QUANTIFICATION AND INCORPORATION OF UNCERTAINTY IN FOREST GROWTH AND YIELD PROJECTIONS USING A BAYESIAN PROBABILISTIC FRAMEWORK

(A DEMONSTRATION FOR PLANTATION COASTAL DOUGLAS-FIR IN THE PACIFIC NORTHWEST, USA)

DUNCAN WILSON[1], VICENTE MONLEON[2], AARON WEISKITTEL[3]

[1] *Oklahoma State University, Department of Natural Resource Ecology & Management, Stillwater, OK, USA, 74078*
[2] *USDA Forest Service, PNRS, FIA Program, Corvallis, OR, USA, 97331*
[3] *University of Maine, Center for Research on Sustainable Forests, Orono, ME, USA, 04469*

ABSTRACT. A Bayesian probabilistic modeling platform was used and evaluated for application in a relatively complex individual-tree growth and yield model for coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco), which was expressed as a mixed discrete and continuous Bayesian Network for annual projections. The modeling platform used a common and open-source Bayesian analysis program (JAGS v3.3.0), and was sufficiently flexible to handle a relatively complex model structure; namely, a differential form, highly dynamic, recursive, hierarchical, non-linear system of equations with rather complex error structures. This novel probabilistic modeling platform met certain desirable criteria, including: (1) accurate and tractable projections that included full error propagation; (2) flexible and comprehensive analytic capabilities; (3) full consideration of hierarchical and multi-level model structures; (4) capacity for random effects calibration; (5) allowance of hypothesis testing and updating knowledge across different system components, simultaneously with varying sources of information (i.e., new data); (6) computational efficiency; and (7) relatively simple implementation as demonstrated in a compiled scripting language. Probabilistic projections of forest growth and yield included all sources of errors and uncertainty (e.g., estimated parameters, state variables, random effects, and residual errors). Cumulative error projections over a 40-year period for three sample Douglas-fir stands were determined. Projection errors for key metrics summed across all trees, such as total basal area and stem density, had coefficient of variations between 4-6% and 7-8%, respectively. Probabilistic projections were markedly different from deterministic projections made with the same model structure. Overall, this novel probabilistic platform showed strong promise as a general platform for ecological modeling, particularly when tractable and analytically correct error projections are required. In particular, the Bayesian probabilistic modeling approach used provided a natural platform for cross-disciplinary research, particularly between social and ecological research domains.

Keywords: forest growth and yield, error propagation, model uncertainty, error budgets, individual tree growth models, coastal Douglas-fir, Oregon, Washington.

1 INTRODUCTION

Ecological modeling is dominated by deterministic approaches, particularly in forestry (Canham et al. 2003, Weiskittel et al. 2011). The structure of the modeling equations is dependent on the system being modeled as well as the objectives, and so it varies considerably between published models. Such models are ubiquitous in ecological research, with examples ranging from projecting animal and plant meta-population dynamics, to process-level vegetation growth, food chains, plant succession and migration, and disturbance risk, as well as social and economic dynamics linked to ecological processes (An 2012). Notably, we do not include agent-based models (Epstein 2006) or models of purely physical processes (e.g., fire dynamics or weather), which can be viewed as more stochastic implementations. In contrast, most ecological models take a difference equa-

tion approach with a discrete time or spatial step, although systems of differential equations are not uncommon (Arnold et al. 1998). While many ecological models incorporate stochastic elements (e.g., weather, mortality events), these discrete elements are usually embedded within an otherwise deterministic framework.

Forest growth and yield (GY) models have a long history of development and are prime examples within a field dominated by difference-form, deterministic approaches (Pienaar and Shiver 1986, Weiskittel et al. 2011). The current challenges in forest GY modeling serve as an illustration for the broader field of ecological modeling. Forest GY models are ubiquitous in forestry, with particular use in economic forecasting, designing silvicultural systems, risk management, harvest scheduling, forecasting fiber supply, carbon dynamics, and ecosystem management (LeMay and Marshall 2001). The structure of forest GY models vary, but the majority were designed as a system of equations that sequentially predict structural attributes from inventory data and variables previously projected by the model. Forest GY models have variable time steps, but an annual interval is often preferred for a variety of reasons (Weiskittel et al. 2011). Similar to temporal resolution, forest GY models can also have different spatial resolutions ranging from individual trees to stand attributes with increasing focus on higher spatial resolutions (Weiskittel et al. 2011). Consequently, individual tree-level models with an annual temporal resolution have become the primary focus for both research and practical implementation (e.g., Kershaw et al. 2017; Weiskittel et al. 2016a, 2017).

Within a forest GY model, a variety of component equations that sequentially predict tree- or stand-level attributes (e.g., tree diameter or live stem density) are used. Most often, these equations are fitted independently and then sequentially linked *post hoc.* The component equations are also often strongly non-linear, and may have correlated residuals across component equations. While simultaneously modeling systems of equations to account for correlated errors has been used in this past (e.g., Hall and Clutter 2004), this method is usually used only for a subset of the equations. As a consequence, estimates of prediction error are generally lacking due to the complexity of the system of equations. In fact, most forest GY models provide no estimate of projection uncertainty and relatively few studies have attempted to address this for forest management–orientated models (e.g., Gertner et al. 1996; Green et al. 1999; MacFarlane et al. 2000; McGarrigle et al. 2013; Kershaw et al. 2017).

Furthermore, ecosystem management has progressed to the point where traditional forest GY projections of total volume or diameter distributions are no longer suf-

ficient (McComb et al. 1993). Many ecosystem dynamics and structural elements are highly stochastic both temporally and spatially. Consequently, they are often not very well represented within a deterministic modeling framework. An example is modeling the understory vegetation composition and dynamics that are an important habitat component for many small mammals and songbirds (Wilson and Puettmann 2007). Vegetation response to thinning is strongly tied to pre-treatment conditions, which are highly variable and largely unrelated to current overstory conditions (Wilson et al. 2009). As such, modeling vegetation within existing forest GY models becomes extraneous when vegetation predictions are largely independent of tree growth. Many similar information needs for ecosystem management simply do not fit within the current GY framework and, as a result, such modeling efforts have progressed independently (e.g., Running and Gower 1991, Scheller et al. 2007) rather than collaboratively.

In this study, we highlight the utility of a probabilistic modeling platform for addressing many of the challenges common to ecological models. We re-parameterized an individual-tree GY model (DF.GOAB) for plantation-grown, coastal Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*; Weiskittel et al. 2007) to demonstrate the advantages of this approach. The model platform is a continuous-node form of a Bayesian Network (Pearl 1988) that we updated (i.e., projected) using current tree- and stand-level structural information along with Bayes' theorem. The results are presented as posterior distributions, which are computed numerically with a Markov Chain Monte Carlo (MCMC) sampling approach. Unlike deterministic models, updating the model is efficient with stochastic elements, which can exponentially increase the computational requirements. Component equations, such as dominant height growth or mortality, are ideally developed using contemporary Bayesian parameter estimation approaches (Gelman et al. 2013), but this is not a strict requirement. These component equations can take almost any form, including continuous or discrete, linear and non-linear functions or distributional states. Error structures can include Gaussian and other exponential families, but also non-standard user-defined distributions, as well as hierarchical structures. Similar approaches to quantifying uncertainty using a Bayesian approach have been previously demonstrated in other disciplines (e.g., Freni and Mannina 2010) yet remain uncommon in forest science.

Because our study focuses on the modeling platform and not on developing the GY model equations or structure, we adopted model forms from a previously developed GY model as an example system (DF.GOAB; Weiskittel et al. 2007). The probabilistic modeling plat-

form is quite general, so we do not attempt to fully present here the model details or assess prediction accuracy with independent data. Instead, our primary objective is to contrast the probabilistic approach with current practices (i.e., deterministic and quasi-stochastic modeling approaches). We also will demonstrate several instances where deterministic and probabilistic models potentially have different projections. First, we present the probabilistic modeling platform and how projections are made, and then how the revised DF.GOAB model is portrayed in a Bayesian Network.

2 Methods

2.1 Bayesian probabilistic modeling platform

The Douglas-fir GY modeling system we chose to illustrate this approach is typical for this class of models, and representative of approaches taken in many terrestrial vegetation dynamics models. The component equations are a mixture of linear and non-linear equations, with weighting and multiple hierarchical random effects, and have a discrete time-step. In deterministic approaches, the estimated equation parameters are typically fixed at the value of their point estimates, often with some level of stochasticity to predict categorical variables such as mortality. Most systems of equations, including the ones in this study, are highly recursive, with predicted variables used as predictors in subsequent time-steps. These approaches ignore the uncertainty in the estimation of both parameters and predicted variables, and the stochastic component of the models equations (Dennis et al. 1985). To address these issues, the system of equations used in the study was fitted to an existing database of Douglas-fir growth, and then portrayed in a Bayesian probabilistic modeling platform. We used a mixed categorical- and continuous-node form of a Bayesian Network (Pearl 1988, Nielsen and Jensen 2009).

There is a large literature on contemporary Bayesian parameter estimation, and since the model projections follow the same methodology, we focus our explanation on where the applications (parameter estimation versus model projections) differ. Models are expressed similarly in both Bayesian Networks and in contemporary Bayesian parameter estimation as directed acyclic graphs (DAG; Nielsen and Jensen 2009). This approach allows very complex models, including hierarchical models and systems of equations, to be expressed as a series of conditional distributions defined in a parent-child relationship (Wikle 2003). Specifically, the full, joint probability distribution for the entire model can be written as the product of conditional distributions, $P(x_1, ..., x_n) = \Pi_{i=1}^n P(x_i|pa(x_i))$, where, x_i is the full suite of variables, and $pa(x_i)$ represents the parents of

x_i. This relatively simple factorization allows complex models to be described through simpler marginal probabilities, and thus estimate individual components separately, while retaining connectivity throughout the entire model.

Models in a Bayesian framework for both parameter estimation and model projections take the same general form in that parameters are represented as random variables with an associated prior probability distribution, with the model structure expressing the form of the likelihood. The contrast between Bayesian parameter estimation and model projections in a Bayesian Network are: (1) parameters in a Bayesian Network are expressed as highly informative priors with means and standard errors for the parameters usually taken from fitted equations and (2) the Bayesian Networks do not necessarily include new data to update the priors. Projections are made for "child" variables based on parent-child hierarchy (i.e., model structure), which is expressed as a series of conditional probabilities defined by the DAG. These generally include state variables as well as parameters estimates and all sources of error. Posterior distributions for projected variables are computed numerically with a Markov Chain Monte Carlo (MCMC) sampling approach since most model forms do not have a closed-form analytical solution. This MCMC approach to a Bayesian Network is not a traditional discrete-state type but is more a type of efficient random-walk first toward and then within the stable posterior distribution (Cappé and Robert 2000). The MCMC approach is automatically contained within the stable posterior distribution when no new data are included.

Model projections without new data are essentially infilling of "missing" model response data, where projections of tree- and stand-level variables over time (as in our study) represent these missing data. If some of this data becomes available, it can be provided to the model, which allows for simultaneous updating of model parameters (e.g., random effect calibration, state or latent variable estimation, etc.). Also, different types of new data will update different model parameters (including error parameters) depending on the structure of the model (Wilson et al. 2009) because systems of equations are represented simultaneously in Bayesian Networks. Here, we describe projections as missing data to underscore how projected responses from the model are conditionally related to the model parameters and state variables.

In Figure 1, the entire GY model is represented as a Bayesian Network, which is functionally very similar to contemporary Bayesian parameter estimation methods, with the exception that the entire system of equations is represented together. Blending new data within Bayesian Networks is quite common, and further blurs

the lines between the two approaches (Clark and Gelfand 2006, Wilson et al. 2009). Within the Bayesian Network, projections are made for each tree over each projection year. This projection on an "annual time step" is a slight misnomer in a Bayesian Network since each iteration of the Markov chain updates every tree for every year in the projection. Thus, 1,000 MCMC iterations does not represent a 1,000 year projection, but rather 1,000 samples from the posterior distribution, with each sample containing a full set of projections for every tree and year combination. The MCMC iterations collected for each variable comprise the marginal posterior distributions. For simplicity, the model parameters, residuals, and prior distributions are all assumed to be uniformly normal in this analysis. This assumption might be important as some of the parameters and residuals may not be normally distributed given the complex non-linearity of the models used. In addition, the wrong prior distribution can lead to incorrect estimations of uncertainty in modeling results (Freni and Mannina 2010).

2.2 Dataset

Data from 65 University of Washington Stand Management Cooperative permanent research installations in Oregon and Washington, USA, and Vancouver Island, British Columbia, Canada were used for analysis. These installations cover a wide range of growing conditions typical of the region. The overall climate is humid oceanic, with a distinct dry summer and a cool, wet winter. The twenty-year mean annual rainfall for these locations ranged from 91 to 293 cm (18–32% occurring during the growing season), and January and July mean temperatures ranged from -2.7 to 6.7°C and 14.7 to 19.2°C, respectively. Variation in precipitation and temperature are strongly related to elevation and distance from the coast. Elevation ranged from 5 to 1160 m above sea level, slope was between 0 and 60%, and all aspects were represented. Soils varied from a moderately-deep sandy loam to a very deep clay loam with mean water holding capacity of 139 ±63 mm (45 – 303 mm).

Since its establishment in 1985, the Stand Management Cooperative (SMC) at the University of Washington (http://www.cfr.washington.edu/research.smc/) has maintained a database representing 435 installations in British Columbia, Washington, and Oregon (Maguire et al., 1991). In this study, we focus on a subset of these data, restricted to pure, plantation-grown Douglas-fir in western Oregon, Washington, and British Columbia; specifically, Douglas-fir plantations extracted from the Type I, II, and III installations. Type I installations are square 0.2 ha plots established in existing plantations that have received designed sets of silvicultural treatments since plot establishment in the late 1980s and early 1990s. Type II installations are square 0.2-ha plots that were installed between 1986 and 1991 in stands approaching commercial thinning age and received different levels of commercial thinning treatments. Type III installations were established between 1985 and 1990 as initial spacing trials with six densities ranging from 247 to 3048 trees per ha. Plot size for Type III installations varied from 0.086 ha at the highest density to 0.202 ha at the lowest density. For this analysis, only the untreated (i.e., control) plots in the Type I, II and III installations were used, resulting in a model fitting database of 167 plots across 65 installations.

At each plot, individual trees were tagged and measured for diameter at breast height (D) with a subsample of trees measured for total height (HT) and height to crown base (HCB). Common variable definitions and abbreviations are given in Table 1. Plots were generally re-measured on a 4-year interval (range 2-8 yrs; Table 2). A subsample of trees were measured for HT and HCB on each plot, but with the exception of estimating missing heights to compute dominant height (HD), only directly measured variables were used in equation fitting.

2.3 Growth and yield equations

Forest GY models have a long tradition of development and use in forestry (Weiskittel et al. 2011). They are designed to project tree- or stand-level growth over time beginning with a discrete starting condition, often taken from a field-measured inventory. These models are fundamentally similar to many terrestrial vegetation and wildlife models, different only in the model focus (tree growth, soil organic matter, wildlife populations size, etc.) and data describing starting conditions.

We modified a previously published system of equations developed to project GY of individual Douglas-fir trees on an annual time-step (DF.GOAB; Weiskittel et al. 2007). This system of equations was derived from model forms used in the ORGANON-SMC GY model for coastal Douglas-fir (Hann et al. 1993, 2003). Model variables are either measured in a field inventory at the start of the projection (t_0), projected for time step t, or derived (Table 3). The primary projected variables are HD at the stand-level, and D, H, HCB and survival at the tree-level, with stand- and tree-level derived variables as model covariates. The system of equations is deterministic and highly recursive, which is common among terrestrial vegetation models (Canham et al. 2003). As in most forest GY models, tree growth is projected based on predicted site quality and stand structural characteristics. Site quality is represented by the projected height growth trajectory of the dominant trees, which is generally insensitive to stem density. The site index (SI)
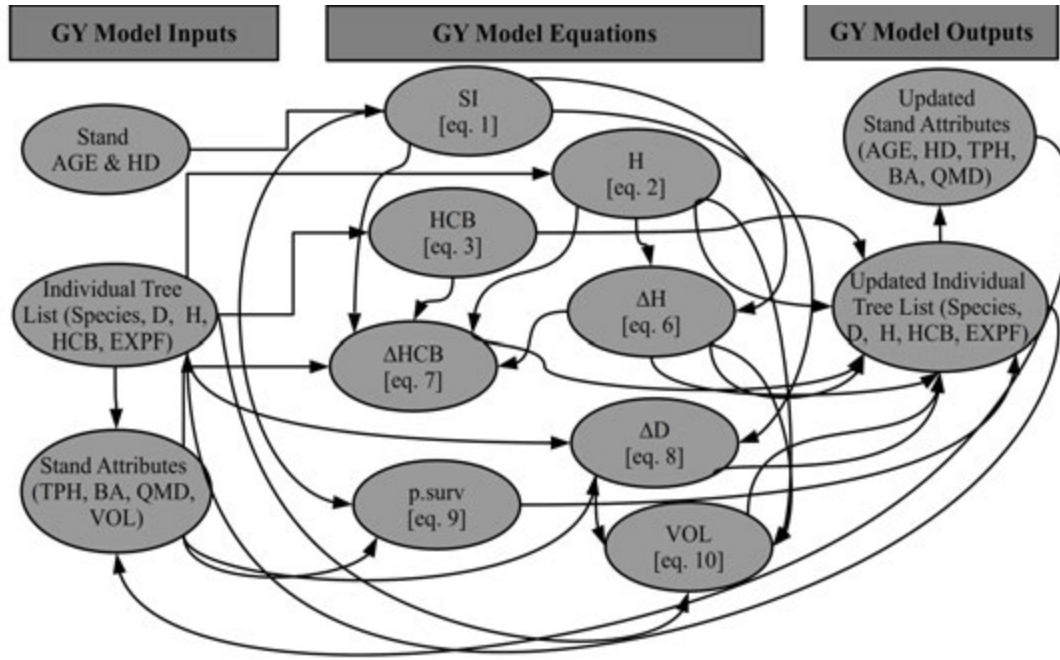
Figure 1: Structure and inter-relationships between model inputs, equations, and outputs for the Douglas-fir growth and yield (GY) model framework used in this analysis.

is then the expected height of the dominant trees at a reference age, taken as 25 years from planting for this study. Dominant height was predicted using a Hossfeld function fitted as a non-linear difference form equation,

$$SI = \frac{Age_{SI}^2}{Age_{SI}\left(\frac{Age_{10}}{HD_{10}} + \alpha_{11}\right) + \frac{\alpha_{10}}{Age_{10}} + \\ + \alpha_{11}Age_{10} + \alpha_{10}} + \epsilon_{SI} \quad (1)$$

where α_{10} are estimated parameters, Age_{si} is the site index age (25 years from planting in this study), Age_0 is the breast-height age at the initial inventory, HD_0 is the dominant height at the initiation inventory, and residual error is weighted by projection period, $\varepsilon_{SI} \sim N$. SI was estimated for each plot and then averaged for an installation.

It is assumed that all trees were measured for the initial diameter, D_0, in the initial inventory. However, H and HCB were only measured on a subsample of trees. Unmeasured individual tree heights, H_i for tree i at the start of the projection (t_0), were estimated with a power function and transformed to the log-log scale. This was done for each inventoried plot separately.

$$ln\left(H_{i0}\right) = \beta_0 + \beta_1 ln\left(D_{i0}\right) + \epsilon_{Hi0} \quad (2)$$

where β_{0-1} are estimated parameters, and residual errors, $\epsilon_H \sim N\left(0, \sigma_H^2\right)$ specific to a single plot. Similarly, the HCB was estimated using the same model form,

$$ln\left(HCB_{i0}\right) = \gamma_0 + \gamma_1 ln\left(D_{i0}\right) + \epsilon_{HCBi0} \quad (3)$$

where γ_{0-1} are estimated parameters, and $\epsilon_{HCB} \sim N\left(0, \sigma_{HCB}^2\right)$. The equation predicting maximum crown width (MCW) for open grown Douglas-fir trees was taken directly from Hann (1999) and applied as a known relationship without any incorporation of uncertainty. Similarly, covariates derived from MCW were treated as known.

Projections were made on an annual time step (t) for each tree. Dominant average height HD of the inventoried plot was projected using [Eq. 1] modified as,

$$HD = \frac{Age_t^2}{Age_t\left(\frac{Age_{t-1}}{HD_{t-1}} + \alpha_{41}\right) + \frac{\alpha_{40}}{Age_{t-1}} + \\ + \alpha_{41}Age_{t-1} + \alpha_{40}} + \epsilon_{HD} \quad (4)$$

where α_{40-41} are estimated parameters, and $\epsilon_{HD} \sim N\left(0, \sigma_{HD}^2\right)$. The Hossfeld function based base-age invariant equation (Cieszewski 2003) was selected because it showed adequate fit to the data and could be algebraically solved for Age_t. Growth effective age (GEA) was defined as the expected age of an individual tree for a given height, had that tree followed the dominant height trajectory (Eq. 1) for the plot estimated SI. For each time step, GEA_{it} was defined as the algebraic solution for Age_t in Eq. 4 based on substituting Age_{SI} for Age_{t-1}, SI for HD_t and H_{it-1} for HD_{t-1}. Potential height growth is then the height increment of an average dominant tree with height, H_{it-1} and age set to GEA_{it-1}

Table 1: Common variable units and definitions used in this analysis.

Variable	Units	Definition
AGE_t	yrs	breast height age at projection year t
AGE_{si}	yrs	Site index base age (25 yrs)
HD	m	Avg. height of the 100 dominant trees per ha.
SI	m	Site index; estimated dominant tree height at breast height age 25
MCW	m	Maximum crown width for open-grown trees for a specified diameter
MCA	m^2	Maximum crown area computed from MCW and assuming a circle
CCF	%	Crown competition factor; max potential crown area as a proportion of plot area
D	cm	Diameter at breast height
H	m	Total tree height
HCB	m	Height to lowest live branch whorl
CL	m	Crown length
CR	proportion	Crown ratio
TPH	# ha^{-1}	Stem density
BA	m^2 ha^{-1}	Basal area as a sum of D^2
DQ	cm	Quadratic mean diameter
GEA	yrs	Growth effective age
PHG	m	Potential height growth
ΔD	cm	Annual diameter increment
ΔH	m	Annual height increment
ΔHCB	m	Annual change in height to crown base
VOL_{ib}	m^3	Total inside bark tree volume

as,

$$PHG_{it} = \frac{(GEA_{it-1} + 1)^2}{(GEA_{it-1} + 1)\mathcal{N}_5 + \alpha_{40}} - H_{it-1}, \quad (5)$$

where: $\mathcal{N}_5 = \frac{GEA_{it-1}}{H_{it-1}} + \alpha_{41} GEA_{it-1} - \frac{\alpha_{40}}{GEA_{it-1}} + \alpha_2 (GEA_{it-1} + 1)$.

Individual tree height growth was then modeled as a function of potential height growth, modified to account for stand- and tree-level covariates, similar to Arney (1985) as,

$$\begin{aligned} \Delta H_{it} &= \left(\frac{H_{it-1}}{HD_{it-1}}\right)^{\psi_{60} + install_{\Delta H} + plot_{\Delta H}} \\ &\times \left(1 - \psi_{61}\left(\frac{CCF_{it-1}}{600}\right)^{\psi_{62}}\right) \\ &\times PHG_{it} + \epsilon_{\Delta H} \end{aligned} \quad (6)$$

where ψ_{60-62} are estimated parameters, $install_{\Delta H}$ and $plot_{\Delta H}$ are random effects assumed distributed Gaussian with means zero and variances $\sigma^2_{\Delta H_{install}}$ and $\sigma^2_{\Delta H_{plot}}$, respectively, and residual error is weighted by H as, $\epsilon_{\Delta H} \sim N\left(0, H_{it-1}^2 \sigma^2_{\Delta H}\right)$.

Recession in HCB was adapted from Hann and Hanus (2004) as,

$$\Delta HCB_{it} = \frac{CL_{it-1} + \Delta H_{it}}{1 + \exp(\mathcal{N}_7)} + \epsilon_{\Delta HCB} \quad (7)$$

where, $\mathcal{N}_7 = \varphi_{70} + install_{\Delta HCB} + plot_{\Delta HCB} + \varphi_{71} ln(CR_{it-1}) + \varphi_{72} CR_{it-1} + \varphi_{73} GEA_{it-1} +$

$\varphi_{74} ln(BA_{it-1}) + \varphi_{75}\frac{CR_{it-1}}{BA_{it-1}}$, and where, φ_{70-75} are estimated parameters, CL is crown length, CR is crown ratio, $install_{\Delta HCB}$ and $plot_{\Delta HCB}$ are random effects assumed distributed Gaussian with means zero and variances $\sigma^2_{\Delta HCB_{install}}$ and $\sigma^2_{\Delta HCB_{plot}}$, respectively, and residual error is weighted by CL as, $\epsilon_{\Delta HCB} \sim N\left(0, CL_{it-1}^2 \sigma^2_{\Delta HCB}\right)$.

Diameter growth was adapted from Hann et al. (2006) as,

$$\Delta D_i = \exp\begin{pmatrix} \theta_{80} + install_{\Delta D} + plot_{\Delta D} \\ +\theta_{86} BA_{it-1}^{0.5} + \theta_{85}\frac{D_{it-1}}{DQ_{it-1}} \\ +\theta_{81} D_{it-1} + \theta_{82} D_{it-1}^2 \\ +\theta_{83} \ln\frac{CR_{it-1} + 0.2}{1.2} \\ +\theta_{84} \ln(SI - 1.37) \end{pmatrix} + \epsilon_{\Delta} D \quad (8)$$

where θ_{80-86} are estimated parameters, $install_{\Delta D}$ and $plot_{\Delta D}$ are random effects assumed distributed Gaussian with means zero and variances $\sigma^2_{\Delta D_{install}}$ and $\sigma^2_{\Delta D_{plot}}$, respectively, and residual error is weighted as $\epsilon_{\Delta DBH} \sim N\left(0, DBH_{it-1}^2 \sigma^2_{\Delta DBH}\right)$.

Individual tree survival probability was modeled with logistic regression,

$$\begin{aligned} logit(p.surv_{it}) &= \delta_{90} + \delta_{91} D_{it-1} + \theta_{92} SI \\ &+ \theta_{93}\left(1 - \frac{D_{it-1}}{DQ_{it-1}} + \theta_{94} BA_{t-1}\right) \end{aligned} \quad (9)$$

where δ_{90-94} are estimated parameters. Crown length and tree height covariates were excluded from the model

Table 2: Attributes of the individual trees and plots used for constructing the equations.

Attribute	Mean	SD	Min	Max
Individual tree measurements (n = 18,660)				
DBH (cm)	11.84	7.91	0.25	58.67
HT (m)	9.13	6.09	0.83	39.74
HCB (m)	2.21	3.72	0.00	28.44
CL (m)	6.92	3.36	0.49	20.91
CR (%)	83.93	16.33	4.36	100.00
BAL (m^2 ha^{-1})	6.53	8.21	0.00	49.64
GEA (yrs)	11.75	7.34	2.13	52.67
PHT (m yr^{-1})	0.89	0.15	0.41	1.19
ΔDBH (cm yr^{-1})	1.18	0.57	0.01	2.49
ΔHT (m yr^{-1})	0.87	0.28	0.01	1.92
ΔHCB (m yr^{-1})	0.44	0.39	0.00	2.44
Vol_{ib} (m^3)	1.68	3.43	0.01	45.57
Individual measurement periods (n = 690)				
Measurement period (yrs)	2.93	1.03	2.00	8.00
BA (m^2 ha^{-1})	10.08	12.33	0.01	52.08
TPH (# ha^{-1})	770.83	437.77	9.95	1832.74
QMD (cm)	10.95	7.85	0.50	37.67
CCF (%)	95.39	87.87	0.18	347.63
HT100 (m)	9.96	7.06	1.46	39.08
AGE_{BH}	8.57	7.36	0.00	42.00
AGE_{TOT}	14.93	7.84	6.00	51.00
SI (m at 25 yrs AGE_{BH})	21.27	2.72	10.94	26.11
ΔBA (m^2 ha^{-1} yr^{-1})	1.45	0.89	0.01	3.73
ΔHT100 (m yr^{-1})	0.93	0.26	0.18	1.91

to increase the number of trees available for fitting the model, since most trees were not directly measured for these metrics.

Finally, tree-level attributes like D and H are commonly used to predict individual tree volume (Vol) so uncertainty in either of these estimates could significantly influence this estimate. Consequently, a volume equation similar in form to Honer et al. (1965) was derived using a subset of destructively sampled Douglas-fir trees ($n = 337$),

$$Vol_{ib} = \frac{D}{\theta_{100} + \frac{\theta_{101}}{H}} \qquad (10)$$

where Vol_{ib} is the total volume inside bark and $\theta_{100-101}$ are estimated parameters.

Except for the survival as well as H and HCB at t_0 equations, all other equations were fitted using the *nlme* package (Pinheiro and Bates, 2000) in R v3.0.2 (R Development Core Team, 2012). The increment equations were annualized from multi-year measurement interval data using an iterative technique described by Cao (2000) and Weiskittel et al. (2007). All parameters were estimated using nonlinear mixed effects with installation and plot within installation as random effects. Since

these data violated the assumption of homogeneous variance, a power variance function of the primary covariate was incorporated into the fitting function for the equations. The power variance function was common to all installations and plots and was defined in this analysis as s^2 (v) $= |v|^{(2x)}$, where v is the variance covariate, s^2 (v) is the variance function evaluated at v, and x is the variance function coefficient. Missing H and HCB data were fitted directly within the Bayesian model using contemporary parameter estimation techniques (Fig. 2). The fitted parameters were given non-informative priors as ˜N(0,10000), with the residual errors given partially informative priors, σ˜uniform(0,20). Parameter estimates, standard errors, correlation between parameters, as well as residual errors were fitted and directly incorporated into projections in the Bayesian Network.

In contrast to other fitted equations, Bayesian methods were used to fit probability of survival since the observation intervals for the survival data varied between two and eight years, making simple logistic regression inappropriate (Cao 2000). The Bayesian approach handled multi-year observations by modeling the period probability of survival as a product of the individual year probabilities within each measurement period, with the

Table 3: Variable states as inventoried, derived or projected for individual tree i or stand level. Variables only used in projections that are set at current values are shown as time step t-1. Recursive variables are those where the current value affects projections are shown in italics.

Source (time step)	Variable
Inventory (t0)	Di, Hi*, HCBi*, Survi Age
Derived (t0)	Cli, CRi, HD, SI, BA, TPH, CCF
Projected (t-1)	GEA, CCF, BA, TPH
Projected (t)	HD, PHG, HMOD, ΔDi, ΔHi, $\Delta HCBi$, Survival
Derived (t)	D, H, HCB, CR, CCF, BA, TPH, HD, DQ

Notes: Constants, parameters and transformed variables are not shown.

*Often only a subsample is measured.

Figure 2: Missing heights were estimated prior to growth projections with two randomly chosen MCMC iterations shown here (circles and triangles). Only a randomly chosen subset of missing heights on plot 1 are shown for clarity. The estimated regression functions for both MCMC iterations are also plotted.

exception that explanatory covariates retained the values at the start of each measurement period. Priors for the parameters were δ_{1-5} ~N(0,1000). This Bayesian model fitting was done separately, and then parameter estimates, standard errors, and correlation between parameters were inputted as data for the Bayesian Network to predict annual survival probabilities.

2.4 Bayesian model projections

To illustrate model projections, inventory data from three Douglas-fir plots were used as starting points in model projections (Table 4). Each plot was pure Douglas-fir, with a full inventory of D and subsampled H and HCB from single 0.202 ha plots. These data were used in a regional GY cross-model comparison (Johnson 2005), so they represented a range of growing conditions

typical for the region and provided a convenient reference.

Table 4: Douglas-fir stand characteristics (Breast Height Age, Dominant Height, Site Index, Basal Area, number of trees per ha, and Quadratic Mean Diameter) for three 0.202 ha inventoried plots serving as starting conditions for the 40-year model projections. The inventory data is from Johnson (2005) and was available online at www.growthmodel.org.

Variable	Units	Plot 1	Plot 2	Plot 3
Age_{BH}	yrs	29	25	19
Ht_{Dom}	m	22.1	18.4	20.6
SI[a]	m at 25 yrs	19.2	18.4	27.2
BA	m^2 ha^{-1}	33.7	31.8	34.8
NT	# ha^{-1}	1502	1809	949
QD	cm	16.9	15	21.6

[a] Eq. 1 estimated.

In a Bayesian Network, parameters are considered random variables and therefore the full distribution of the parameters are provided, including their means and variances. The component equations fitted to the available Douglas-fir dataset were described by parameters, correlations between parameters, and finally errors, which included the standard errors for parameters, random effects, and random residual errors. All of these statistics were inputted to the model as initial conditions and/or more formally as prior distributions. For each MCMC iteration, component equation parameters were drawn from a multivariate normal (MVN) distribution specified by the parameter means, standard errors, and correlation between parameters. Similarly, random effects were drawn for each MCMC iteration for each component equation as a distributed Gaussian normal with mean zero and the estimated variance component. Parameters and random effects were randomly drawn at the start of each MCMC iteration and then kept con-

stant for each projection year and tree within that single iteration. Residual errors, however, were randomly generated for each tree and projection year within a single MCMC iteration. Residual errors were randomly generated based on the estimated residual errors, but then applied (i.e., added) on the unweighted scale. Individual tree survival in a projection year was randomly determined based on the computed mean probability of survival. A few restrictions were added to the model structure such as not allowing D, H, or HCB to decrease with increasing age, although HD was not similarly constrained. Survival was tracked for individual trees to ensure consistency. Finally, plot average or summation variables such as DQ or TPH were expanded to a per-ha basis and computed within each iteration, and therefore also expressed as posterior densities.

All projections were made with JAGS v3.3.0 (Plummer 2003). Each projection started with the inventory list from the three plots. Initial H and HCB were measured on approximately 15% of the plot trees, so these variables were estimated for trees with missing data using equations 2 and 3 embedded within the Bayesian Network. The base projection of 40 years included all sources of error and uncertainty, including a stochastically applied survival function. A fixed random number "seed" was used for repeatability of results. An initial 25,000 MCMC iterations were used as a "burn-in" period, and discarded. Projections were comprised of 40,000 additional MCMC iterations, retaining every 10th iteration to reduce serial correlation. These 4,000 MCMC iterations formed the posterior distributions of the projected variables. Note that this burn-in period and serial correlation concerns were only necessary due to missing H and HCB data being estimated directly within the projection model. Additionally, we randomly chose 25% of these MCMC iterations to illustrate model behavior.

2.5 Error budgets

One of the principal advantages of a Bayesian Network is the automatic propagation of errors throughout the model. This is achieved since each variable is connected through the conditional relationships depicted by the Bayesian Network. We examined the error budgets of the Douglas-fir GY model by investigating how uncertainty in the component equations (i.e., residual variance, parameter error, random effects, or state uncertainty) influenced system-wide error. Forty-year projections were made for each of the three plots that included all errors (i.e., base projection), where all errors were set to null values (i.e., a deterministic model with all errors set to zero), and finally, where individual error terms were set to null values. In sum, 18 different 40-year projections were run for each plot. Mortality in all of the projections, including the deterministic model, was stochastically applied using the same approach as the base projection and used as a basis of comparison.

3 RESULTS

In this section, we focus on assessing the adequacy of the modeling platform with emphasis on error propagation and model projections. This is done to illustrate the dynamics and flexibility of the Bayesian modeling platform, which is then further expanded on in the Discussion to other useful aspects relevant to ecological modeling.

3.1 Bayesian Networks as a modeling platform

We found the Bayesian modeling platform to be adequately flexible for handling a relatively small, but rather complex ecological model. The model specification in the Bayesian Network is generally straightforward (Supplemental Materials S1), with component equations depicted by easily understandable scripts in the programs WinBUGS or JAGS. One of our primary goals was to develop a modeling platform that would easily incorporate error budgets. As such, the models used contain several distinct errors types, and each was propagated automatically through the Bayesian Network. The models were then compiled and simulated forward 40 years in JAGS for between 60-90 min, depending on the number of simulated trees (192 to 366). Simulations took considerably longer in WinBUGS (the compiling step only) or failed entirely for large datasets, a recognized glitch in WinBUGS with highly recursive models. Results from several smaller models (fewer trees and/or projection yrs) were compared between WinBUGS and JAGS, and were for all purposes, indistinguishable. All model projections were therefore made with JAGS running on a high performance computer cluster at Oklahoma State University. Single JAGS runs, however, were easily done on a desktop computer with adequate memory.

In general, the individual equations fit well and are consistent with expectations (Supplemental Materials S2). Plot-level projected variables are comparable to six other regional growth and yield models widely used for coastal Douglas-fir, including ORGANON (Hann et al. 1993), FVS (Crookston and Dixon 2005), DF-SIM (Curtis et al. 1981), FPS (Arney 2005), DF.GOAB (Weiskittel et al. 2007), SPS (Arney 1985), and TASS (Mitchell 1975). Each model used the same three initial plots for projections (Table 4), and we compared our run against non-thinned and non-fertilized growth projections after 20 years to match their results. Our base-projection dominant heights were 7.8%, on average, higher than six other common Douglas-fir GY models

used in the region, which is likely due to the plantation-specific (versus regional) dominant height growth equation used in our study. Our base-projection stand BA was 1.9 and 3.7% higher relative to the six model average for plots 1 and 2, respectively, while it was 17.7% lower for plot 3. The source of this difference for plot 3 is unclear; however, two of the growth models, including ORGANON (Hann et al. 2003), showed similar projections for plot 3 (within 3.5%). With respect to the predicted 20-year periodic annual increment (PAI), which includes uncertainty from both D and H, the revised DF.GOAB generally had the highest value (except for plot 3), but it was generally near the median estimate from the other models (Table 5).

Table 5: Derived periodic annual increment (m^3 ha^{-1} yr^{-1}) of various growth and yield (GY) models in the Pacific Northwest for three plots after 20 years of prediction. The GY models included DFSIM (Curtis et al. 1981), FPS (Arney 2005), FVS (Crookston and Dixon 2005), DF.GOAB (Weiskittel et al. 2007), SMC-ORGANON (Hann et al. 2003, 2006), SPS (Arney 1985), and TASS (Mitchell 1975) as well as the revised DF.GOAB developed in this analysis, which is presented with a 95% credible interval. For comparison, the overall median with a 95% confidence interval (CI) was computed.

GY Model	Predicted PAI (m^3 ha^{-1} yr^{-1})		
	Plot 1	Plot 2	Plot 3
DFSIM	18.9	19.3	35.6
FPS	15.3	17.8	33.8
FVS	15.8	13.4	17.2
DF.GOAB	16.8	18.8	19.9
DF.GOAB	22.6	24.3	28.7
Revised	22.6	24.3	28.7
(This Study)	20.6, 24.6	22.0, 26.5	24.7, 32.8
ORGANON	17.8	16.2	23.9
SPS	17.9	21.6	30.5
TASS	17.0	20.8	30.4
Median with	17.4	18.5	29.5
95% CI	14.8, 24.8	15.4, 27.6	24.6, 43.6

3.2 Error representation

All sources of uncertainty in the system of equations are represented in the Bayesian probabilistic modeling platform. Parameter errors are represented as MVN for each fitted equation, but are considered independent across equations. The database for equation fitting is extensive, so standard errors on parameter estimates are relatively small. However, residual errors and random

effect are relatively large for reference to model performance (Supplemental Materials S3).

Each iteration of the MCMC represents a full 40-year simulation for all projected trees, and the model took a random MVN draw as the parameter estimates for each iteration. Parameter estimates fully retain the pair-wise linear correlations between parameters, and assign these unique parameter draws to project all individual trees across all years for that single MCMC iteration. Different MCMC iterations use various parameter draws from the same MVN distribution. Similarly, random error is partitioned during equation fitting into installation and plot variance components, along with residual error. Each component is considered independent within and between equations (Littell et al. 2006). The Bayesian Network is flexible enough to model this hierarchical error structure, which is necessary given the typical objective of projecting specific field inventoried plots as discrete units rather than as random draws from a population. This also allows for the potential to calibrate individual plot- and installation-level random effects. However, without calibration data the installation- and plot-level random effects would be applied as full unconditional error terms, additive to residual variance. We assessed installation- and plot-level random effects for across equation correlations, but found these to be mainly non-significant, with the largest correlation coefficient (-0.51) between $plot_{\Delta H}$ and $plot_{\Delta D}$. Residual errors also had low correlations.

The model is also able to incorporate uncertainty in starting conditions and propagate this error throughout the system of equations. These types of GY models primarily use difference form equations, predicting change from a starting condition, rather than states. In our GY model, starting conditions for each tree included D, H, and HCB, where only D was recorded in the field for each tree, with sub-samples for H and HCB. Missing values for H and HCB are then predicted using plot specific regressions. Individual-tree H and HCB estimates are relatively imprecise, and so were entered into the data as actual estimates within the Bayesian model rather than assuming the fitted means from the regression line (Fig. 2). Using a plot-specific equation and data to estimate H and HCB [eqs. 2-3], the Bayesian platform fitted the parameter estimates through a contemporary MCMC approach with non-informative priors (Gelman 2003). We ran a sufficiently long (25,000 burn-in iterations) MCMC and checked to ensure convergence to a stable posterior distribution and good mixing of the chain. Further MCMC iterations in the projections were assumed to adequately represent the posterior distributions for the H and HCB parameters. If a regional H and HCB equation were used (e.g., Robinson and Wykoff 2004), the Bayesian platform is flexible enough to simul-

taneously update priors for the plot random effect (i.e., calibration) using the H and HCB data for the specific plot. It should also be noted that log-bias corrections (Flewelling and Pienaar 1981) are not necessary in a Bayesian Network since the back-transformations to the nominal scale are made at the individual tree-level predictions (Stow et al. 2006).

We modeled individual-tree mortality with logistic regression. For each projection year the model predicted mortality (yes/no) based on the estimated probability of mortality from the fitted model. Randomly drawn parameter estimates were constant across all projection years within a single MCMC iteration. Predicted mortality for each tree in a projection year was taken as a random draw from a Bernoulli distribution with the parameter equal to the projected probability of mortality. Within a single MCMC iteration (i.e., over all yrs), we constrained tree status (live v. dead) to remain dead in future years if a mortality event had occurred. Mortality was the only discrete variable in the model.

Site index is the only required measure of site quality for the ecological system. Predictions of site index for our illustration plots were relatively imprecise, despite the actual age of the plantations being close to the site index base age of 25 years (Fig. 3). Precision will decrease for stand ages further from this base, particularly for younger stands. The site index for plot 2 was known exactly (18.4 m at 25 yrs) because the stand was 25 years old. As with all other estimated variables and parameters, site index is represented in our model as a posterior distribution of possible values. This is distinctly different from other forest GY models that apply a point estimate of site index as a known value. Lastly, for two derived variables (MCW and MCA) we took parameter estimates from the literature, but since these were given without standard errors or correlation between parameters, these parameters were entered into the model deterministically. All derived variables in this model, except some in the starting year, are represented as posterior distributions since they were computed from variables with modeled uncertainty. Certain starting year derived variables, such as BA_0, are based on fully measured data (and thus assumed to be known with certainty and without any measurement error), which can be an important source of uncertainty in forests (e.g., Gertner 1990).

3.3   Error propagation

Error propagation occurred automatically across component equations and projection years for all sources and types of errors, including parameter estimation error, residual error, and random effects, as well as uncertainty in stand- and tree-level starting conditions. Plot-level
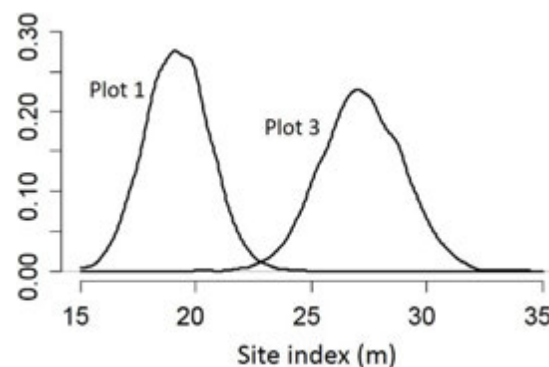
Figure 3: Derived posterior distributions of site index for plots 1 and 3. Site index for plot 2 was known precisely since it was 25 years old.

projections showed error propagation across years, with the uncertainty increasing over time (Fig. 4). For illustration, stand-level variable projections for Plots 1 and 3 are shown for 25 randomly selected MCMC iterations, plotted relative to the mean and 95% credible interval (Fig. 4). For consistency, the same MCMC iterations were used across all variables. Plot 2 behaved similarly to Plot 1, and is not shown. HD projections [Eq. 4] showed wider credible intervals (Fig. 4), despite precise parameter estimates (Supplemental Materials S3). The primary influence was a relatively large residual error that was additively applied to this non-linear equation in successive projection years. This uncertainty in HD projection resulted in the somewhat wider credible intervals for site index (Fig. 3). Projections for stand-level variables, BA, DQ, and TPH, were somewhat less variable. The 95% credible intervals for BA were approximately symmetric, but increased over the projection period to between 14% and 23% of the mean at 40 years for these plots (Fig. 3). DQ showed similar projection errors averaging 18.1% of the mean, while TPH was somewhat less precise, ranging between 24% and 31% of the mean projection.

Individual tree-level projection errors were greater than stand-level errors (Fig. 4). Individual tree projections for D, H, and HCB are influenced by random effects and residual errors, as well as parameter errors. In addition, most trees had an estimated initial H and HCB values, which contributed to projection error. The stand-level projections are a summation or average of the individual trees, so much of the individual tree projection uncertainty is offset by other trees at the stand-level. The tree-level projections are assumed to be conditionally independent in the model, and the relatively large plots (0.202 ha) with a large number of measured trees (192 to 366) allow for considerable averaging. Tree-level projections were not strictly independent, since plot-
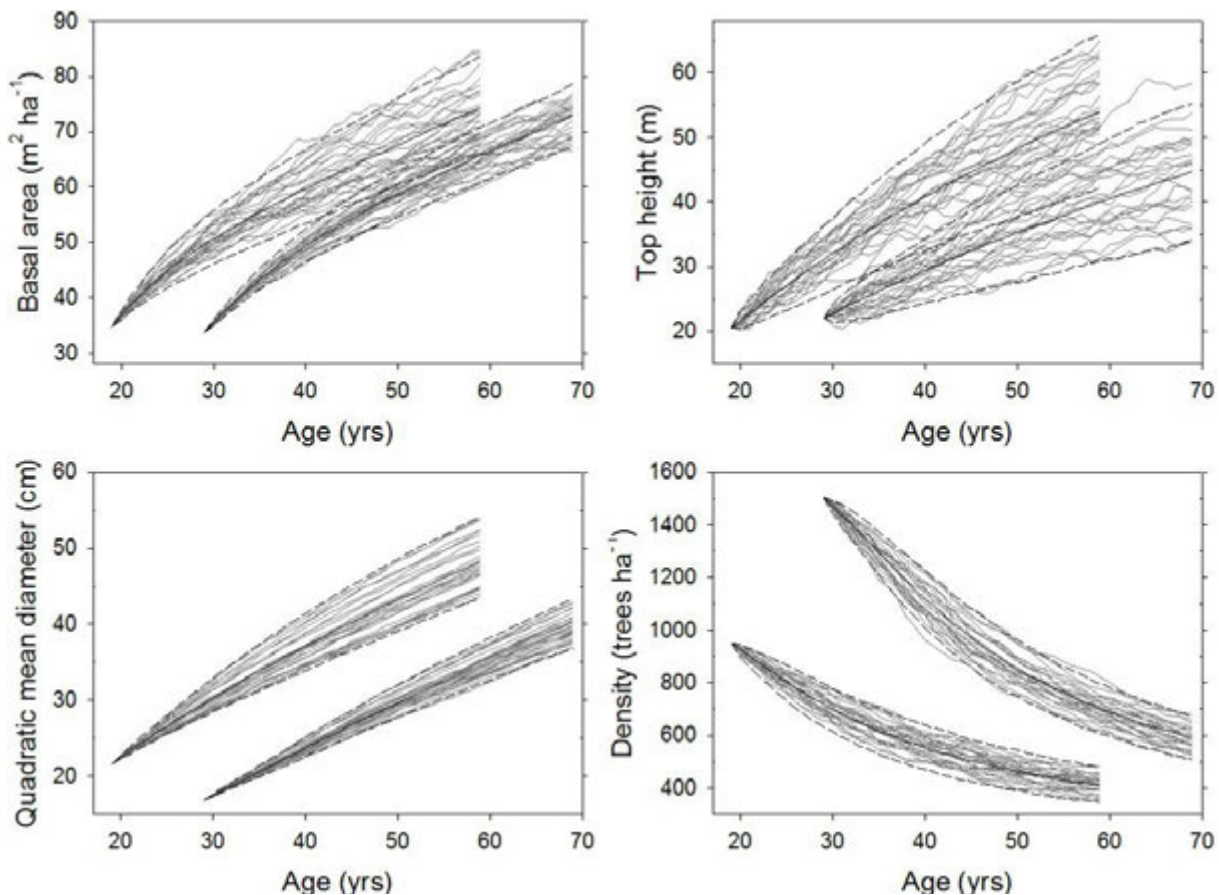
Figure 4: Base 40-yr projections for Plots 1 and 3 of pure Douglas-fir plantations with initial ages of 29 and 19 yrs, respectively. Plot 2 behaved similarly to Plot 1, so was not shown for clarity. Shown are posterior means (solid heavy line), and 95% credible intervals (dashed lines) for plot-level characteristics. Base projections included all sources of error. Also shown are 25 randomly chosen MCMC iterations from the 4,000 iterations comprising the sampled posterior distributions, keeping the same iterations for all variables.

level variables (e.g., BA and CCF) influence these tree-level projections. However, within a single projection step (i.e., age), the trees were considered independent. Also, Figure 4 illustrates the variability in the mortality projections. As expected from ecological understanding, mortality tends to be higher in smaller diameter trees (Fig. 5), but there is also stochastic variability between MCMC iterations. It should also be noted that errors due to parameter and random effects do not average across trees, since these are applied to all trees identically in an MCMC iteration.

3.4 Error budgets

Increasingly wider credible intervals with projection age was simply inherent to the recursive nature of these equations, where prediction error in one year (either tree-level or plot-level variables) were propagated forward to the next year, and so on (Fig. 4). Residual er-

ror was considered independent across trees and years, which is an error structure that serves to reduce across-year effects through partial cancellation depending on the random draw. We did not employ an error structure that was constrained across the entire simulation period, or gradually changing (e.g., autoregressive). Thus, the contribution of residual errors to the system-wide prediction error was generally higher than when accounting for other sources of error (Table 6).

A full error budget for this forest GY model is beyond the scope of this paper, since our primary objective was to develop the modeling framework and evaluate the several advantages it can provide. Projection errors for selected plot-level variables are shown in Table 6 for the full base-projection for Plot 1 only, at the 40th projection year (i.e., all errors included), and individually reduced models, removing one source of error for a single component equation at a time. System-level error is clearly
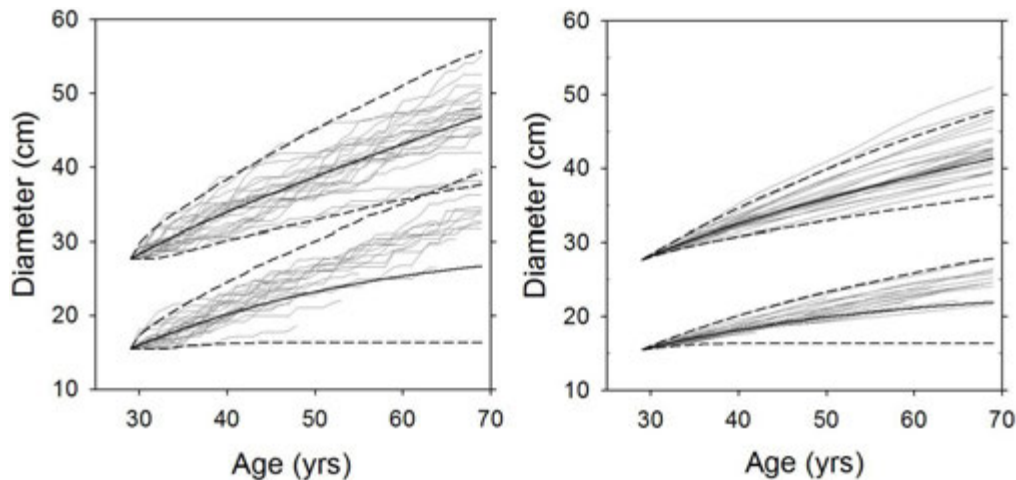
Figure 5: Diameter growth projections for arbitrarily chosen trees 1 and 2 from Plot 1 for the base projection (left) and projection where residual error for diameter growth was set to zero (right). Shown are posterior means (solid heavy line), and 95% credible intervals (dashed lines). Also shown are 25 randomly chosen MCMC iterations from the 4,000 iterations comprising the sampled posterior distributions. Mortality events are shown as an ending line prior to the end of the projection. The credible intervals reflect no diameter growth after a mortality event, so are wider than a dynamic approach that only tracks live trees.

not a simple additive process across all error components (Table 6). HD residual error showed the greatest effect, reducing RMSE from 5.6 to 0.52 m. All HD projections were relatively similar to the deterministic projection, as shown by very small percentage differences. Similarly, H_0 and HCB_0 estimates were similar to deterministic estimates, with the RMSE contributing the most to the variability (Table 6), rather than parameter error. However, all of these equations were non-recursive, and were not dynamic across component equations.

Projections of BA and TPH with and without including residual errors on ΔD illustrate the marked differences between approaches (Fig. 6). The only difference between the two projections is that residual error was set to 0.00001 (i.e., essentially zero). Doing so paradoxically increased projection error, but also strongly reduced the projection difference relative to deterministic projections (Table 6). All three plots behaved similarly, and a simple examination of the component equations explains this model behavior.

The individual tree-level ΔD equation shows a highly recursive and dynamic projection function. In simplified terms, diameter growth in each time-step is a function of the starting diameter, other tree-level variables also influenced by D (H and HCB), as well as plot-level derived variables involving D, H, and HCB. However, the dominant influence causing the differences shown in Figure 5 is due to a strong interaction between an annual-time step residual error on ΔD and the highly recursive and dynamic form of the ΔD equation. As trees in-

crease in diameter, they tend to grow faster, provided all other variables remain unchanged. This is generally accepted in these types of ecological systems, and is reflected in our parameter estimates (Supplemental Materials S3). Thus, by adding residual errors to individual tree growth, if a tree had a positive error (higher ΔD in a year than the expected mean) this larger D would result in a positive feedback, with higher expected ΔD in the subsequent year. Because ΔD is a highly non-linear equation, errors across years do not, on average, sum to zero for individual trees; that is, they do not "cancel out." Neither do these errors cancel out for derived variables such as BA. However, the dynamic nature of the projection system restricted "run-away" D projections since this variable is strongly and negatively related to stand-level BA. Similarly, run-away D and BA were prevented due to a strong interaction with mortality, where the probability of mortality strongly increased with increasing stand BA. There was a strong increase in projected mortality for the projections, including residual error on ΔD (Fig. 5), which resulted in lower projection error. Regarding these comparisons to deterministic projections, it is not clear which projection scenario (i.e., including residual error on ΔD or not) is closer to the actual tree D and BA dynamics. This topic is expanded on in the Discussion (section 4.2). The other stand-level variables (DQ and TPH) behaved similarly, since each one's component equations were similarly non-linear and recursive. The deterministic projection was more similar to the scenario without residual error on ΔD for the reasons stated.

Table 6: Effects of individually removed sources of error for a single component equation, with the component equation and source listed. Results are 40-yr projections for Plot 1 only showing residual error (RMSE), and percentage difference (Diff%) compared to a purely deterministic projection.

Component (Eq #)	Source of error	HD (m)		BA (m^2 ha^{-1})		DQ (cm)		TPH (# ha^{-1})	
		RMSE[a]	Diff%[b]	RMSE	%	RMSE	Diff%	RMSE	Diff%
All	none	5.6	-0.21	2.82	21.17	1.64	18.45	41.18	-18.2
HD (4)	parameters	5.58	-0.23	2.83	21.15	1.65	18.47	41.36	-18.2
HD (4)	residual error	0.52	0.03	2.67	21.16	1.38	18.42	28.65	-18.3
H (2)	parameters	5.54	-0.06	2.8	21.21	1.59	18.46	40.5	-18.1
H (2)	residual error	5.65	-0.19	2.76	21.16	1.58	18.44	40.29	-18.2
HCB (3)	parameters	5.77	0.03	2.8	21.13	1.62	18.48	41.01	-18.3
HCB (3)	residual error	5.63	0.2	2.88	21.22	1.64	18.46	41.23	-18.1
HMOD (6)	parameter	5.6	-0.21	2.83	21.15	1.64	18.46	41.08	-18.2
HMOD (6)	random effects	5.65	-0.21	2.81	21.28	1.62	18.58	41.1	-18.4
ΔH (6)	residual error	5.62	-0.05	2.53	19.07	1.48	16.39	41.02	-15.4
ΔHCB (7)	parameter	5.64	-0.08	2.8	21.28	1.65	18.53	40.92	-18.2
ΔHCB (7)	random effects	5.65	-0.21	2.53	21.4	1.43	18.73	39.09	-18.6
ΔHCB (7)	residual error	5.6	-0.21	4.09	24.77	2.4	22.2	45.31	-23.7
ΔD (8)	random effects	5.65	-0.21	2.24	21.1	1.21	18.35	37.35	-18.1
ΔD (8)	parameters	5.65	-0.21	2.81	21.28	1.62	18.59	41.35	-18.4
ΔD (8)	residual error	5.62	-0.05	3.96	-3.71	2.41	-2.79	68.08	2.6
SURV (9)	parameters	5.65	-0.21	2.71	21.29	1.61	18.57	39.66	-18.3

[a] RMSE is the standard deviation of the posterior distribution for the 40th projection year.

[b] Difference relative to a deterministic projection, computed as 100 * (projection mean - deterministic mean) / projection mean, where the two variables are the means of the posterior distributions. Only one deterministic projection was required.

Random effects and parameter errors did not cause similar differences between our projections and the deterministic one. These types of errors entered into projections in a fundamentally different way, since they were chosen stochastically once and only once at the beginning of each MCMC iterations, and so were fixed across all projection years, whereas residual error was stochastically drawn in each projection step (i.e., each projection year).

4 Discussion

This example from a common class of ecological models, namely a recursive system of discrete, difference form equations, was easily and adequately represented within a Bayesian Network. The MCMC approach allowed for mixed discrete and continuous variables, and was sufficiently flexible to model a rather complex error structure of our example ecological system. We specifically chose to investigate a Bayesian Network as a modeling platform. The main criteria for this platform were: (1) accurate and tractable projections that include full error propagation, (2) flexible and comprehensive analytic capabilities, (3) allows hierarchical and multi-level model structures, (4) capability for random effects calibration, (5) capability for hypothesis testing and updating knowledge across different system components, simultaneously with different sources of information (i.e., new data), (6) computationally fast, and (7) relatively simple implementation, preferably in a scripting language. Finally, while not one of the formal criteria, the probability interpretation of the posterior distributions provides a clear advantage for use within decision support systems (Borsuk et al. 2004) and is the basis for the natural link between Bayesian approaches and risk assessment (Berger 1985).

There are several off-the-shelf programs that implement MCMC in a Bayesian context, including Win-BUGS and JAGS. Commercial packages for implementing Bayesian Networks were not evaluated, but have been used successfully elsewhere (e.g., Borsuk et al. 2004). Most of these commercial packages are limited to either categorical or Gaussian distributions, and are limited in their capacity to simultaneously update parameters with new data. Our choice of open-source or freely available software was guided by the need for both a Bayesian network platform and the desire for simultaneous Bayesian parameter updating used in calibration and parameter estimation, as was done in this study. We also required a Linux version for access to high performance computing facilities, which we would expect to be a common requirement for research use.
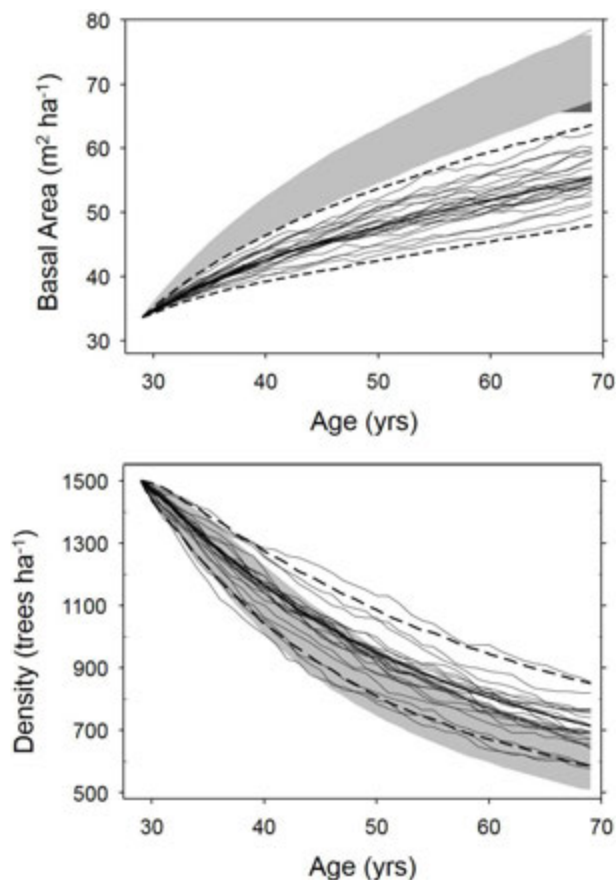
Figure 6: Forty year projections for Plot 1, a pure Douglas-fir plantation, where residual error for diameter growth was set to zero. Shown are posterior means (solid heavy line), and 95% credible intervals (dashed lines), as well as 25 randomly chosen MCMC iterations. For reference, 95% credible intervals for the base projection (all sources of error included) are shown as gray outline, which are identical to results from Fig. 4.

The JAGS program was one possible choice, and was shown to be thoroughly adequate. Alternately, for porting existing ecological models to a Bayesian Network, bespoke code could be written for MCMC samplers (Hojsgaard et al. 2012), and this approach might be preferable for deployment of models past the development stage as it simplifies the amount of recoding of the original model.

The primary difficulty with representing a relatively simple Douglas-fir forest GY model as a Bayesian Network arose from the highly recursive nature of the equations. The graph moralizing step in constructing the model dependencies is not unduly numerically intensive (Nielsen and Jensen 2009), but tended to cause certain Bayesian software programs to have very long "compile" times, WinBUGS and OpenBUGS in particular. Once the model was compiled in JAGS, running 40,000 iter-

ations of the MCMC took 60 to 90 minutes, depending on the number of trees simulated. This compiling issue with highly recursive equations is a known problem with these software packages, but was not exhibited for JAGS. Finally, due to the highly connected nature of Bayesian Networks, the models contained an extremely large number of variables (>100,000), with each variable being represented as a posterior distribution of possible values. These included all parameters, random effects, and residual errors, but also each predicted variable for each projection year was considered a unique variable. These included all predicted H_0 and HCB_0, as well as separate D, H, HCB and survival projections for each tree over each of the forty years in the projection. Derived variables at the tree- or plot-level such as CCF and BA had similar representation, each being expressed as posterior distributions for each time step. This highly connected nature of Bayesian Networks adds to their flexibility, but also to their computational load.

4.1 Error budgets

Plot 1 BA was projected to be 72.8±2.82 m^2 ha^{-1} (mean ±sd) after 40 years, with 95% CI of 67.3 to 78.5 (Fig. 4). Similar statistically valid errors are available for every projected variable on each plot. Automatic error propagation across the system of equations is a distinct advantage of the Bayesian modeling platform. Few existing ecological models include error estimates with the projections (Larocque et al. 2008), but there are notable exceptions (e.g., Borsuk et al. 2004, Kershaw et al. 2017). We feel this is a major shortcoming in ecological modeling, since the usual manner in which model projections are assessed is through estimates of projection error. Without these, it is quite difficult to evaluate the adequacy of point estimates, or make between-model comparisons. Several methods have been developed to provide error estimates, and these can be classified into Monte Carlo–type simulations or analytical solutions (Lo 2005). Bayesian Networks provide an analytical solution (i.e., a full posterior distribution) that is represented through MCMC sampling. Accurate and tractable error estimates are increasingly required for ecological projections, particularly for carbon sequestration where carbon credits are often tied to an estimated lower confidence bound as a conservative measure (Smith and Heath 2001).

Error characterization in ecological models is most often done as a sensitivity analysis, where individual parameters are varied (typically extreme values) and the model re-run sequentially to provide information on which parameters most strongly influence predictions (e.g. Simons-Legaard et al. 2015). In contrast, simple Monte Carlo methods to fully develop error estimates

for a system of equations generally require tens, if not hundreds of thousands of simulations (O'Neill and Gardner 1979). Further, characterizing autocorrelation between parameters and other non-independent variables will necessarily add complexity in specifying the simulation space. There are, however, several methods that have been developed to reduce the number of Monte Carlo simulations, notably orthogonal polynomials (e.g., Gertner et al. 1996, Parysow et al. 2000). Results and efficiency of these methods would be comparable to the Bayesian Network approach used in this study, but with their utility restricted to estimating error budgets.

In its simplest form, a Bayesian Network approach to characterize error propagation can be viewed as simply Monte Carlo simulations of a well-conditioned system of equations. That is, if there is no parameter estimation through new data (e.g., calibration in this context) or other sources of information, then the dependencies of the posterior distributions are explicitly defined as $\Pi(parents)$. Markov Chain Monte Carlo sequentially samples each variable conditional on the current estimate of the other variables, and the dependencies specified in the component equations. The posterior distributions (e.g., Fig. 4) are a collection of equally plausible predictions and form statistically valid error estimates. The primary contrast between Bayesian Network and simple Monte Carlo approaches is in the use of stochastic draws. The Monte Carlo approach explicitly make a random draw from a known (or presumed known) distribution. In cases where parameters are non-independent (e.g., MVN) then random draws for these parameters are conditioned on their covariances. In contrast, random numbers in MCMC are used to compute a proposed step within a random walk (Cappé and Robert 2000). Assuming the current MCMC iteration is sampling from the true posterior distribution for the model, then each subsequent iteration (step) of the MCMC can be used to characterize the posterior distributions. In the simplest Bayesian Network case (i.e., no new information is provided), when starting value for the chain are chosen with the same error distributions and conditional dependencies as in the specified model, then even the first iteration is describing the stationary posterior distribution (Pearl 2000).

A disadvantage to a Bayesian Network approach for error budgets is the inability to provide optimal solutions for reducing system error based on a loss function (Lo 2005). This requires an analytical solution for error propagation that Gaussian error propagation (Lo 2005) can provide. There are several other analytic approaches to error budgets, including orthogonal polynomials (Parysow et al. 2000) and first-order Taylor series approximations of variance (Mowrer 1991). In addition, Bayesian synthesis (Raftery et al. 1995, Green et al. 2000) is well suited to error characterization in similarly structured ecological models. Such an approach is ideally suited to characterizing uncertainty where the model parameters are not estimated from data, which is common in mechanistic models (Green et al. 1999). Bayesian synthesis can be used where the mechanistic model structure and parameters are combined with field-collected data from higher-level projections; e.g., BA and TPH at the starting age in our model. The joint pre-model distribution of all input (model parameters) and output (usually field data) are updated using contemporary Bayesian methods and algorithms to provide a post-model distribution. However, this approach was not applied in this study due to abundant process-level data being available for fitting the component equations.

4.2 Model development and error structure specification

It became quickly evident during model development that incorporating residual errors had a marked influence on projections (Fig. 5). This occurred across all equations that had a recursive component (e.g., ΔD), and it carried forward to derived variables. Using a forest plantation system, Kangas (1997) showed a similarly increasing pattern of projection error over time from a simple Monte Carlo error propagation approach over 50-year projections. Kangas (1997) also indicated a difference between Monte Carlo means and deterministic projections, although not as large as in our study. Previous studies have also demonstrated differences compared to deterministic projections (Dennis et al. 1985; Gertner 1991). In particular, both Mowere (1991) and Kangas (1996) showed clear differences in projections from a structurally similar forest GY model to ours, when residual errors were included. However, while their component equations for ΔD were also recursive, they attributed the resulting difference to non-linear component equations rather than inherent recursiveness. We submit that the difference is due to both non-linear and recursive equation forms.

As mentioned previously, it was not clear which projection of BA (if either) was closer to actual growth dynamics (Fig. 6). None of the earlier studies demonstrating this effect attributed this to a recursive equation, but instead to strictly non-linear equation forms. However, tree growth, as with most, if not all, perennial plants, as well as many animal and other organisms are inherently recursive, even in a competitive, resource scarce environment (Harper 1977, Kremer 1983). This was the basis for the Chapman-Richards growth function (Pienaar and Turnbull 1973), as well as simplified Lotka-Volterra dynamics. Overall, we consider the system of equations used in our model to fairly well represent growth dy-

namics. There are both positive and negative feedback elements with each of the recursive equations, which act to constrain individual tree- and plot-level growth rates. Other sources causing differences in our model include non-recursive equation components such as an estimated site index. The site index predictions, however, were similar compared to the deterministic predictions despite having a strongly non-linear form. This was due to state variables being used in site index computations (i.e., top-height and age were field measured without likely significant measurement error). However, site index was still estimated with relatively high error (Fig. 2) and this variable enters into growth increment equations non-linearly, which can therefore result in different mean projections from deterministic ones (Table 6). Although these impacts were considerably smaller than the impacts of including residual errors to other component equations, they remain a potentially significant source of uncertainty.

In this study, we simply noted the marked difference between probabilistic and deterministic growth projections, as have other studies. This is clearly a topic that needs to be resolved, as Bayesian model projections are becoming more common (Clark and Gelfand 2006). The relatively sparse modeling dataset makes it difficult to determine which projection approach is most accurate. We have numerous tree measurements across 167 plots; however, the mean total measurement period for a plot was 12.8 years, with a sd of 4.4. With such short time-series, prediction error is difficult to assess for recursive equations. There is considerable additional effort required to fully and accurately characterize the error structures, even for a relatively simple ecological model as this. Nevertheless, we do not view this as a disadvantage as prediction accuracy does depend on correct model specification, including correct error specification. In addition, model building with a specific aim to reduce projection error will help refine data collection efforts.

4.3   Research integration

The ability of Bayesian Networks to learn and communicate information by representing the model systems conditional probabilities (Pearl 1988) is, in our opinion, the greatest advantage of the probabilistic modeling framework described in this study. It is also the principal feature that distinguishes Bayesian Networks over alternate modeling platforms. Many of the other attributes, such as error propagation, have clear corollaries with other analytic techniques; Bayesian Networks provide this collection of advantages in a single platform, with existing MCMC programs greatly simplifying implementation. In addition, Bayesian Networks also allow for very natural and easily understood integration

among discipline-specific models. There has always been a need to integrate research across disciplines, but most modeling efforts remain discipline specific. The need for integrating research across once disparate disciplines is becoming critical as ecological risks associated with climate change become more apparent (Willows and Connell 2003).

A direct application of this approach is potentially where site index estimates are unavailable for an established stand; for example, due to past disease or physical damage to dominant trees so that the dominant trees were not "free to grow" over the entire stand development, and do not represent site potential. In these cases, the GY models are often applied regardless, but using an approximate estimate of site index or a regional average. However, in the probabilistic model developed in this study, it is entirely possible to substitute field collected ΔD or ΔH information to directly estimate the site index, such that information or "message-passing" occurs opposite of the direction of influence (Pearl 1988). For example, a weakly informative uniform distribution for site index could be used as a prior distribution that is then updated with new response data in a conventional Bayesian approach. Multiple types of data can then be used simultaneously to predict site index.

This information passing and learning feature of Bayesian Networks as well as the ability to incorporate new data into model projections forms much of the basis for "model-data fusion" approaches in ecological modeling (Fox et al. 2009, Hobbs and Ogle 2011), where model parameters and to some extent model structure are in constant flux as new data become available. Model-data fusion approaches are fundamentally similar to the probabilistic modeling approach taken in our study, although we emphasized model projections and error budgets. The Bayesian Network as constructed could immediately be applied to investigate typical model-data fusion questions (Hobbs and Ogle 2011). Data assimilation is another related approach usually implemented around a Kalman filter approach (Houtekamer and Mitchell 1998). The somewhat seamless interplay between contemporary Bayesian parameter estimation and model projection should be emphasized; both were done simultaneously in this study. The flexible Bayesian parameter estimation approach (Berger 2000, Wikle 2003, Clark and Gelfand 2006) carries forward into a similarly flexible modeling platform for projections within a Bayesian Network. Our example of a relatively complex system of equations was only one class of ecological model that this framework can express.

Importantly, the information passing and learning feature of Bayesian Networks could be applied across very different disciplines, such as combining economics and

political science with ecological or hydrology models. The conditional probability representation of the ecological or social processes provides a common language across research domains, while the casual representation allows formal hypothesis testing (in a scientific sense) across research domains. Formal hypothesis testing across research domains is further aided by the probabilistic interpretation of the posterior distributions. A flexible probabilistic platform is critical here. Several good examples exist of Bayesian Networks in ecological modeling; however, most of these are constrained to categorical-only variables (e.g., ARIES model of ecosystem services; Villa et al. 2009) or have limited ability to include new data for learning or hypothesis testing (e.g., Marcot et al. 2006). The simplest case of cross-domain hypothesis testing, or learning, comes when projected outcomes of one model are independent variables (i.e., with no parents) in a different model. For more complex cross-discipline integration, parameter identifiability will be challenging, and will need to be addressed through multiple data sources to isolate the processes of interest.

5 Summary

There are many sources of uncertainty present (e.g., systematic error, measurement error, parameter uncertainty) in ecological models (Uusitalo et al. 2015). Bayesian Networks are becoming an important approach for quantifying and understanding this uncertainty (e.g., Uusitalo 2007, Barton et al. 2012). To our knowledge, this is the first development and use of a Bayesian Network to quantify uncertainty in a commonly used individual-tree forest GY modeling framework. The overall projections were consistent with several other regional GY models in the Pacific Northwest (e.g., Johnson 2005); however, the projections showed a relatively high degree of uncertainty in both estimates of stand- and tree-level attributes even though the framework only addressed uncertainty in the underlying prediction equations and not other additional factors like measurement error. This uncertainty significantly increased with the length of the projection, which is an important finding as most GY projections for both scientific and practical planning purposes are often 50 to 100 years in length. The model forecasts were highly sensitive to error associated with ΔD given that it is the primary variable in this type of modeling framework. This relatively high uncertainty can complicate interpretation of derived GY model outputs, which are generally used for forest financial assessments or planning of specific management activities (e.g., Weiskittel et al. 2016b). Recognition of this uncertainty and effective incorporation into derived outputs is necessary for realistic and optimal representation of the system being forecasted. Overall, the Bayesian probabilistic framework presented here highlights the importance of uncertainty in deterministic modeling and is flexible enough to be adapted to other regions, modeling approaches, or ecological systems.

References

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling.* 229:25-36.

Arney, J. D. (1985). A modeling strategy for the growth projection of managed stands. *Canadian Journal of Forest Research, 15*(3), 511-518.

Arney, J.D. (2005). Forest Projection System. Forest Biometrics Institute. Portland, OR, USA. http://www.forestbiometrics.com/

Arnold, J. G., & Fohrer, N. (2005). SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes, 19*(3), 563-572.

Barton, D.N., Kuikka, S., Varis, O., Uusitalo, L., Henriksen, H.J., Borsuk, M., de la Hera, A., Farmani, R., Johnson, S. & Linnell, J.D. (2012). Bayesian networks in environmental and resource management. *Integrated Environmental Assessment and Management, 8*(3), 418-429.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis.* Springer.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, *95*(452), 1269-1276.

Borsuk, M. E., Stow, C. A., & Reckhow, K. H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, *173*(2), 219-239.

Canham, C.D., J.J. Cole and W.K. Lauenroth. (2003). *Models in Ecosystem Science*. Princeton University Press. 504 pp.

Cao, V. (2000). Prediction of annual diameter growth and survival for individual trees from periodic measurements. *Forest Science*, *46*(1), 127-131.

Cappé, O., & Robert, C. P. (2000). Markov chain Monte Carlo: 10 years and still running!. *Journal of the American Statistical Association*, *95*(452), 1282-1286.

Cieszewski, C. J. (2003). Developing a well-behaved dynamic site equation using a modified Hossfeld IV function Y3=(axm)/(c+ xm-1), a simplified mixed-model and scant subalpine fir data. *Forest science*, *49*(4), 539-554.

Clark, J. S., & Gelfand, A. E. (2006). A future for models and data in environmental science. *Trends in Ecology & Evolution*, *21*(7), 375-380.

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*(1), 60-80.

Curtis, R. O., Clendenen, G. W., & DeMars, D. J. (1981). *A new stand simulator for coast Douglas-fir: DFSIM user's guide.* US Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. 79 p.

Dennis, B., Brown, B. E., Stage, A. R., Burkhart, H. E., & Clark, S. (1985). Problems of modeling growth and yield of renewable resources. *The American Statistician*, *39*(4), 374-383.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling.* Princeton University Press.

Flewelling, J. W., & Pienaar, L. V. (1981). Multiplicative regression with lognormal errors. *Forest Science*, *27*(2), 281-289.

Freni G., & Mannina G. (2010). Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology,* 392, (1-2), 31-39.

Fox, A., Williams, M., Richardson, A. D., et al. (2009). The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10), 1597-1615.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* CRC Press.

Gertner, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6), 800-804.

Gertner, G. (1991). Prediction bias and response surface curvature. *Forest Science*, 37(3), 755-765.

Gertner, G., Parysow, P., & Guan, B. (1996). Projection variance partitioning of a conceptual forest growth model with orthogonal polynomials. *Forest Science*, 42(4), 474-486.

Green, E. J., MacFarlane, D. W., Valentine, H. T., & Strawderman, W. E. (1999). Assessing uncertainty in a stand growth model by Bayesian synthesis. *Forest Science*, 45(4), 528-538.

Green, E. J., MacFarlane, D. W., & Valentine, H. T. (2000). Bayesian synthesis for quantifying uncertainty in predictions from process models. *Tree Physiology*, 20(5-6), 415-419.

Hall, D. B., & Clutter, M. (2004). Multivariate multilevel nonlinear mixed effects models for timber yield predictions. *Biometrics*, 60(1), 16-24.

Hann, D. W., Olsen, C. L., & Hester, A. S. (1993). *ORGANON user's manual.* Department of Forest Resources, Oregon State University.

Hann, D. W. (1999). An adjustable predictor of crown profile for stand-grown Douglas-fir trees. *Forest Science*, 45(2), 217-225.

Hann, D. W., & Hanus, M. L. (2004). Evaluation of nonspatial approaches and equation forms used to predict tree crown recession. *Canadian Journal of Forest Research*, 34(10), 1993-2003.

Hann, D. W., Marshall, D. D., & Hanus, M. L. (2003). *Equations for predicting height-to-crown-base, 5-year diameter-growth rate, 5-year height-growth rate, 5-year mortality rate, and maximum size-density trajectory for Douglas-fir and western hemlock in the coastal region of the Pacific Northwest.* Forest Research Laboratory, Oregon State University. Corvallis, OR, USA.

Hann, D. W., Marshall, D. D., & Hanus, M. L. (2006). *Reanalysis of the SMC-ORGANON equations for diameter-growth rate, height-growth rate, and mortality rate of Douglas-fir.* Forest Research Laboratory, Oregon State University. Corvallis, OR, USA.

Harper, J. l., 1977. *Population biology of plants.* Academic Press, London, UK.

Hobbs, N. T., & Ogle, K. (2011). Introducing data-model assimilation to students of ecology. *Ecological Applications*, 21(5), 1537-1545.

Højsgaard, S., Edwards, D., & Lauritzen, S. (2012). *Graphical models with R.* Springer.

Honer, T. G. (1965). A new total cubic foot volume function. *The Forestry Chronicle*, *41*(4), 476-493.

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3).

Johnson G.P. 2005. Growth model runoff II. Portland, OR: Growth Model Users Group. 61 p. Available online: www.growthmodel.org.

Kangas, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4), 281-290.

Kangas, A. S. (1997). On the prediction bias and variance in long-term growth projections. *Forest Ecology and Management*, 96(3), 207-216.

Kershaw, J.A., Weiskittel, A.R., Lavigne, M.B. and McGarrigle, E. (2017). An imputation/copula-based stochastic individual tree growth model for mixed species Acadian forests: a case study using the Nova Scotia permanent sample plot network. *Forest Ecosystems*, 4(1), 1-15.

Kremer, J. N. (1983). Ecological implications of parameter uncertainty in stochastic simulation. *Ecological Modelling*, 18(3), 187-207.

Larocque, G. R., Bhatti, J. S., Boutin, R., & Chertov, O. (2008). Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation.*Ecological Modelling*, 219(3), 400-412.

LeMay, V. and Marshall, P. (eds), 2001. Forest Modelling for Ecosystem Management, Forest Certification and Sustainable Management. Proceedings of the Conference held in Vancouver BC, Canada, 12-17 August 2001. University of British Columbia, Vancouver, 506 pp.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). SAS for mixed models. SAS Institute Inc. Cary, NC, USA.

Lo, E. (2005). Gaussian error propagation applied to ecological data: post-ice-storm-downed woody biomass. *Ecological Monographs*, 75(4), 451-466.

MacFarlane, D.W., Green, E.J. & Valentine, H.T. 2000. Incorporating uncertainty into the parameters of a forest process model. *Ecological Modelling*, 134(1), 27-40.

Maguire, D. A., Bennett, W. S., Kershaw, J. A., Gonyea, R., & Chappell, H. N. (1991). *Establishment report: Stand Management Cooperative silviculture project field installations.* College of Forestry, University of Washington, Seattle, WA, 42 p.

Marcot, B. G., Steventon, J. D., Sutherland, G. D., & McCann, R. K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, 36(12), 3063-3074.

McComb, W. C., Spies, T. A., & Emmingham, W. H. (1993). Douglas-fir forests: managing for timber and mature-forest habitat. *Journal of Forestry*, 91:31-42.

McGarrigle E, Kershaw JA Jr, Ducey MJ, Lavigne MB (2013) A new approach to modeling stand-level dynamics based on informed random walks: influence of bandwidth and sample size. *Forestry* 86, 377–389

Mitchell, K. J. (1975). Dynamics and simulated yield of Douglas-fir. *Forest Science Monograph*, 17, 1–39.

Mowrer, H. T. (1991). Estimating components of propagated variance in growth simulation model projections. *Canadian Journal of Forest Research*, 21(3), 379-386.

Nielsen, T. D., & Jensen, F. V. (2009). *Bayesian networks and decision graphs.* Springer.

O'Neill, R. V., & Gardner, R. H. (1979). Sources of uncertainty in ecological models. *Methodology in Systems Modelling and Simulation.* North-Holland, Amsterdam, pp. 447-463.

Parysow, P., Gertner, G., & Westervelt, J. (2000). Efficient approximation for building error budgets for process models. *Ecological Modelling*, 135(2), 111-125.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann. San Francisco, CA.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference* (Vol. 29). MIT Press. Cambridge, MA.

Pienaar, L. V., & Turnbull, K. J. (1973). The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. *Forest Science*, 19(1), 2-22.

Pienaar, L. V., & Shiver, B. D. (1986). Basal area prediction and projection equations for pine plantations. *Forest Science*, 32(3), 626-633.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. Springer. New York, NY, USA.

Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. pp. 20-22.

R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Raftery, A. E., Givens, G. H., & Zeh, J. E. (1995). Inference from a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association*, 90(430), 402-416.

Robinson, A. P., & Wykoff, W. R. (2004). Imputing missing height measures using a mixed-effects modeling strategy. *Canadian Journal of Forest Research*, 34(12), 2492-2500.

Running, S. W., & Gower, S. T. (1991). FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. *Tree Physiology*, 9(1-2), 147-160.

Scheller, R. M., Domingo, J. B., Sturtevant, B. R., Williams, J. S., Rudy, A., Gustafson, E. J., & Mladenoff, D. J. (2007). Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecological Modelling*, 201(3), 409-419.

Simons-Legaard, E., Legaard, K. and Weiskittel, A. (2015). Predicting aboveground biomass with LANDIS-II: A global and temporal analysis of parameter sensitivity. *Ecological Modelling*, 313, 325-332.

Smith, J. E., & Heath, L. S. (2001). Identifying influences on model uncertainty: an application using a forest carbon budget model. *Environmental Management*, 27(2), 253-267.

Stow, C. A., Reckhow, K. H., & Qian, S. S. (2006). A Bayesian approach to retransformation bias in transformed regression. *Ecology*, 87(6), 1472-1477.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203(3-4), 312-318.

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software*, 63, 24-31.

Villa, F., Ceroni, M., Bagstad, K., Johnson, G., & Krivov, S. (2009). ARIES (Artificial Intelligence for Ecosystem Services): A new tool for ecosystem services assessment, planning, and valuation. In: *11Th annual BIOECON conference on economic instruments to enhance the conservation and sustainable use of biodiversity, conference proceedings.* Venice, Italy.

Weiskittel, A. R., Garber, S. M., Johnson, G. P., Maguire, D. A., & Monserud, R. A. (2007). Annualized diameter and height growth equations for Pacific Northwest plantation-grown Douglas-fir, western hemlock, and red alder. *Forest Ecology and Management*, 250(3), 266-278.

Weiskittel, A.R., Kershaw Jr, J.A., Crookston, N.L., & Hennigar, C.R. (2017). The Acadian Variant of the Forest Vegetation Simulator: Continued development and evaluation. In: Keyser, C. & Keyser, T., Eds. Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. US Forest Service. 10-13 pp.

Weiskittel, A.R., Kuehne, C., McTague, J.P., & Oppenheimer, M. (2016a). Development and evaluation of an individual tree growth and yield model for the mixed species forest of the Adirondacks Region of New York. USA. *Forest Ecosystems*, 3, 26.

Weiskittel, A.R., Wilson, D.S., Kuehne, C. (2016b). Forecasting Douglas-fir response to forest management: evaluating alternative approaches and growth model projection uncertainty. 2016 Western Mensurationists Annual Meeting. Skamania Lodge, Stevenson, WA. Available online at http://www.westernmensurationists.org/m2016/Weiskittel_Aaron.pdf

Weiskittel, A.R., Hann, D.W., Kershaw Jr., J.A., & Vanclay, J.K. (2011). *Forest Growth and Yield Modeling*. Wiley. 430 pp.

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6), 1382-1394.

Willows, R.I. & Connell, R.K. (Eds.). (2003). *Climate adaptation: Risk, uncertainty and decision-making.* UKCIP Technical Report. UKCIP, Oxford, UK.

Wilson, D. S., & Puettmann, K. J. (2007). Density management and biodiversity in young Douglas-fir forests: challenges of managing across scales. *Forest Ecology and Management*, 246(1), 123-134.

Wilson, D. S., Stoddard, M. A., & Puettmann, K. J. (2008). Monitoring amphibian populations with incomplete survey information using a Bayesian probabilistic model. E*cological Modelling*, 214(2), 210-218.

Wilson, D. S., Anderson, P. D., & Puettmann, K. J. (2009). Evaluating the consistency of understorey vegetation response to forest thinning through synthetic analysis of operational-scale experiments. *Forestry*, 82(5), 583-596.

A Supplemental Materials

A.1   Supplemental Materials S1

JAGS script that implements the Bayesian probabilistic modeling platform for the developed individual-tree growth and yield model.

A.2   Supplemental Materials S2

Model form and fit statistics for each of the equations used in the analysis.

A.3   Supplemental Materials S3

Parameter estimates, standard errors, and associated p-values for the models used in this analysis.