, the predominant species in our stands, was slightly better (Table 4) than for the other species.

By exploring model sensitivity and accuracy we can glean insight into both forest ecology and model performance. For

example, our analysis revealed that the estimate of total surface fuels was more sensitive to whether actual weather conditions were entered into the model or whether the default conditions were accepted than was torching index. Two main explanations exist. Either the total surface fuels in a stand after a wildfire truly were sensitive to the weather conditions during the fire or the sensitivity of the model output to weather reflects the mathematical or process structure of FVS. Torching index in the FFE-FVS model is a calculation designed to represent the wind speed at which a surface fire is expected to ignite the crown layer, which depends on the fuels, fuel moisture, canopy base height, and topography (slope steepness). Johnson et al. (2011) found that the torching index estimated by FFE-FVS was insensitive to the application or absence of a surface fuel treatment after thinning. The insensitivity of modeled output to changes in input data can reflect coarse thresholds or categorizations within the internal mathematics of the model. When FFE-FVS aggregates measured fuels information into a discrete fuel model, the model becomes insensitive to changes in on-the-ground fuels because relatively broad fuel conditions could occupy a single fuel model. Such insensitivity does not necessarily reflect a real world lack of response but, rather, gaps in our knowledge of how fine-scale differences in inputs might best be propagated through the model.

The combined findings may reveal limitations in the use of discrete fuel model classes as input to FFE-FVS. Stand conditions that might produce variable fire behavior and effects get lumped into one fuel class which, in turn, produces identical predictions. Similar explanations may explain why derived variables that depend on fuel estimates are less sensitive to changing scenarios than the measured variables in our analyses. If so, our results suggest how changes in the structure of key input variables might be able to improve model performance. Like model sensitivity, model insensitivity may reflect a true lack of relationship in the real world or it may result from a missing or poorly-specified mathematical relationship in the model.

##### 4.2. Model domain

Models are useful for developing hypotheses and/or making decisions. Assessing their performance can improve their utility. The design of a model performance assessment, including which criteria are chosen, directly defines the scope of inference for an evaluation. The full potential domain of a model such as FVS is enormous, especially when the extensions are taken into account. The adequacy of a large model can be defined over a smaller domain constrained by a discrete set of data and questions. This understanding can aid in building the set of domains over which a model is adequate, the kinds of questions that it can answer, and the context in which a question can be answered.

The approach we adopted was to evaluate FVS at a local, stand-scale. The full domain of model parameters is difficult to evaluate, so we chose a relatively small set over which to evaluate model performance across model scenarios. The domain of our analysis was the east Cascades variant of the model for tree mortality and fuels. In so doing, we have begun to piece together the applicable domain of FFE-FVS, which requires both definition of when the model performs well and when the model performs poorly. A model is expected to fail outside of its applicable domain (Rykiel, 1996), yet if that space is not delineated then it is likely the model will be used in contexts for which it can be expected to fail. Although we cannot recommend that these results be extrapolated to all applications of FFE-FVS, our analyses did reveal conditions under which the model worked best and conditions for which the model performed poorly. We expect other studies will provide additional information until a clearer picture exists of model performance across the domain of FFE-FVS.

#### 4.3. Plot-scale data versus stand-scale model

Scale is a consideration in applying any model. FVS and FFE-FVS are stand level forest growth and fire effects models that have been used to simulate stand-scale dynamics for decades. These models are generally run at the stand scale by pooling input data from plots within the stand. Variability across plots can be incorporated into analyses in different ways (Hummel and Cunningham, 2006), but is not commonly done. Where plots are quite variable, we might not expect the stand average to be similar to what is observed on any one plot. In our analysis, we used the range of observed plot-scale data to bound the range for which we considered the model to provide accurate predictions. We observed for several of our metrics that the variability of the metric based on observed data was greater than the range of simulated metrics using the stochastic component of FFE-FVS (RANNSEED) (Fig. 5). This discrepancy might simply suggest increasing variability in the RANNSEED function to better reflect on-the-ground conditions. The increased variability of the observed data might also reflect patterns of fire behavior that are not yet captured in FFE-FVS. The Cold Springs fire burned stands unevenly, which increased the variance across plots within a stand. When the FFE-FVS model is run at the stand level, as is common in management applications, it does not capture such within-stand variation. A performance analysis at a plot level would produce different results, but not reflect the way FVS and FFE-FVS are commonly used.

Variability in plot-scale data will also be affected by the size of the plot. Within our analyses, all plots were the same size. However, if researchers were to use significantly smaller plots, they should expect even greater variability in observed data. Larger plots should reduce variability in observed data across plots. How field data are collected, including plot size and number, will therefore impact how well model outputs might be expected to compare to observed data.

#### 4.4. Management implications

Whether model performance can be considered adequate depends on several factors. A model performance assessment needs to be consistent with model use. The decision about what output to evaluate or what criteria to use should be linked to expectations and needs about model performance. We selected metrics for a national, federally supported, individual tree forest growth and fire effects model that were related to a previous analysis for the study area that used the same model. Results from that previous analysis suggested that most of the trees removed by silvicultural treatments designed to support fire and habitat objectives in the forest reserve, while generating enough revenue to break-even, would be medium-sized shade tolerant conifers like grand fir (Hummel and Barbour, 2007). Results from our current analysis indicate that model performance was at least fair for these species, in these size classes. This is important, because management decisions were made that authorized density reduction treatments (USDA, 2003). The treatments focused on removing trees less than 25 cm. For trees of this size class, our results indicate that the model performed well, and that model performance was relatively insensitive to the use of particular model inputs (Table 4).

Our results suggest that using site-specific information over model default values could significantly improve the accuracy of simulated values. In particular, recording and inputting actual weather data is likely to improve model performance. Since the model was less sensitive to the inclusion of detailed fuels tallies, it may be more useful to collect quick fuels data (i.e. recording which photo best corresponds to visual inspection) over several plots than to collect labor-intensive fuels data at fewer plots. Of course there

may be other compelling reasons to collect detailed fuels data. The small size of our sample means it would be incautious to extrapolate our results to other forest conditions, field situations, and model variants without additional performance assessments.

#### References

- Agee, J.K., 1993. Fire Ecology of Pacific Northwest Forests. Island Press, Washington DC.
- Ager, A.A., McMahan, A., Hayes, J.L., Smith, E.L., 2007. Modeling the effects of thinning on bark beetle impacts and wildfire potential in the Blue Mountains of eastern Oregon. *Landscape and Urban Planning* 80, 301–311.
- Anderson, H.E., 1982. Aids to Determining Fuel Models for Estimating Fire Behavior. USDA Forest Service INT General Technical Report INT-GTR-122. Ogden UT. <<http://www.treesearch.fs.fed.us/pubs/6447>>.
- Barreiro, S., Tomé, M., 2011. SIMPLOT: simulating the impacts of fire severity on sustainability of eucalyptus forests in Portugal. *Ecological Indicators* 11, 36–45.
- Box, G.E.P., Draper, N., 1987. Empirical Model Building and Response Surfaces. Wiley, John & Sons.
- Bull, E.L., Parks, C.G., Torgersen, T.R., 1997. Trees and Logs Important to Wildlife in the Interior Columbia River Basin. USDA Forest Service Pacific Northwest Research Station General Technical Report PNW-GTR-391, p. 55.
- Calkin, D., Hummel, S., Agee, J.K., 2005. Modeling the tradeoffs between fire threat reduction and late-seral forest structure. *Canadian Journal of Forest Research* 35, 2562–2574.
- Cawrse, D., Keyser, C., Keyser, T., Sanchez Meador, A., Smith-Mateja, E., Van Dyck, M., 2010. Forest Vegetation Simulator Model Validation Protocols. In: United States Department of Agriculture Forest Service, Fort Collins, CO, p. 10. <[http://www.fs.fed.us/fmsc/ftp/fvs/docs/steering/FVS\\_Model\\_Validation\\_Protocols.pdf](http://www.fs.fed.us/fmsc/ftp/fvs/docs/steering/FVS_Model_Validation_Protocols.pdf)>.
- Crookston, N.L., Dixon, G.E., 2005. The forest vegetation simulator: a review of its structure, content, and applications. *Computers and Electronics in Agriculture* 49, 60–80.
- Crookston, N.L., Stage, A.R., 1991. User's Guide to the Parallel Processing Extension of the Prognosis Model. USDA Forest Service INT General Technical Report INT-GTR-281. <<http://www.fs.fed.us/fmsc/fvs/documents/userguides.shtml>>.
- Crookston, N.L., Colbert, J.J., Thomas, P.W., Sheehan, K.A., Kemp, W.P., 1990. User's Guide to the Western Spruce Budworm Modeling System. USDA Forest Service INT Research Station General Technical Report GTR-INT-274. Ogden, UT, p. 75.
- David, L., 2005. Dwarf Mistletoe Impact Modeling System: User Guide and Reference Manual Nonspatial Model. Unpublished Report Updated 2012: USDA Forest Service. <<http://www.fs.fed.us/fmsc/fvs/documents/userguides.shtml>>.
- Diaz, N.M., High, C.T., Mellen, T.K., Smith, D.E., Topik, C., 1997. Plant association and management guide for the mountain hemlock zone. Gifford Pinchot and Mt. Hood National Forests. R6-MTH-GP-TP-08-95. USDA Forest Service Pacific Northwest, Region, p. 101.
- Dixon, G.E., 2010. Essential FVS: A User's Guide to the Forest Vegetation Simulator. In: U.S. Department of Agriculture, F.S. (Ed.). Forest Management Service Center, Fort Collins, CO.
- FMSC, 2011a. Service Center Website. <<http://www.fs.fed.us/fmsc/fvs/index.shtml>> (accessed 11.11).
- FMSC, 2011b. Annual Report. <<http://www.fs.fed.us/fmsc/ftp/fmsc/annualreport/FMSCAnnualStaffReport2011.pdf>> (accessed 11.11).
- Ford, E.D., Kennedy, M.C., 2011. Assessment of uncertainty in functional-structural plant models. *Annals of Botany* 108 (6), 1043–1053.
- Frankel, S.J., 1998 (tech co-ord). User's Guide to The Western Root Disease Model, v.3.0. USDA Forest Service Pacific Southwest Research Station General Technical Report PSW-GTR-165. Albany, CA. <<http://www.fs.fed.us/fmsc/fvs/documents/userguides.shtml>>.
- Franklin, J.F., Dyrness, C.T., 1988. Natural vegetation of Oregon and Washington. In: U.S. Department of Agriculture, F.S., Pacific Northwest Research Station (Ed.), Portland, Oregon.
- Gregg, T.F., Hummel, S., 2002. Assessing sampling uncertainty in FVS projections using a bootstrap resampling method. In: Crookston, N.L., Havis, R.N. (Eds.), Second Forest Vegetation Simulator Conference, February 12–14, Fort Collins, CO. Proc RMRS-P-25 Ogden UT: USDA Forest Service Rocky Mountain Research Station Proceedings, pp. 164–167.
- Hamilton, Jr., D.A., 1991. Implications of random variation in the stand prognosis model. In: Research Note – US Department of Agriculture, Forest Service, p. 11.
- Hornberger, G.M., Cosby, B.J., 1985. Selection of parameter values in environmental models using sparse data: a case study. *Applied Mathematics and Computation* 17, 335–355.
- Hood, S.M., Mchugh, C.W., Ryan, K.C., Reinhardt, E., Smith, S.L., 2007. Evaluation of a post-fire tree mortality model for western USA conifers. *International Journal of Wildland Fire* 16, 679–689.
- Hoover, C.M., Rebain, S.A., 2010. Forest carbon estimation using the forest vegetation simulator: seven things you need to know. In: U.S. Department of Agriculture, F.S. (Ed.).

- Hummel, S., Agee, J.K., 2003. Western spruce budworm defoliation effects on forest structure and potential fire behavior. *Northwest Science: Official Publication of the Northwest Scientific Association* 77, 159–169.
- Hummel, S., Barbour, R.J., 2005. Landscape silviculture for late-successional reserve management. In: Powers, R.F. (Ed.), USDA Forest Service, Pacific Southwest Research Station General Technical Report PSW-GTR-203. Restoring Fire-Adapted Ecosystems: Proceedings of the 2005 National Silviculture Workshop. Tahoe City, CA, pp. 157–169.
- Hummel, S., Calkin, D.E., 2005. Costs of landscape silviculture for fire and habitat management. *Forest Ecology and Management* 207, 385–404.
- Hummel, S., Cunningham, P., 2006. Estimating variation in a landscape simulation of forest structure. *Forest Ecology and Management* 228, 135–144.
- James, P.M.A., Fortin, M.-J., Sturtevant, B.R., Fall, A., Kneeshaw, D., 2011. Modelling spatial interactions among fire, spruce budworm, and logging in the Boreal Forest. *Ecosystems* 14, 60–75.
- Johnson, M.C., Kennedy, M.C., Peterson, D.L., 2011. Simulating fuel treatment effects in dry forests of the western United States: testing the principles of a fire-safe forest. *Canadian Journal of Forest Research* 41, 1018–1030.
- Kennedy, M.C., Ford, E.D., 2011. Using multicriteria analysis of simulation models to understand complex biological systems. *BioScience* 61 (12), 994–1004.
- Kennedy, M.C., Ford, E.D., Hinckley, T.M., 2010. Defining how aging *Pseudotsuga* and *Abies* compensate for multiple stresses through multi-criteria assessment of a functional-structural model. *Tree Physiology* 30 (1), 3–22.
- Komuro, R., Ford, E.D., Reynolds, J.H., 2006. The use of multi-criteria assessment in developing a process model. *Ecological Modelling* 197, 320–330.
- Lichatowich, J., Mobernd, L., Lestelle, L., Vogel, T., 1995. An approach to the diagnosis and treatment of depleted Pacific salmon populations in Pacific Northwest watersheds. *Fisheries* 20, 10–18.
- Maxwell, W.G., Ward, F.R., 1980. Photo Series for Quantifying Natural Forest Residues in Common Vegetation Types of the Pacific Northwest. USDA Forest Service PNW General Technical Report PNW-GTR-105.
- McElhany, P., Steel, E.A., Jensen, D., Avery, K., Yoder, N., Busack, C., Thompson, B., 2010. Dealing with uncertainty in ecosystem models: lessons from a complex salmon model. *Ecological Applications* 20, 465–482.
- McGarvey, D.J., Johnston, J.M., Barber, M.C., 2010. Predicting fish densities in lotic systems: a simple modeling approach. *Journal of the North American Benthological Society* 29 (4), 1212–1227.
- Pielke, R.A., Conant, R.T., 2003. Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology* 84, 1351–1358.
- Rebain, S., 2010 (comp). The Fire and Fuels Extension to the Forest Vegetation Simulator: Updated Model Documentation. Internal Report: USDA Forest Service. <<http://www.fs.fed.us/fmcs/fvs/documents/userguides.shtml>>.
- Regan, H.M., Coyvan, M., Burgman, M.A., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12, 618–628.
- Reynolds, J.H., Ford, E.D., 1999. Multi-criteria assessment of ecological process models. *Ecology* 80, 538–553.
- Rothley, K.D., Schmitz, O.J., Cohon, J.L., 1997. Foraging to balance conflicting demands: novel insights from grasshoppers under predation risk. *Behavioral Ecology* 8 (5), 551–559.
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling* 90, 229–244.
- Schmitz, O.J., Cohon, J.L., Rothley, K.D., Beckerman, A.P., 1998. Reconciling variability and optimal behaviour using multiple criteria in optimization models. *Evolutionary Ecology* 12, 73–94.
- Stage, A.R., 1973. Prognosis Model for Stand Development. Intermountain Forest & Range Experiment Station, Forest Service, U.S. Dept. of Agriculture, Ogden, Utah.
- Steel, E.A., McElhany, P., Yoder, N.J., Purser, M.D., Malone, K., Thompsen, B.E., Avery, K.A., Jensen, D., Blair, G., Busack, C., Bowen, M.D., Hubble, J., Kantz, T., Mobernd, L., 2009. Making the best use of modeled data: multiple approaches to sensitivity analysis. *Fisheries* 34 (7), 330–339.
- Thompson, J., Spies, T., 2010. Factors associated with crown damage following recurring mixed-severity wildfires and post-fire management in southwestern Oregon. *Landscape Ecology* 25, 775–789.
- Topik, C., 1989. Plant association and management guide for the grand fir zone, Gifford Pinchot National Forest. R6-Ecol-TP-006-88. Portland OR. USDA Forest Service Pacific Northwest, Region, p. 99.
- Turley, M.C., Ford, E.D., 2009. Definition and calculation of uncertainty in ecological process models. *Ecological Modelling* 220, 1968–1983.
- USDA, 2003. Final Environmental Impact Statement (FEIS): Gotchen Risk Reduction and Restoration Project. Mt. Adams Ranger District, Skamania and Yakima Counties, WA. USDA Forest Service, p. 329.
- USDA, 2010. Final Environmental Impact Statement (FEIS): EXF Thinning, Fuels Reduction, and Research Project, Pringle Falls Experimental Forest, Deschutes County OR. Appendix B. USDA Forest Service, PNW Research Station, p. 282.
- Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research* 43, W01411.
- Waddell, K.L., 2002. Sampling coarse woody debris for multiple attributes in extensive resource inventories. *Ecological Indicators* 1, 139–153.