RESEARCH ARTICLE

Bridging implementation gaps to connect large ecological datasets and complex models

Ann M. Raiho[1] | E. Fleur Nicklen[2] | Adrianna C. Foster[3] | Carl A. Roland[2] |
Mevin B. Hooten[1,4,5]

[1]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA

[2]Denali National Park and Preserve, National Park Service, Fairbanks, Alaska, USA

[3]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA

[4]Department of Statistics, Colorado State University, Fort Collins, Colorado, USA

[5]Colorado Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Fort Collins, Colorado, USA

Correspondence
Ann M. Raiho, 1474 Campus Delivery, Fort Collins, CO, USA.
Email: ann.raiho@gmail.com

Funding information
DoD SERDP contract, Grant/Award Number: RC18-1183; National Parks Service; NASA Arctic Boreal Vulnerability Experiment (ABoVE), Grant/Award Number: 80NSSC19M0112; NSF DEB, Grant/Award Number: 1927177

Abstract

Merging robust statistical methods with complex simulation models is a frontier for improving ecological inference and forecasting. However, bringing these tools together is not always straightforward. Matching data with model output, determining starting conditions, and addressing high dimensionality are some of the complexities that arise when attempting to incorporate ecological field data with mechanistic models directly using sophisticated statistical methods. To illustrate these complexities and pragmatic paths forward, we present an analysis using tree-ring basal area reconstructions in Denali National Park (DNPP) to constrain successional trajectories of two spruce species (*Picea mariana* and *Picea glauca*) simulated by a forest gap model, University of Virginia Forest Model Enhanced—UVAFME. Through this process, we provide preliminary ecological inference about the long-term competitive dynamics between slow-growing *P. mariana* and relatively faster-growing *P. glauca*. Incorporating tree-ring data into UVAFME allowed us to estimate a bias correction for stand age with improved parameter estimates. We found that higher parameter values for *P. mariana* minimum growth under stress and *P. glauca* maximum growth rate were key to improving simulations of coexistence, agreeing with recent research that faster-growing *P. glauca* may outcompete *P. mariana* under climate change scenarios. The implementation challenges we highlight are a crucial part of the conversation for how to bring models together with data to improve ecological inference and forecasting.

KEYWORDS
boreal forest, Denali National Park, dendroecology, forest gap models, model calibration, *Picea glauca*, *Picea mariana*, surrogate modeling

1 | INTRODUCTION

Increasingly, computational simulations of complex individual-level patterns are used to support ecological understanding and forecasting (Dietze et al., 2018). Simultaneously, modern statistical techniques, such as machine learning, are rapidly being developed to provide predictions using "big" data in real time (Christin et al., 2019; Luo et al., 2011; Peters et al., 2014). Computational models alone

can be incomplete ecosystem representations because of possible misspecifications or omissions of processes, and machine-learning approaches alone may not provide insights about mechanisms. Similarly, ecological field data can never fully observe an ecological system. Combining these techniques to understand mechanisms and improve predictions is a promising path forward (Reichstein et al., 2019; Wikle & Hooten, 2010). Examples of such ecological applications (Fer et al., 2018; Oberpriller et al., 2021; Pietzsch et al., 2020; Speich et al., 2021; Tao et al., 2020) are critical for moving beyond implementation barriers and solving urgent ecological problems by bringing these promising tools to larger application-based audiences.

We focus on an urgent ecological problem in interior Alaska where changes in the environment, such as permafrost degradation (Hinzman et al., 2006; Osterkamp, 2005; Osterkamp et al., 2009; Osterkamp & Romanovsky, 1999; Sniderhan & Baltzer, 2016), have occurred over large areas and are beginning to alter boreal forest distributions (Jorgenson et al., 2001) including within national parks. In Denali National Park, Nicklen et al. (2021) found that the climate growth responses of two dominant spruce (*Picea*) species, but especially *Picea glauca*, depended on the presence of near-surface permafrost. Warm summer conditions increased *P. glauca* radial growth on well-drained soils, but strongly reduced growth when growing on near-surface permafrost. *Picea mariana*, currently occupying permafrost terrain, may suffer range contraction under climate change as competitor *P. glauca* expands into more suitable terrain as permafrost thaws (ACIA, 2005; Wirth et al., 2008). Contrary to these predictions, increased low-severity fire frequency (e.g., fraction surface carbon consumed) may allow for continued self-replacement successional trajectories (Johnstone & Chapin, 2006; Johnstone et al., 2010; Kane et al., 2007). A comprehensive assessment of the current status of forests in Denali National Park in Roland et al. (2013) showed that *P. mariana* and *P. glauca* do not typically co-occur. However, Roland et al. (2013) also found that sites where *P. glauca* were most common were also the sites where *P. mariana* achieved greatest diameters. This suggests that the current exclusion of *P. mariana* from areas dominated by *P. glauca* may be related to competition between the two species not necessarily niche limitations. *Picea mariana* is slow to grow early in succession (Gutsell & Johnson, 2002; Wagner & Robinson, 2006) and may be hidden from current assessments that only consider modern data for future predictions (Foster et al., 2019). However, these predictions focus on observations from the modern landscape and do not incorporate long-term (i.e., centennial) data, which may be relevant for predicting long-term system behavior. To disentangle this ecological problem and provide national park managers and visitors with predictions of forest response to climate change, we can improve mechanisms portrayed within ecosystem models by combining model simulations with long-term field data.

Combining observations with computational models statistically can be formally conducted through model calibration. In purely statistical models, this process is simply referred to as "model fitting" where uncertainty in a selected model parameterization can be accounted for in a statistical context where data, process, and parameter models may be developed to make posterior parameter inference (Berliner, 1996). But, it is often computationally infeasible to embed a fully computational model within this type of statistical model directly. For computational models, many different model calibration approaches can be taken ranging from informal to formal statistical techniques. A more informal approach involves "tuning" parameters where the user alters parameters until the output more closely matches the observations. More formal techniques for model calibration involve either informing parameters directly with data (LeBauer et al., 2013) or iteratively updating parameters or parameter distributions as data become available (Fer et al., 2018; Oberpriller et al., 2021; Pietzsch et al., 2020; Speich et al., 2021; Tao et al., 2020). Similarly, modelers may choose to calibrate all parameters in a computational model, which we refer to as "full model calibration," or a modeler can choose to focus on the most sensitive model parameters using an "a priori" parameter sensitivity analysis.

In this study, we perform calibration of an agent-based model focusing on the most sensitive parameters. Agent-based models are computational models used in disease, animal, forest ecology, and beyond to simulate the patterns of individuals, then calculate how the aggregation of those individual patterns leads to emergent system properties (Banks & Hooten, 2021; Hooten et al., 2020). For example, human mobility in an agent-based model may lead to population-level spread of a disease. Connecting data with individual-based models can be difficult because it is unclear how the individual data relate to the emergent process of interest, such as vector-borne disease spread among a population (Perkins et al., 2019). In agent-based models of human mobility, data are not typically used directly to constrain the human movement process itself, but instead data can be used to constrain parameters that drive the spread of the disease between individual agents (Venkatramanan et al., 2018). Similarly, in forest ecology, data collected at the individual tree level can be used to constrain growth parameters in individual-based models, which leads to an improved understanding of long-term forest change.

Constraining mechanistic ecological simulation model parameters, however, remains challenging. Mechanistic ecological simulators often have detailed process representations that are aimed at improving realism at the cost of making models computationally expensive. Parallelization of model calculations can greatly reduce computational times. Yet, during model development stages, especially for the non-specialist, this is often not a worthwhile solution because it requires a large investment in computing resources and development of parallelization or simulation code. In this instance, we consider a "non-specialist" a person who could expand model-based ecological interpretation and hypothesis testing but does not have any experience with simulation modeling (Fer et al., 2021). A "non-specialist" may be an expert in another area that is needed to improve model simulation capabilities, but a "non-specialist" does not specialize in the simulation model code or implementation itself (e.g., a formally trained statistician or forest ecologist). Software packages exist that remove both computational and statistical barriers for performing ecological model calibration as a non-specialist

(Fer et al., 2018). However, this software, while integral to model calibration research, is in early stages of development, has yet to be applied in a broad ecological context, and is therefore not applicable to all branches of ecology. Software packages exist that remove statistical barriers (Hartig et al., 2019; Speich et al., 2021), but these packages require the user to perform the model simulations and make consequential decisions about the experimental design of the model simulations. To solve complex ecological problems, ecologists should be involved in the simulation modeling improvement process and more examples from branches of ecology using a variety of statistical methods are needed.

Approximating the computer model using a statistical surrogate or emulator has become a popular solution to model calibration (Gramacy, 2020). Surrogates are statistical representations of computer models and are developed using the paired input and output from a computer model to train a statistical surrogate that is used in the calibration process. There are many types of statistical models that can be used as surrogates such as deep neural networks, generalized additive models, and Gaussian process models. Gaussian process models (GPs) have become popular because GPs are highly accurate at out-of-sample prediction, allow for uncertainty quantification, and are analytically tractable (Bijak et al., 2013; Heitmann et al., 2013; Santner et al., 2018; Verrelst et al., 2016). GPs have been used as successful computer model surrogates in many contexts including engineering, physics, and Earth system modeling. Only recently have they been proposed as a useful framework for calibration of agent-based models in ecology (Banks & Hooten, 2021; Hooten et al., 2020). Furthermore, it has been shown that GPs can account for model structural error and model parameter error simultaneously (Oberpriller et al., 2021), offering an advantage to conventional model calibration techniques (Hartig et al., 2011). As GPs become increasingly popular, many statistical software packages built for the non-specialist are being developed (Erickson et al., 2018; Fer et al., 2018; Liu et al., 2020). While these packages remove implementation barriers, some statistical choices may not be obvious for a non-specialist because they are based on either open scientific questions or practical rules of the model calibration process.

The non-specialist may have to make several choices about how to create the simulations or morph existing simulation output to match the data-generating process. These choices affect the parameter estimates and are rarely discussed. To highlight these challenges in an accessible context, we present an example using tree-ring data collected in Denali National Park to calibrate species parameters in a state-of-the-art individual-based forest model (University of Virginia Forest Model Enhanced, UVAFME; Foster et al., 2019) illustrating a method not commonly used in ecology that may connect an ecological computer model with field data using a Gaussian process surrogate. This example demonstrates the complexities of three statistical modeling choices that are common in many ecological settings: (1) matching the field observations with model output, (2) determining the correct model start time (i.e., initial conditions) from data, and (3) fitting a high-dimensional Gaussian process surrogate to calibrate the model parameters. This example based on UVAFME

involves forest succession, a foundational ecological process, in a region where competitive relationships between species may drive biome-level changes (Wirth et al., 2008). Successional trajectories can be observed in tree-ring data but are seldom used in conjunction with models of succession because of the issues listed above. We calibrate UVAFME using tree-ring-derived basal area to gain meaningful ecological inference and show how we pragmatically overcome the three aforementioned issues in the procedure.

2 | MATERIALS AND METHODS

2.1 | Study area and tree-ring data

Denali National Park (DNPP) comprises a variety of habitats because of wide elevation and topographic gradients (i.e., slopes and aspects). Tree species in DNPP have unique life history characteristics and thus occupy largely distinct habitats with a few exceptions. The most common tree species in DNPP is *Picea glauca*, followed by *Picea mariana*. *Picea glauca* is a relatively faster-growing species that typically occupies well-drained sites with warmer, more basic, and deeply thawed soils, while *P. mariana* tends to occupy areas influenced by near-surface permafrost with thin active layers and cold, wet, acidic soils (Islam et al., 2003; Roland et al., 2013; Van Cleve et al., 1983). Forests in DNPP are under increasing pressure from climate change as temperatures, growing season length, and disturbance frequency and severity increase (Kelly et al., 2013; Roland et al., 2019). These accelerated changes have resulted in predictions that the boreal forest, including DNPP, is at the edge of a biome-level tipping point (Scheffer et al., 2012) where water-limited spruce-dominated forests could be replaced by post-fire early recruiting broadleaf-dominated forests. Landscape features, such as organic matter depth or wetness rating, are seldom directly considered in regional predictions of species distributions and may offer refugia or buffer from extreme changes (Raiho et al., In Revision; Scharf et al., 2021).

To predict future distributions of *P. mariana* and *P. glauca*, we need a mechanistic understanding of the competitive relationship between the two species informed by long-term data. Because Denali National Park occupies a large region (~4.7 million acres), data for model calibration are typically collected from satellite observations or from field sites along roads. However, these modern snapshot data do not hold information about successional trajectories or competitive dynamics between species. Tree-ring data are well suited to help us gain insights about forest succession (Fritts & Swetnam, 1989). The National Park Service collected tree rings for *P. glauca* and *P. mariana* across a wide spatial domain in Denali at 233 sites. These records average 103.2 years long (SD = 49.7) and 2.6 cored trees (SD = 1.5) per site. Plot and tree sampling methods are described in Nicklen et al. (2016). We measured growth rings to 0.001 mm using either CooRecorder (Cybis Elektronik & Data AB), WinDENDRO (Regent Instruments Inc.), or a sliding scale. Ring widths were visually cross-dated using CDendro 8.1 (Cybis Elektronik & Data AB), and

cross-dating was validated with COFECHA 12K XP (Holmes, 1983). Basal area increment, the estimated area of wood produced by each tree in each year of growth, was calculated using ring widths and tree radius with the outside-in approach (Biondi, 1999; Johnson & Abrams, 2009). We used the sum of the ring widths as the radius unless the measured tree radius ((diameter at core/2) – average bark width) was greater than ring width sum, in which case we used the measured radius. Cores with missing outer rings were not used. We use species-level total basal area, which was calculated by summing across the basal area increment measurements by year and species at each plot. For comparison between the tree-ring and model simulation datasets, we remove small trees from UVAFME simulations and perform the same basal area calculation by summing across basal area increment within UVAFME outputs. Across 233 sites, there are only 10 sites that contain tree core data from at least one individual of each species (Figure A9: Appendix S1). We refer to these sites as the "coexistence sites" and focus our model calibration on those sites for a multi-species simulation model calibration.

Tree rings are critical for observing long-term forest succession at a particular location. However, there are many uncertainties associated with tree-ring basal area reconstructions (Alexander et al., 2018). In particular, tree-ring accuracy is reduced for earlier dates because only the trees that are alive at present were cored and do not represent trees that may have died in the past in the same location. This is known as the "fading record" because there is less information in the earlier tree rings about the forest stand (Babst et al., 2014; Nehrbass-Ahles et al., 2014). While this problem is specific to tree-ring data, we highlight it in this manuscript to show how this type of model-data mismatch can be overcome.

2.2 | Individual-based forest simulation model

University of Virginia Forest Model Enhanced (UVAFME) is a forest gap model that was originally developed as an enhancement for alpine and boreal forests on previous forest gap models (Shugart et al., 2020). The basic principle underlying UVAFME is competition for light, water, and nutrients between individual trees. Changing demographics of individual trees then feeds back to the ecosystem states to alter the availability of light, water, and nutrients in the following years. Each year, individual trees are limited by the least of these factors, resulting in realistic forest stand dynamics over successional timescales as limiting factors change over time. UVAFME has been applied in the boreal forest (Yan & Shugart, 2005) and has been further extended to simulate boreal forest dynamics across regions (Foster et al., 2019). Most notably, UVAFME has the ability to simulate permafrost change over time, which greatly affects forest stand development from individual tree to regional spatial levels in Alaska (Shur & Jorgenson, 2007). Permafrost is parameterized in UVAFME as a limiting growth factor that affects tree species differently if they are parameterized as tolerant or intolerant. The current parameterization of permafrost includes fixed rates for tolerant or intolerant growth response to active layer depth. Because growth response to

active layer depth is not well known at the individual tree level, we incorporated these rates as unknown parameters in our sensitivity analysis described in the upcoming section. There are 23 species-specific parameters in UVAFME. The default parameterization for the boreal region of Alaska is fully described in Foster et al. (2019). We used site condition settings from the plot-level data collected in Roland et al. (2019). Forest gap models have been developed to represent the expected long-term (i.e., centennial to millennial) forest community dynamics given climate normal conditions (Bugmann, 2001). Following common forest gap model implementation, we used the default climate subroutines available within UVAFME and drew random monthly temperature and precipitation from climate normal distributions (Group, 2009, 2020). Site-level settings including initial nitrogen pools were assigned based on field data collected in Roland et al. (2013) and Nicklen et al. (2019). We also follow conventional forest gap model practice by starting all of our simulations at bare ground.

2.3 | Sensitivity analysis

To fully emulate UVAFME using a Gaussian process surrogate, we considered all data inputs and data outputs to create a statistical representation of the mathematical model. UVAFME has 23 species parameters, 24 meteorological drivers per year run, and 10 site-specific conditions that are defined at the onset of the modeling process as inputs. Individual growth is incremented annually, and there are also between 10 and 100 belowground pools that are updated. For a 100-year computer model simulation, there are 2456 inputs (= (23 × 2) + (24 × 100) + 10) and between 1000 and 1,000,000 outputs depending on the number of individuals simulated. It is not pragmatic to vary each input individually to adequately represent all dimensions of the model. We must reduce the dimensions of the problem to gain meaningful ecological inference. For our purposes, we are interested in learning about parameters driving competition between spruce species, so we omit meteorological drivers and model input settings hereafter. We performed a sensitivity analysis over the species-level parameters to determine the parameters that were most important to constrain. Sensitivity analysis is a common method for reducing model dimensions by choosing parameters that are most likely to affect the output of interest (LeBauer et al., 2013).

We performed a sensitivity analysis on parameter sets for _P. mariana_ and _P. glauca_ in UVAFME. To do this, we ran eight simulations for each of the 23 parameters for each of the two species, varying one parameter at a time across an informative uniform prior centered on default values with width set based on expert knowledge of the range of realistic values from Foster et al. (2019). This resulted in 85,744 forest simulations (= 2 × 23 × 8 × 233) across 233 sites in Denali National Park. Starting at bare ground, we ran each simulation for 500 years to span the longest tree-ring record (346 years). Across each species set of simulations ($n = 42,872$), we calculated the first principal component score for each basal area time series using the "princomp" function in the R computing environment (R Core Team, 2019; Venables & Ripley,

2013). See Appendix A8.2 for pseudocode. We then determined the parameter sets that were most sensitive across sites by calculating which parameter had the largest variance across first principal component scores by parameter at each site. We chose the first principal component score as our sensitivity diagnostic because it is a simplified representation of the shape of the trajectory of basal area over time. The parameter that most changed the first principal component score variance and therefore the simulation trajectories was identified as the most sensitive for a given site–species combination.

2.4 | Model calibration with a Gaussian process surrogate

We estimated species-specific parameters for the most sensitive parameters at each coexistence site using tree-ring data to calibrate UVAFME and accurately simulate competitive dynamics between *P. glauca* and *P. mariana* over a centennial time frame. We did this with three steps listed here and described below: (1) Input grid—we simulated forest trajectories across a grid of the most important parameters determined by the sensitivity analysis using UVAFME at each site; (2) optimal coring year—we determined the best set of trajectories by calculating historical basal area trajectories for an alive tree subset each year of the simulation and using model selection to determine the optimal coring year in the simulations; (3) parameter estimation—we fit a surrogate model to the alive tree subset at the optimal coring year and estimated the best parameter combination given the tree-ring observations.

2.4.1 | Input grid

We determined the species parameter input grid for the simulations at each site using the most important site parameters from the sensitivity analysis. We varied parameters similar to the sensitivity analysis across a uniform prior centered on the species parameterizations with prior spread determined from expert knowledge from Foster et al. (2019). We selected 36 parameter combinations per site using a Latin-hypercube sampling design implemented with the "lhs" R package (Carnell & Carnell, 2016; Lin & Tang, 2015). This resulted in 36 forest basal area simulations per coexistence site. We chose 36 parameter combinations after initial attempts with fewer simulations (e.g., $5 \times 5 = 25$) where our parameter estimation process resulted in poor basal area predictions. More simulations could be added in the future, but we did not add more simulations because we met our objectives with 36 simulations at each coexistence site for this example of model calibration.

2.4.2 | Optimal coring year procedure

As discussed previously (Section 2.1), tree-ring observation uncertainty includes uncertainty from the fading record where earlier tree

rings are an inaccurate representation of forest stand basal area. Tree rings also provide the best data available for understanding long-term competitive interactions between individuals. Dendroecologists have developed sophisticated statistical models for estimating the uncertainty associated with basal area reconstructions from tree-ring records (Alexander et al., 2018; Dye et al., 2016) with the aim of incorporating these records into forest simulation models. While these methods are statistically robust and necessary for some data assimilation methodologies (Raiho et al., 2020), we used an alternative approach where we sampled the data and the model using the same methodology. As discussed in Section 2.1, we summed across the basal area increment measurement assuming that all trees were measured within a plot. In accordance with this assumption, we also thinned the model output to match the data-generating process where, at each time step, we calculated the basal area trajectory only for trees that would have been alive and large (>10 cm) when cored (i.e., the "alive tree subset"; Figure 1a). This subset varied with coring time because we assumed that the observation year, while known in calendar date, was unknown in successional time. This is similar to system proxy models used in paleoclimatology (Evans et al., 2013) and remote sensing (Jacquemoud et al., 2000) that are created to simulate an observable proxy from mathematical model output. We highlight that UVAFME has a mechanistic basis for constructing tree-ring observations from the model output where the base mechanism is annual individual diameter increment given limiting growth factors determined from the environment (see Section 2.2).

We focused on calibrating the forest simulation model using tree-ring data. To do this, we determined the optimal coring time within all model simulations according to the alive tree subsets. We consider optimal coring time as the simulation year at which the alive tree subset historical BA trajectories best match the tree-ring BA reconstructions. We fit generalized additive models (GAMs; Hastie, 2017) to the tree-ring-derived basal area data by species using UVAFME species basal area model output as a predictor and additional temporal smoothing to account for residual dependence in the data. Our predictor variable was comprised of the alive tree subset for each species resulting from the UVAFME simulation output based on the simulated coring year across the 500 simulation years. We scored these models based on predictive ability (using AIC; Akaike, 1973) to identify the optimal simulated coring year for each site. We relied on these years when making statistical inference at the next stages of our modeling procedure. This process allowed us to find the historical trajectories that most closely matched our data so that we could focus statistical power on parameter inference using the surrogate model. In a fully Bayesian setting (Berliner, 1996), the initial conditions would be determined based on model-data feedback during the model-fitting process. As discussed earlier, this is not an option in our example because of complexity in the simulation models. Another option would be to include optimal coring time in the model calibration process, but because we modularized our process as described in the next section, we illustrate a pragmatic method for determining initial conditions and leave estimating initial condition start time as a future direction.
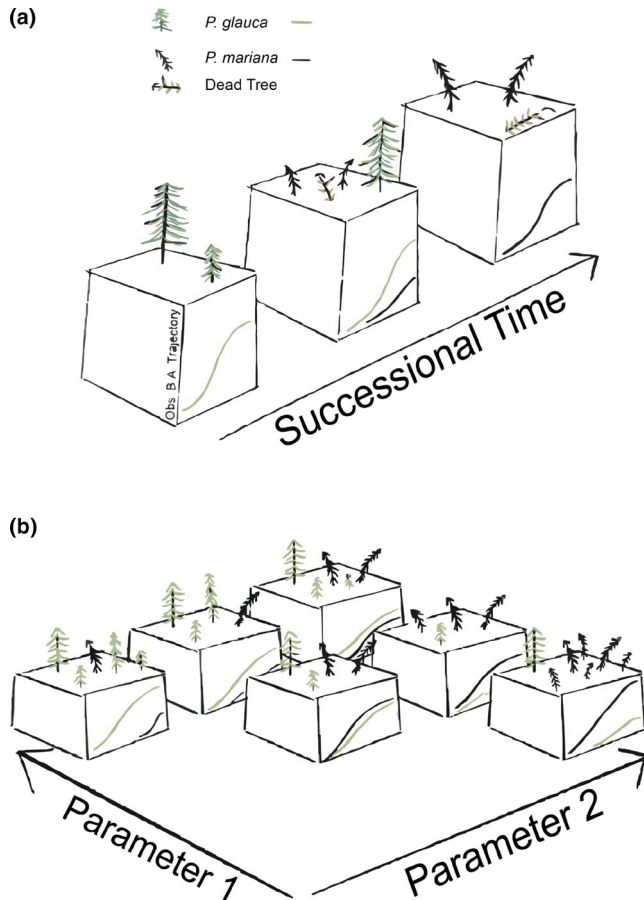
(a)

(b)

FIGURE 1 (a) Cartoon of one simulation of a forest stand at three time decadal steps $t = 1, 2,$ or 3. In this example, *Picea glauca* is dominant at $t = 1$, both species are present at $t = 2$, and *Picea mariana* is dominant at $t = 3$. Observed basal area trajectories of the alive tree subset for each species (black for *P. mariana* and green for *P. glauca*) are shown on the side of each cube and illustrate how the basal area trajectory observations may change depending on when the trees were cored during successional time. (b) Consistently labeled with cartoon (a). Each cube shows the alive tree subset basal area reconstructions for a given simulated forest stand. Forest stands vary with parameter sets used in the computer experiment

2.4.3 | Parameter estimation

Our aim is to provide an ecological example of a commonly used surrogate modeling (Bayarri et al., 2009; Higdon et al., 2008; Tuo & Jeff Wu, 2016; Tuo & Wu, 2015) known as the "Kennedy and O'Hagan" framework (Kennedy & O'Hagan, 2001; Liu et al., 2009), building on the surrogate modeling strategy Gramacy (2020). The basic premise uses a Gaussian process model (also known as "kriging"), which focuses the majority of statistical learning on the covariance matrix such that

$$Y \sim \text{MVN}(0, \Sigma) \quad (1)$$

$$\Sigma = \exp\left\{ -\sum_{k=1}^{p} \frac{(x_k - x'_k)^2}{\theta} \right\} + \nu I_{x=x'} \quad (2)$$

where the mean of the response variable (Y) is assumed to be 0, and the covariance is a function of the distances between the inputs with range (θ) controlling smoothness and nugget (ν) controlling the size of the discontinuity. While our modeling experiments are deterministic, see Lee et al. (2011) for nugget recommendations in surrogate modeling experiments. We use a Gaussian process in two different parts of the model calibration process described below.

At each of the 10 coexistence sites, we estimated two UVAFME model parameters, one for each species. This algorithm assumes the forest model simulations ($y^M(\cdot, \cdot)$) are deterministic and underlie the observations in the field data (Y^F) with an estimated bias ($b(\cdot)$) and process variance (ϵ) such that,

$$Y^F = y^M(\cdot, u^*) + b(\cdot) + \epsilon, \quad (3)$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$, and u^* is an input parameter value. We repeated the following steps over a full grid (i.e., 20×20) of possible parameter values at each coexistence site. We chose a 20×20 grid to increase the simulation parameter grid resolution ($N = 36$ to $N = 400$) and demonstrated the ability of this approach to estimate parameters with only a small subset of simulations (i.e., 9% of full grid). First, we built the surrogate of the forest simulation model ($\hat{y}^M$) by fitting a Gaussian process to the input (X_{n_M}, u) and output (Y_{n_M}) of the model experiments ($n_M = 1, \dots, N, N = 36$). Specifically, we fit a Gaussian process model with heteroskedastic noise to represent the increasing variability over time in the forest model simulations using the Matern covariance structure (Binois et al., 2018). Second, we made a prediction ($\hat{Y}_{n_F}^{M|u}$) from the surrogate given a set of parameters (u^*). Third, we calculated the residuals between these predictions and the tree-ring data (i.e., $\hat{Y}_{n_F}^{M|u} - Y^F$). Fourth, we fit another Gaussian process model to these residuals to estimate bias ($\hat{b}$) to calculate the negative log likelihood of the parameters (u^*) given the bias. In this step, we were simply optimizing to minimize bias. Finally, across the 20×20 parameter grid, we determined which parameter set has the lowest negative log likelihood. From this procedure, we obtained a 400 pixel parameter likelihood surface, the most likely parameter combination ($\hat{u}$) according to the data, and the data-estimated model bias correction ($\hat{b}$). With the estimated bias correction, we also obtained a bias square error estimate from the residual prediction and were able to show a bias-corrected basal area prediction by adding the prediction from the original surrogate to the bias. We repeated the steps across all coexistence sites to obtain optimal parameter sets for the most important species parameters determined from the sensitivity analysis. We specifically chose to use a maximum-likelihood approach instead of a Bayesian approach to provide an ecological example of a thrifty model calibration using a surrogate model. In using maximum likelihood, we reduced our computational time 10-fold, because these calculations take only a few minutes or seconds on a personal computer. We implemented the statistical surrogate model using the "laGP" package version 1.5.5 and the "hetGP" package version 1.1.3 in the R computing environment version 3.6.2 (Binois & Gramacy, 2019; Gramacy, 2016; R Core Team, 2019b). We also provide pseudocode in Appendix A8.2.

3 | RESULTS

3.1 | Sensitive parameters

Parameters governing life expectancy (average maximum age (years), Max. Age) and species climate suitability (minimum growing degree-days for growth; DD Min) dominated the sensitivity analysis of *P. glauca*. Max. Age is used in many of the allometric equations and is used to calculate an annual background mortality rate. If the growing degree-days for a year (cumulative sum of degrees C above 5°C) are below DD Min for a species, the species DBH increment growth is 0 for the year. The parameter determining the minimum diameter increment before a tree is considered "stressed" and marked for potential stress-induced mortality (DBH Min) dominated the sensitivity analysis for *P. mariana* (Figure 2). Trees have a chance (21%–33%) of dying if they have below-minimum growth for at least 2 years. Across the 233 sites, the most sensitive parameters for *P. glauca* and *P. mariana* were life expectancy (Max. Age, 23% of sites) and minimum annual increment for diameter at breast height (DBH Min, 77% of sites). The second most sensitive parameters for *P. glauca* and *P. mariana* were growing degree-days minimum required for growth (DD Min, 17% of sites) and leaf mass per unit area (LC, 10% of sites). Both sensitivity analyses showed spatial patterns from lowland (northwest) to higher elevation sites (southeast) where both species were sensitive to minimum growing degree-days at higher elevations. The coexistence sites (Figure 2, open circles) showed somewhat different patterns for *P. glauca* with six sites having the maximum growth parameter (*g*) as the most sensitive parameter. The maximum growth parameter (*g*) determines the speed of initial DBH increment growth. Thus, faster-growing species are typically parameterized to have a higher *g*. Over all 233 sites, *g* was the most sensitive parameter for only 14% of sites for *P. glauca*. Sensitive parameters for the coexistence sites for *P. mariana* showed similar patterns to the overall dataset with seven of the 10 sites having minimum annual increment for diameter at breast height (DBH Min) as the most sensitive parameter. Of the 6 coexistence sites where *g* is the most sensitive for *P. glauca*, 4 of them also had DBH Min as the most sensitive parameter for *P. mariana*.

3.2 | Optimal coring times

The majority of forest simulation site–year combinations at the coexistence sites showed *P. mariana* dominance over *P. glauca* (95.4%). This was caused by a combination of poor recruitment by *P. glauca*, nitrogen limitation (Figures A10 and A11: Appendix S1), and maximum age restrictions. However, *P. glauca* persisted in most model simulations between 0 and 300 years (Figure A12: Appendix S1). This pattern in the model output led to similar estimates of optimal coring years across sites (Figure 3 vertical lines, mean = 161.85 simulation year, SD = 96.3 simulation year) because, earlier in the simulations, both *P. mariana* and *P. glauca* were present in the model output. For example in Figure 4 left, at site E. CHITS16, *P. mariana*

(blue) was dominant from year 0 to year ~120; then, *P. glauca* (yellow) became dominant from ~120 to ~280 on average across the 36 simulations for this site. During the optimal coring year analysis for E. CHITS16, we found that year 113 was the optimal coring year for model calibration. Figure 4 right top shows the alive tree subset for the optimal coring year contrasted with Figure 4 right bottom, which shows a suboptimal alive tree subset. Similar graphics for determining the optimal alive tree subset across sites is shown in Figure A12: Appendix S1.

Using this process, we were able to estimate stand age. We found that stand age was related to chronology age, but as chronology age increased, the estimated stand age diverged from the chronology age (Figure 5, dashed line). This suggests that our approach may be able to account for bias in large tree sampling to improve estimates of stand age by reducing temporal bias (Gutsell & Johnson, 2002; Speer, 2010). However, this is a preliminary finding and more sites may be necessary to quantify the shape and magnitude of this bias.

3.3 | Parameter estimates

We estimated parameter likelihood surfaces (11× the resolution of the original simulations) at 10 sites for the two species most sensitive parameters using a GP surrogate for UVAFME (Figure 6). For the most sensitive parameters at the coexistence sites, neither resulted in parameter estimates near the default parameterization. We found that *P. glauca* growth parameter (*g*) was estimated to be high at the majority of the coexistence sites (Figure 7 upper left). We assessed the effect of *g* for *P. glauca* on the final basal model output area across the coexistence site simulations and found that increasing *g* values lead to increased *P. glauca* basal area in the simulations (Figure 7 bottom left). Higher estimates of *g* correspond to observations of tree-ring data (Figure A9: Appendix S1, gray) that typically show greater *P. glauca* basal area than represented in the model simulations. Across the six coexistence sites where DBH Min was the most important parameter for *P. mariana*, none were estimated near the default value. Similar to *g* in *P. glauca*, DBH Min for *P. mariana* was typically estimated to be larger (Figure 7 upper right). However, the relationship between DBH Min and simulated *P. mariana* basal area is weaker than the relationship between *g* and *P. glauca* basal area (Figure 7 bottom right versus bottom left). The likelihood surfaces in parameter space (Figure 6) were multi-modal and demonstrate that, while it is important to reduce dimensionality in the model-fitting process, it may also be important to include additional parameters for these sites in the future.

After determining the most likely model parameterization according to the data, we were able to adjust model predictions from the GP surrogate given information from the tree-ring data (Figure 8). In some cases (Figure 8, top two rows), the optimal parameterization led to GP predictions that closely matched the data with low model prediction uncertainty, while in other cases (Figure 8, third and fifth rows), the predictions suggest that further model calibration may be necessary because there were both poor prediction using
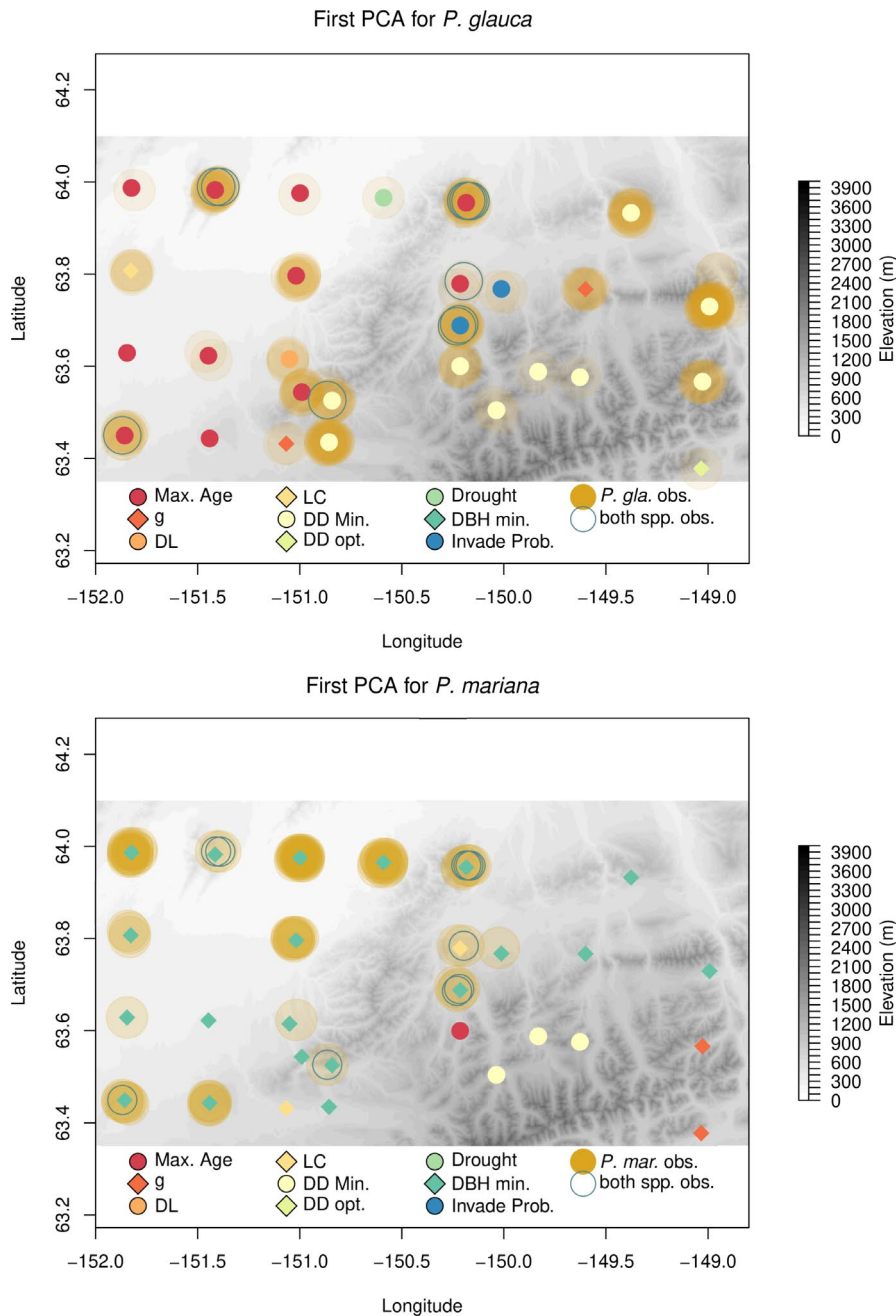
First PCA for *P. glauca*

First PCA for *P. mariana*

FIGURE 2 Elevation map of northern Denali National Park. Colored points show the most sensitive species parameter from the forest simulation model for each site cluster for *Picea glauca* (top) and *Picea mariana*. Tree-ring observations are indicated behind the colored points in transparent yellow where increasing saturation indicates more observations of a particular species. Circles indicate the 10 coexistence site locations where both spruce species were observed when tree rings were collected. The most sensitive species parameters are as follows: life expectancy (Max. Age), growth parameter (*g*), scalar parameter for leaf area to diameter relationship (DL), leaf mass per unit area (LC), growing degree-days minimum (DD Min), growing degree-days optimum (DD opt.), drought tolerance (drought), diameter at breast height minimum growth increment (DBH Min), and probability of seed dispersal into the plot from outside the plot (Invade Prob.)

the optimal parameters and wide uncertainty estimates. In general, we found that model predictions were less similar to the data for *P. glauca* than for *P. mariana* (Figure 8, white versus gray) aligning with our results from the optimal coring years where coexistence of *P. glauca* and *P. mariana* was somewhat uncommon in our simulations because *P. glauca* was suppressed by poor recruitment and nutrient limitations.

4 | DISCUSSION

DNPP faces climate changes that have been predicted to dramatically alter species composition and distribution (Chapin et al., 2004). *Picea glauca* is currently the most abundant tree species and may

expand into *P. mariana* sites in the near future (Nicklen et al., 2021). Wirth et al. (2008) hypothesized that this expansion may cause a range contraction for *P. mariana* because *P. glauca* grows more quickly and may outcompete slower-growing *P. mariana* during early post-fire succession. However, *P. mariana* is also more tolerant of wet edaphic conditions and may itself expand into new sites (Roland et al., 2013). Currently, it is difficult to observe the competitive dynamics between *P. mariana* and *P. glauca* because they do not typically co-occur (Nicklen et al., 2021). We used long-term data to constrain the growth between the two species at sites where they do currently co-occur within DNPP to better understand how they may affect each other in the future. Our model calibration results show that *P. glauca* may grow faster and *P. mariana* may become growth-stressed sooner than current UVAFME default parameterizations (Figure 7) agreeing
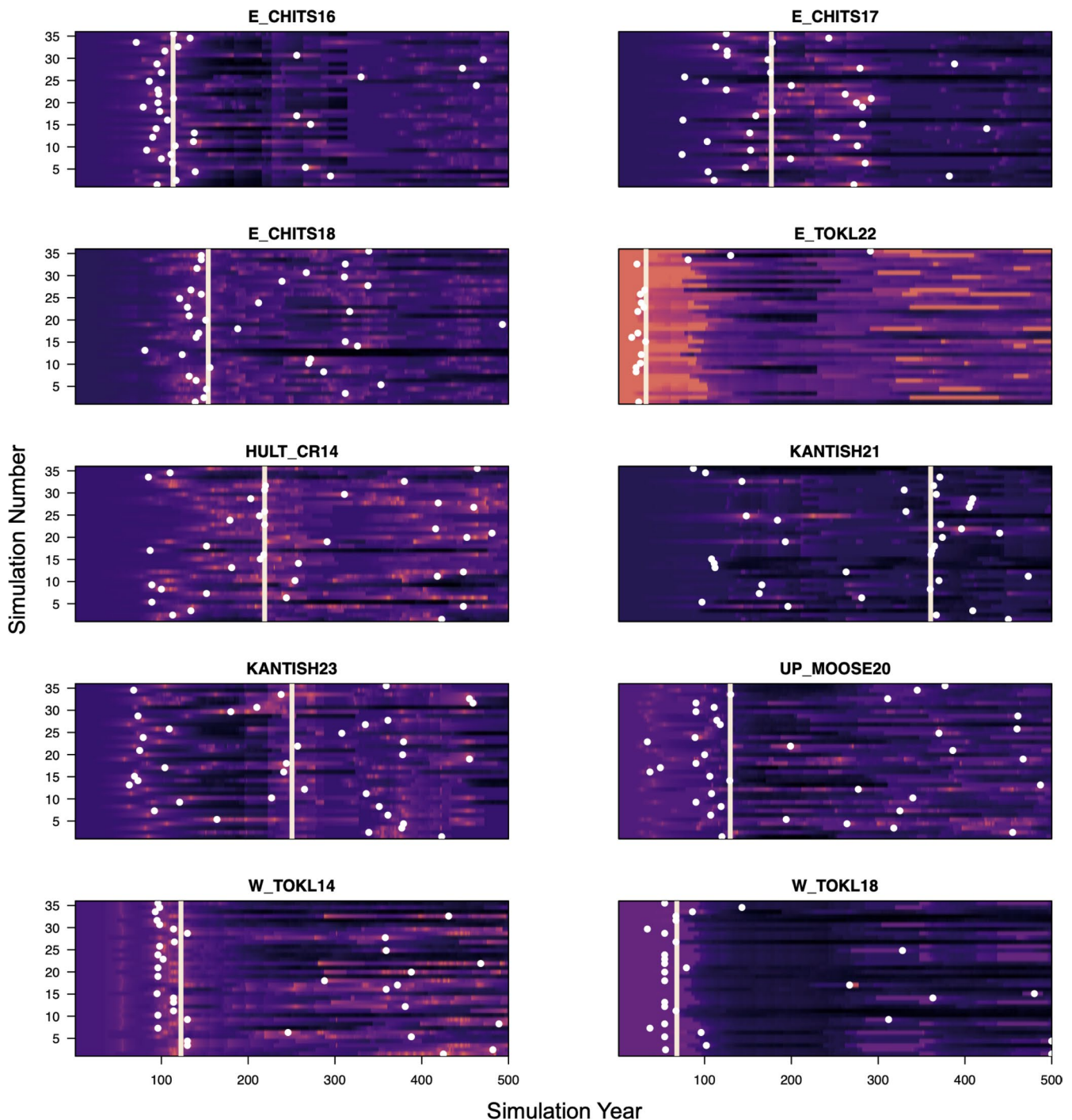
FIGURE 3 Optimal coring time results at the coexistence sites. White points represent the lowest AIC values for each simulation at a given site. Vertical white lines represent the median of the white points also called the optimal coring time. Plots are colored by AIC values with lighter colors representing lower AIC values. AIC color palette is relative to each site

with assessments that *P. glauca* may have a competitive advantage over *P. mariana*. Many sites were estimated to be older than tree rings alone would predict (Figure 5). Similarly, across all 233 sites, the maximum age parameter (i.e., life expectancy) of *P. glauca* was the most sensitive model parameter. Emphasis on maximum age contradicts the assumption that early succession is the most vulnerable time for *P. mariana* in competition with *P. glauca* and suggests that understanding which environmental factors lead to long-lived

P. mariana may help predict future *P. mariana* distribution. Many experiments have been conducted that could inform these growth and longevity parameters under a variety of growing condition scenarios across sites (e.g., Johnstone et al., 2010).

Improving parameter representations may lead to better predictions from forest gap models and may also help ecologists pinpoint under-studied mechanisms such as long-term growth rate trade-offs between species. For instance, *g* has not been significantly altered
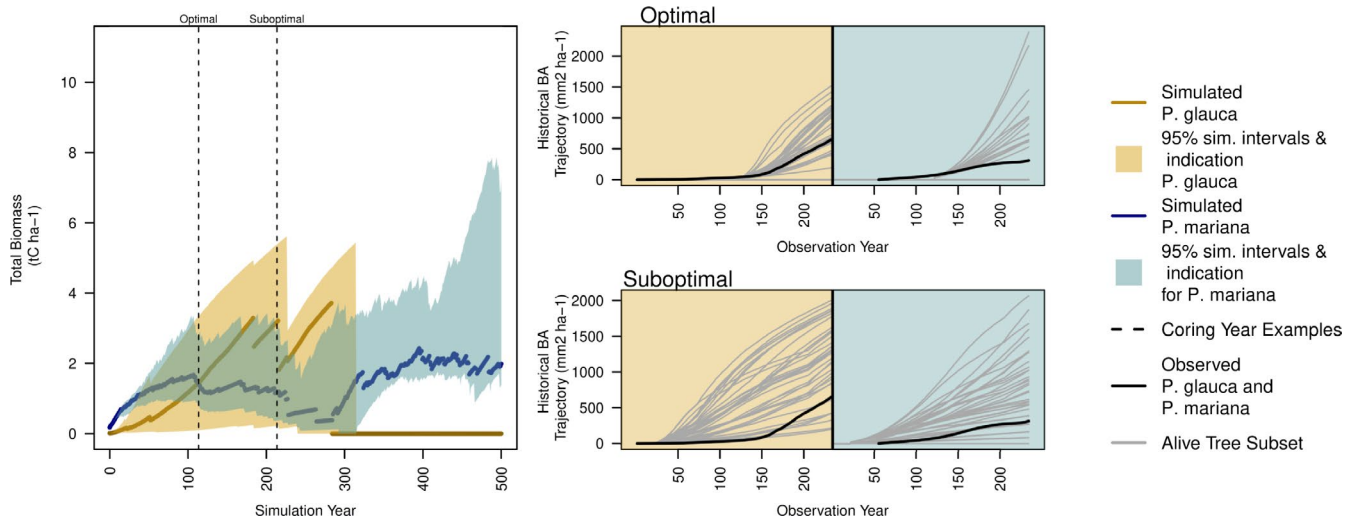
FIGURE 4 Left: Biomass trajectories over a 500-year simulation for *Picea glauca* (yellow) and *Picea mariana* (blue). Line represents median across biomass trajectories, and shading represents 95% quantiles. Vertical lines represent optimal coring year chosen by the model selection procedure and example of suboptimal coring year. Right: Alive tree subsets (gray) overlaid with field-collected tree-ring data (black). The top plot represents the optimal alive tree subset collected from model output in year 113 of the model simulation, while the bottom plot represents a suboptimal example of the alive tree subsets collected from the model output in year 213
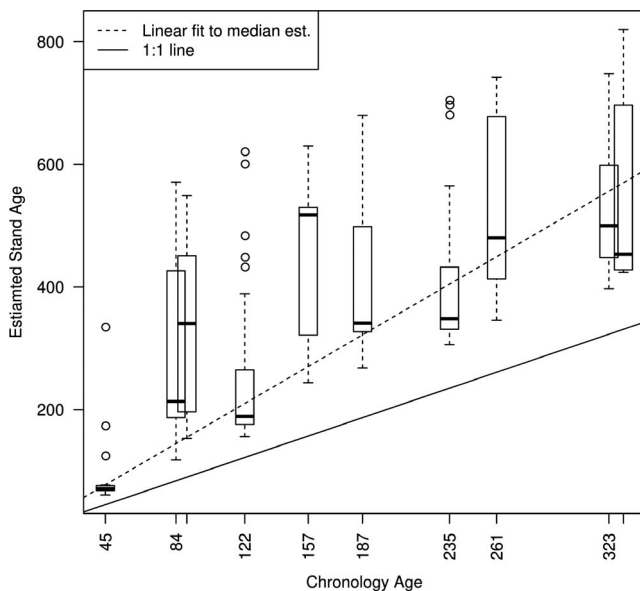

FIGURE 5 Chronology age determined by the number of tree rings collected at a site by estimated stand age determined by the optimal coring year selection process. Box plots represent uncertainty in estimated stand age between 36 simulations performed at each site. Solid black line indicates an assumption that chronology age and estimated stand age are synonymous, and dashed line represents the possible increasing bias in that assumption as chronology age increases

in UVAFME since its implementation in Botkin et al. (1972) and may have too strong of an influence over individual tree growth. Simulating coexistence between tree species is not specific to this model but is a difficult problem in global vegetation modeling in general because trade-offs between species or plant functional types over time often result in one vegetation type dominating the

simulation (Arora & Boer, 2006; Gravel et al., 2011; Turnbull et al., 2013). However, over the process of forest succession, species have opportunities to take advantage of different available resources (Falster et al., 2017). These successional trade-offs are mechanistic and have been proposed as a path forward for improving global vegetation representation and predictions (Fisher & Koven, 2020), but the first step may be to understand which parameters are driving long-term coexistence between species. While our example is not comprehensive, our approach provides a pragmatic path forward for improving multi-species representations in vegetation simulation models using long-term data.

Altering the model output allowed us to efficiently estimate model parameters while confronting uncertainties associated with the data-generating process, in particular, the fading record (Babst et al., 2014; Nehrbass-Ahles et al., 2014). In ecology, field observations are often proxies for the true quantity of interest. To accurately represent the true quantity of interest, ecologists have two options: alter the observation model to account for measurement error or alter the model output to match the observation. Both of these options involve modeling a translation between the field observation and the quantity of interest, which can be referred to generally as the "data model." In our tree-ring example, much effort has been put into constructing data models that translate tree-ring widths into basal area or biomass "data products" to match forest simulation output (Alexander et al., 2018; Dye et al., 2016). Simultaneously, system proxy models in paleoclimatology or inverse transfer models in remote sensing have been developed to translate model output into observational proxies (Evans et al., 2013; Jacquemoud et al., 2000). We adopt this latter strategy and translate the individual-level model output into species-level basal area data to match the tree-ring observations (Figures 1 and 4). This approach leverages the ecological mechanisms in UVAFME where each individual tree
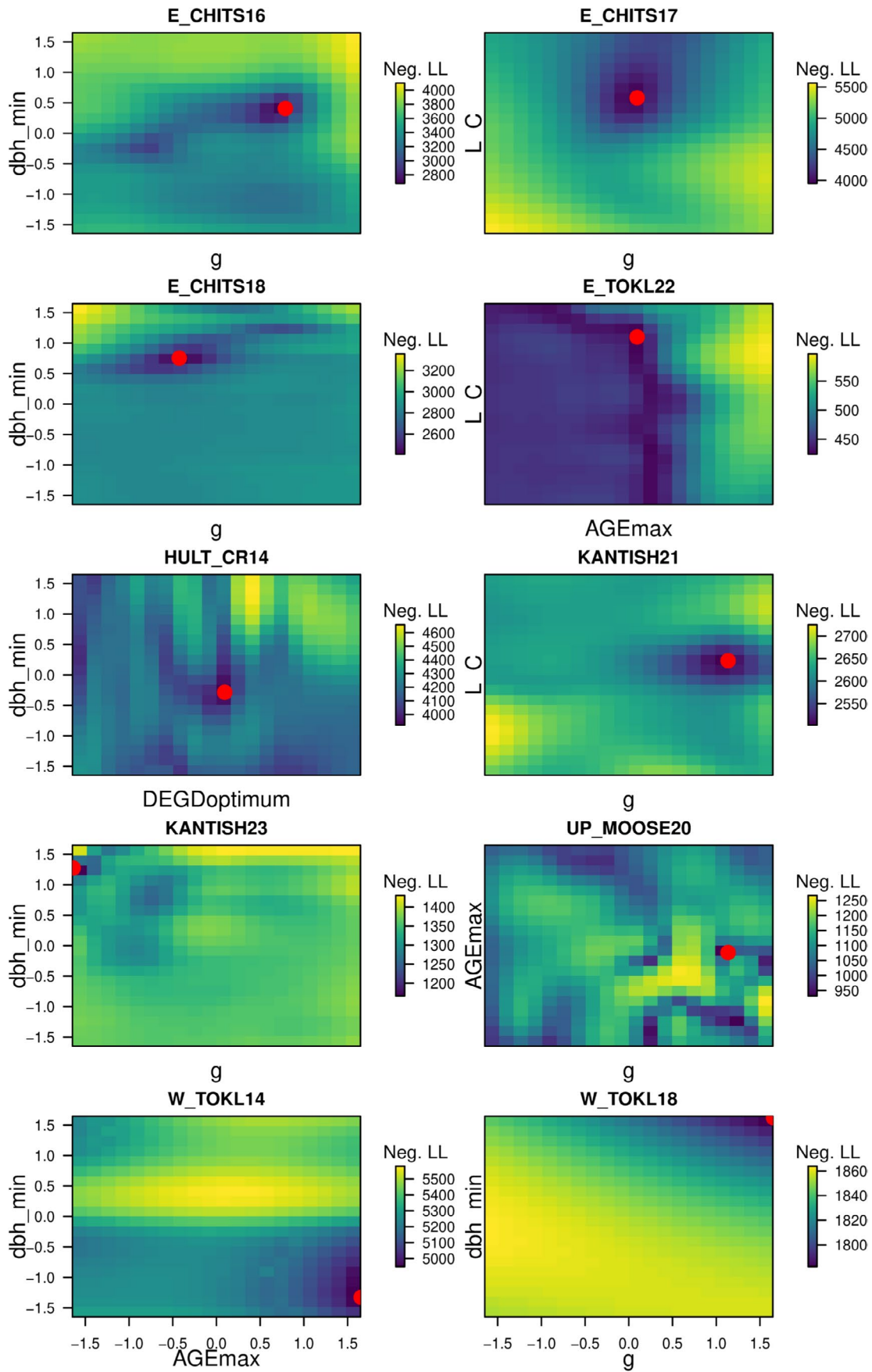
FIGURE 6 Negative log-likelihood surfaces between the two most sensitive parameters for each site. Red points indicate lowest negative log-likelihood value on each surface. Each figure is centered on the default parameter values
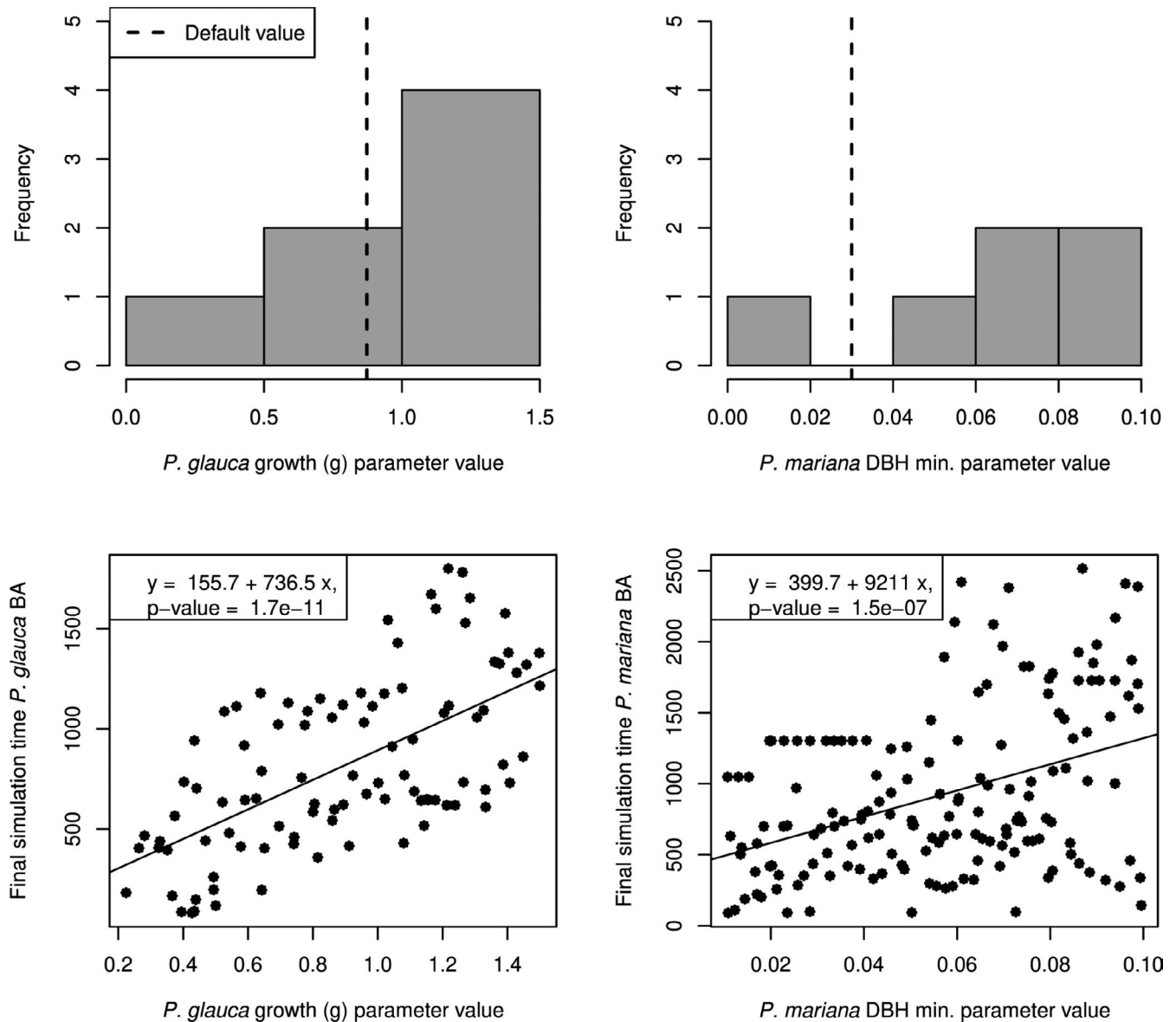
FIGURE 7 Top: Histograms of parameter estimates across coexistence sites from modularized Kennedy and O'Hagan algorithm (Liu et al., 2009). Default parameterization from Foster et al. (2019) shown as a vertical dashed line. Bottom: Relationship between chosen species parameter values and final basal area across simulations at coexistence sites where each point represents a single simulation. Diagonal lines represent fitted linear relationship with coefficient estimates shown in the top-left boxes. Bottom left: The maximum growth parameter (*g*) affects the speed of initial DBH increment growth where higher values of *g* suggest faster-growing trees. Bottom right: DBH Min determines the minimum diameter increment before a tree is considered "stressed" and flagged for potential stress-induced mortality where higher values of DBH Min indicate lower stress tolerance

grows annual tree-ring increments and allowed us to more easily determine which simulations to use for model calibration.

Our approach relies on simulations from a computer experiment spanning a range of output covering what was observed in the data. While our approach for estimating optimal coring time was pragmatic, sometimes (e.g., Figure 4 top right, *P. mariana*), the optimal alive tree subset did not fully overlap the historical trajectories from the data. The GP approach allowed us to interpolate these missing components of the simulations in most cases, but at some of the coexistence sites (e.g., UP. MOOSE 20), we may need to include more parameters when building the GP to fully calibrate UVAFME to

simulate the dynamics between *P. glauca* and *P. mariana*. Iterative approaches have been suggested for vegetation model calibration (Fer et al., 2018; Hartig et al., 2011). A related procedure for performing optimal design of the computer experiment to identify the next set of simulations could be incorporated to improve the GP surrogate for a particular site (Williams et al., 2018). This approach has been useful in engineering (Ju et al., 2018) but remains to be tested in an ecological setting where model output and field data observations may be mismatched.

Like field data, most mechanistic model output does not represent the true process of interest. However, unlike field data,
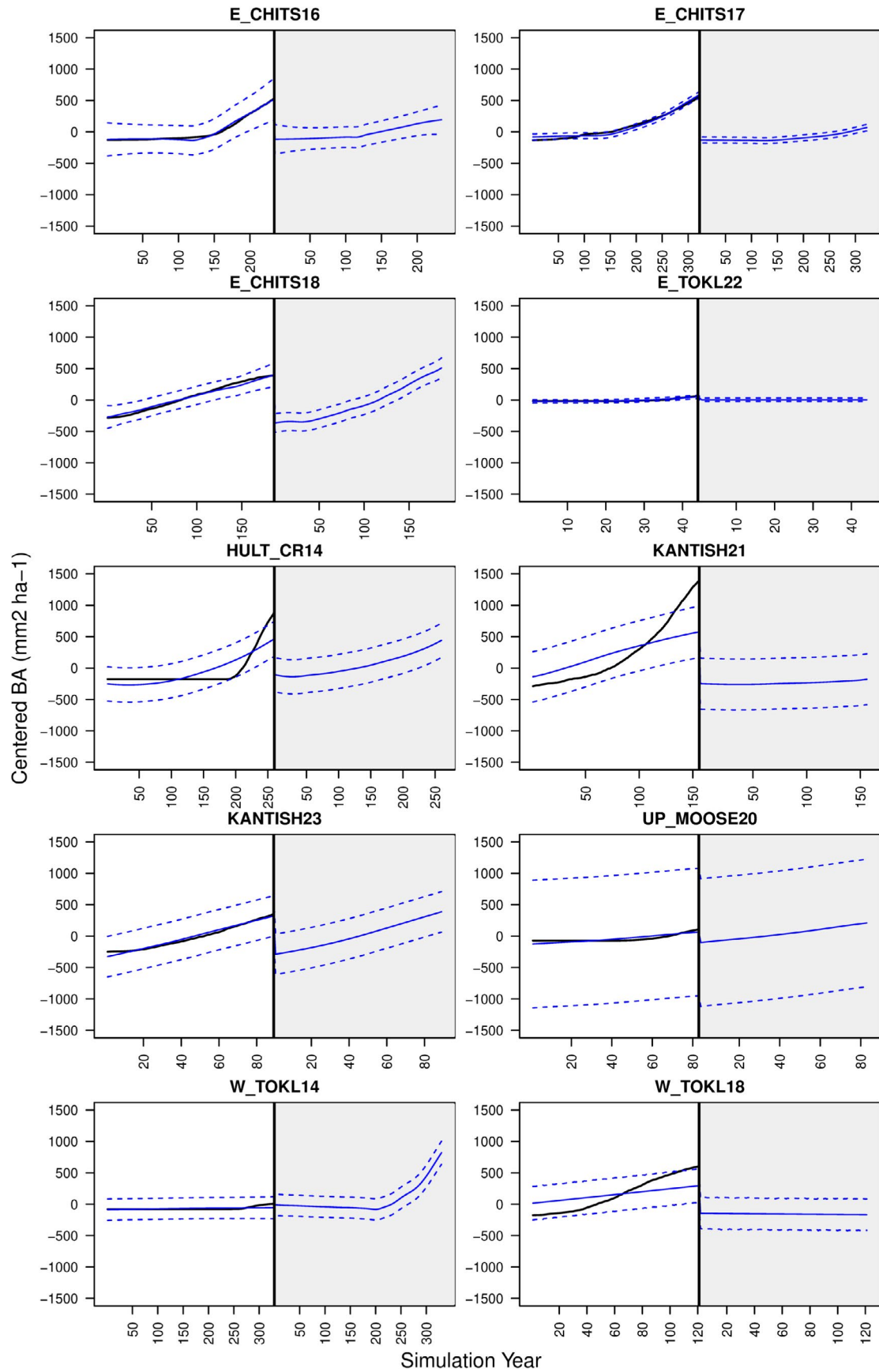
FIGURE 8 Adjusted model predictions (blue) via model GP surrogate with optimal parameters overlaid with data (black). Solid blue line is the median prediction, and dashed lines are 95% confidence intervals. Gray shading indicates *Picea mariana*, while no shading indicates *Picea glauca*

simulation models have the ability to represent many dimensions of the observable process, including missing data or emergent phenomena. This can make it difficult to determine which model subspace the data represent. In our example, the subspace of the modeled process that was difficult to determine was the coring sample date, which relies on knowing when the data were collected during the forest stand development process. Important influence of initialization and/or boundary conditions on prediction is a common property in ecological systems (Hastings, 2001). The general strategy in statistical models is to estimate the start or end time based on the data. Yet, embedding a mathematical model into a statistical model is not always pragmatic. A common method that simulation modelers use to confront this problem is to assume that the start time ($t = 1$) is known or that the start time begins after an equilibrium state (Carvalhais et al., 2008; Elzein et al., 2020). For forest stand development, this assumption is met by initializing the model at bare ground and running the model for many years before analyzing the model output. Our approach shows that this assumption may be negatively biasing the stand ages in our example (Figure 5) and that the stand initialization time can be optimized using the data. A future direction of our approach is to include the optimal coring data within the model calibration procedure itself, but this approach would sacrifice computational time in the estimation process and may not yield different results. Many studies have found that forests are sensitive to initial conditions (Raiho et al., 2020; Temperli et al., 2013). Our approach provides a way to estimate stand initiation with tree-ring data and forest simulation models together to leverage the long-term observations while accounting for missing processes with a mechanistic model for forest stand development.

All simulations for this study were performed on a portable computer. This is atypical of an individual-based model where computational times typically force the modeler to use distributed computing resources (Hooten et al., 2020). For the non-specialist, this requirement may be prohibitive or discouraging, but it is possible to use our approach without performing the model simulations oneself. A surrogate GP may be fit to any inputs and outputs of a computer experiment and used to create an emulator for the computer model. Ecosystem model developers invest large amounts of time learning new coding languages, researching the history of simulation model development, and investing in understanding model intricacies. While model developers are extremely important to the continued improvement of the mechanisms within the simulation models, many research teams have pointed to the need for more multi-model comparisons (Renwick et al., 2018; Thomson et al., 2006) and more sophisticated assimilation of field data (Fer et al., 2021). This type of work can be completed by a statistician or quantitative ecologist who may not have an interest in the development of a particular model code base. Beginning ecologists who have a specific interest in constraining many different types of simulation models with field data may find that exploring GP modeling with preexisting inputs and outputs available on public archiving sites is a more efficient entry point. This strategy may help create a larger workforce for fitting ecological models to field data by more efficiently introducing students to data assimilation without requiring them to perform the model simulations themselves.

Bringing computer simulation models together with field data has great potential for improving ecological inference and forecasting for a variety of applications (Christin et al., 2019; Dietze et al., 2018; Luo et al., 2011; Peters et al., 2014). In interior Alaska, rapidly improving understanding and quantitative predictions of forest distributions is necessary for ecological management under changing climate. Ecologists can increase the pace of model improvements by continuing to build on a growing suite of existing model calibration tools (Fer et al., 2018; Oberpriller et al., 2021; Pietzsch et al., 2020; Speich et al., 2021; Tao et al., 2020) and team members by trying new methods in different contexts. We provided the first example of a forest model calibration using tree-ring data allowing us to demonstrate three pragmatic approaches to proceed with model calibration: connecting model outputs to data-generating processes, determining data-driven starting conditions from a suite of model simulations, and reducing a high-dimensional model calibration problem _a priori_ to model fitting for computational efficiency. These approaches are a subset of a growing toolbox for building robust connections between computer models and field data in ecology.

AUTHOR CONTRIBUTIONS

Ann M. Raiho: Conceptualization (lead); formal analysis (lead); methodology (lead); visualization (lead); writing—original draft (lead); writing—review and editing (lead). **E. Fleur Nicklen:** Conceptualization (supporting); data curation (lead); formal analysis (supporting); visualization (supporting); writing—review and editing (equal). **Adrianna Foster:** Conceptualization (supporting); formal analysis (supporting); methodology (supporting); software (lead); writing—review and editing (equal). **Carl A. Roland:** Conceptualization (supporting); data curation (equal); funding acquisition (lead); project administration (lead); supervision (equal); writing—review and editing (equal). **Mevin B. Hooten:** Conceptualization (equal); formal analysis (supporting); funding acquisition (supporting); methodology (supporting); supervision (lead); writing—review and editing (equal).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ann M. Raiho 🔾 https://orcid.org/0000-0002-2552-3399
Mevin B. Hooten 🔾 https://orcid.org/0000-0002-1614-723X

REFERENCES

ACIA, E. (2005). *Arctic climate impact assessment*. ACIA Overview report.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60, 255–265. https://doi.org/10.1093/biomet/60.2.255

Alexander, M. R., Rollinson, C. R., Babst, F., Trouet, V., & Moore, D. J. (2018). Relative influences of multiple sources of uncertainty on cumulative and incremental tree-ring-derived aboveground biomass estimates. *Trees*, 32, 265–276. https://doi.org/10.1007/s00468-017-1629-0

Arora, V. K., & Boer, G. J. (2006). Simulating competition and coexistence between plant functional types in a dynamic vegetation model. *Earth Interactions*, 10, 1–30. https://doi.org/10.1175/EI170.1

Babst, F., Alexander, M. R., Szejner, P., Bouriaud, O., Klesse, S., Roden, J., Ciais, P., Poulter, B., Frank, D., Moore, D. J. P., & Trouet, V. (2014). A tree-ring perspective on the terrestrial carbon cycle. *Oecologia*, 176, 307–322. https://doi.org/10.1007/s00442-014-3031-6

Banks, D. L., & Hooten, M. B. (2021). Statistical challenges in agent-based modeling. *The American Statistician*, 75(3), 235–242. https://doi.org/10.1080/00031305.2021.1900914

Bayarri, M., Berger, J., & Liu, F. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4, 119–150.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In K. M. Hanson & R. N. Silver (Eds.), *Maximum entropy and Bayesian methods*. Maximum entropy and Bayesian methods. Fundamental theories of physics (An international book series on the fundamental theories of physics: Their clarification, development and application), (vol. 79). Springer. https://doi.org/10.1007/978-94-011-5430-7_3

Bijak, J., Hilton, J., Silverman, E., & Cao, V. D. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the United Kingdom with Gaussian process emulators. *Demographic Research*, 29, 729–766. https://doi.org/10.4054/DemRes.2013.29.27

Binois, M., & Gramacy, R. (2019). hetgp: Heteroskedastic Gaussian process modeling and sequential design in R. *Journal of Statistical Software*, 98, 1–44.

Binois, M., Gramacy, R. B., & Ludkovski, M. (2018). Practical heteroscedastic Gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27, 808–821. https://doi.org/10.1080/10618600.2018.1458625

Biondi, F. (1999). Comparing tree-ring chronologies and repeated timber inventories as forest monitoring tools. *Ecological Applications*, 9, 216–227.

Botkin, D. B., Janak, J. F., & Wallis, J. R. (1972). Some ecological consequences of a computer model of forest growth. *The Journal of Ecology*, 1, 849–872. https://doi.org/10.2307/2258570

Bugmann, H. (2001). A review of forest gap models. *Climatic Change*, 51, 259–305.

Carnell, R., & Carnell, M. R. (2016). *Package 'lhs'*. CRAN. https://cran.rproject.org/web/packages/lhs/lhs.pdf

Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G. J., Pereira, J. S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., & Valentini, R. (2008). Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. *Global Biogeochemical Cycles*, 22(2), 1–16. https://doi.org/10.1029/2007GB003033

Chapin, F. S., 3rd, Peterson, G., Berkes, F., Callaghan, T. V., Angelstam, P., Apps, M., Beier, C., Bergeron, Y., Crépin, A.-S., Danell, K., Elmqvist, T., Folke, C., Forbes, B., Fresco, N., Juday, G., Niemelä, J., Shvidenko, A., & Whiteman, G. (2004). Resilience and vulnerability of northern regions to social and environmental change. *Ambio*, 33, 344–349. https://doi.org/10.1579/0044-7447-33.6.344

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10, 1632–1644. https://doi.org/10.1111/2041-210X.13256

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., Larsen, L. G., Loescher, H. W., Lunch, C. K., Pijanowski, B. C., Randerson, J. T., Read, E. K., Tredennick, A. T., Vargas, R., Weathers, K. C., & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 1424–1432. https://doi.org/10.1073/pnas.1710231115

Dye, A., Barker Plotkin, A., Bishop, D., Pederson, N., Poulter, B., & Hessl, A. (2016). Comparing tree-ring and permanent plot estimates of aboveground net primary production in three eastern US forests. *Ecosphere*, 7, e01454. https://doi.org/10.1002/ecs2.1454

Elzein, T., Larocque, G. R., Sirois, L., & Arseneault, D. (2020). Comparing the predictions of gap model with vegetation and disturbance data in south-eastern Canadian mixed forests. *Forest Ecology and Management*, 455, 117649. https://doi.org/10.1016/j.foreco.2019.117649

Erickson, C. B., Ankenman, B. E., & Sanchez, S. M. (2018). Comparison of Gaussian process modeling software. *European Journal of Operational Research*, 266, 179–192. https://doi.org/10.1016/j.ejor.2017.10.002

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., & Anchukaitis, K. J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, 76, 16–28. https://doi.org/10.1016/j.quascirev.2013.05.024

Falster, D. S., Brännström, Å., Westoby, M., & Dieckmann, U. (2017). Multitrait successional forest dynamics enable diverse competitive coexistence. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E2719–E2728. https://doi.org/10.1073/pnas.1610206114

Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., Desai, A., Duveneck, M. J., Fisher, J. B., Haynes, K. D., Hoffman, F. M., Johnston, M. R., Kooper, R., LeBauer, D. S., Mantooth, J., Parton, W. J., Poulter, B., Quaife, T., Raiho, A., … Dietze, M. C. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, 27, 13–26. https://doi.org/10.1111/gcb.15409

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, 15, 5801–5830. https://doi.org/10.5194/bg-15-5801-2018

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, 12, e2018MS001453. https://doi.org/10.1029/2018MS001453

Foster, A. C., Armstrong, A. H., Shuman, J. K., Shugart, H. H., Rogers, B. M., Mack, M. C., Goetz, S. J., & Ranson, K. J. (2019). Importance of tree-and species-level interactions with wildfire, climate, and soils in interior Alaska: Implications for forest change under a warming climate. *Ecological Modelling*, 409, 108765.

Fritts, H. C., & Swetnam, T. W. (1989). Dendroecology: A tool for evaluating variations in past and present forest environments. *Advances in Ecological Research*, 19, 111–188.

Gramacy, R. B. (2016). laGP: Large-scale spatial modeling via local approximate Gaussian processes in R. *Journal of Statistical Software*, 72, 1–46.

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC Press.

Gravel, D., Guichard, F., & Hochberg, M. E. (2011). Species coexistence in a variable world. *Ecology Letters*, 14, 828–839. https://doi.org/10.1111/j.1461-0248.2011.01643.x

Group, P. C (2009). *Mean monthly temperature for Alaska 1971–2000; annual mean average temperature for Alaska 1971–2000*. Oregon State University. http://prism.oregonstate.edu

Group, P. C. (2020). *Alaska*. http://prism.oregonstate.edu/projects/alaska.php

Gutsell, S., & Johnson, E. (2002). Accurately ageing trees and examining their height-growth rates: Implications for interpreting forest dynamics. *Journal of Ecology*, *90*, 153–166. https://doi.org/10.1046/j.0022-0477.2001.00646.x

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models–theory and application. *Ecology Letters*, *14*, 816–827. https://doi.org/10.1111/j.1461-0248.2011.01640.x

Hartig, F., Minunno, F., & Paul, S. (2019). *BayesianTools: General-purpose MCMC and SMC samplers and tools for Bayesian statistics*. R package version 0.1.7. https://github.com/florianhartig/BayesianTools

Hastie, T. J. (2017). *Generalized additive models*. Routledge.

Hastings, A. (2001). Transient dynamics and persistence of ecological systems. *Ecology Letters*, *4*, 215–220. https://doi.org/10.1046/j.1461-0248.2001.00220.x

Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. (2013). The coyote universe extended: Precision emulation of the matter power spectrum. *The Astrophysical Journal*, *780*, 111. https://doi.org/10.1088/0004-637X/780/1/111

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, *103*, 570–583. https://doi.org/10.1198/016214507000000888

Hinzman, L. D., Viereck, L. A., Adams, P. C., Romanovsky, V. E., & Yoshikawa, K. (2006). *Climate and permafrost dynamics of the Alaskan boreal forest* (pp. 39–61). Alaska's Changing Boreal Forest.

Holmes, R. L. (1983). *Program COFECHA user's manual*. Laboratory of Tree-Ring Research, The University of Arizona.

Hooten, M., Wikle, C., & Schwob, M. (2020). Statistical implementations of agent-based demographic models. *International Statistical Review*, *88*, 441–461. https://doi.org/10.1111/insr.12399

Islam, M. A., MacDonald, S. E., & Zwiazek, J. J. (2003). Responses of black spruce (*Picea mariana*) and tamarack (*Larix laricina*) to flooding and ethylene. *Tree Physiology*, *23*, 545–552. https://doi.org/10.1093/treephys/23.8.545

Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J.-P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, *74*, 471–481. https://doi.org/10.1016/S0034-4257(00)00139-5

Johnson, S. E., & Abrams, M. D. (2009). Age class, longevity and growth rate relationships: Protracted growth increases in old trees in the eastern United States. *Tree Physiology*, *29*, 1317–1328.

Johnstone, J. F., & Chapin, F. S. (2006). Effects of soil burn severity on post-fire tree recruitment in boreal forest. *Ecosystems*, *9*, 14–31. https://doi.org/10.1007/s10021-004-0042-x

Johnstone, J. F., Hollingsworth, T. N., Chapin, F. S., & Mack, M. C. (2010). Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. *Global Change Biology*, *16*, 1281–1295. https://doi.org/10.1111/j.1365-2486.2009.02051.x

Jorgenson, M. T., Racine, C. H., Walters, J. C., & Osterkamp, T. E. (2001). Permafrost degradation and ecological changes associated with a warming climate in central Alaska. *Climatic Change*, *48*, 551–579.

Ju, L., Zhang, J., Meng, L., Wu, L., & Zeng, L. (2018). An adaptive Gaussian process-based iterative ensemble smoother for data assimilation. *Advances in Water Resources*, *115*, 125–135. https://doi.org/10.1016/j.advwatres.2018.03.010

Kane, E., Kasischke, E., Valentine, D., Turetsky, M., & McGuire, A. (2007). Topographic influences on wildfire consumption of soil organic carbon in interior Alaska: Implications for black carbon accumulation. *Journal of Geophysical Research: Biogeosciences*, *112*, 1–11. https://doi.org/10.1029/2007JG000458

Kelly, R., Chipman, M. L., Higuera, P. E., Stefanova, I., Brubaker, L. B., & Hu, F. S. (2013). Recent burning of boreal forests exceeds fire regime limits of the past 10,000 years. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 13055–13060.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 425–464. https://doi.org/10.1111/1467-9868.00294

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., & Dietze, M. C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, *83*, 133–154. https://doi.org/10.1890/12-0137.1

Lee, H., Gramacy, R., Linkletter, C., & Gray, G. (2011). Optimization subject to hidden constraints via statistical emulation. *Pacific Journal of Optimization*, *7*, 467–478.

Lin, C. D., & Tang, B. (2015). Latin hypercubes and space-filling designs. In *Handbook of design and analysis of experiments* (vol. 1, pp. 593–625).

Liu, F., Bayarri, M., & Berger, J. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, *4*, 119–150.

Liu, H., Ong, Y.-S., Shen, X., & Cai, J. (2020). When Gaussian process meets big data: A review of scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, *31*, 4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109

Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., & Schimel, D. S. (2011). Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications*, *21*, 1429–1442. https://doi.org/10.1890/09-1275.1

Nehrbass-Ahles, C., Babst, F., Klesse, S., Nötzli, M., Bouriaud, O., Neukom, R., Dobbertin, M., & Frank, D. (2014). The influence of sampling design on tree-ring-based quantification of forest growth. *Global Change Biology*, *20*, 2867–2885. https://doi.org/10.1111/gcb.12599

Nicklen, E. F., Roland, C. A., Csank, A. Z., Wilmking, M., Ruess, R. W., & Muldoon, L. A. (2019). Stand basal area and solar radiation amplify white spruce climate sensitivity in interior Alaska: Evidence from carbon isotopes and tree rings. *Global Change Biology*, *25*, 911–926. https://doi.org/10.1111/gcb.14511

Nicklen, E. F., Roland, C. A., Ruess, R. W., Scharnweber, T., & Wilmking, M. (2021). Divergent responses to permafrost and precipitation reveal mechanisms for the spatial variation of two sympatric spruce. *Ecosphere*, *12*, e03622. https://doi.org/10.1002/ecs2.3622

Nicklen, E. F., Roland, C. A., Ruess, R. W., Schmidt, J. H., & Lloyd, A. H. (2016). Local site conditions drive climate–growth responses of *Picea mariana* and *Picea glauca* in interior Alaska. *Ecosphere*, *7*, e01507.

Oberpriller, J., Cameron, D. R., Dietze, M. C., & Hartig, F. (2021). Towards robust statistical inference for complex computer models. *Ecology Letters*, *1*, 1–11. https://doi.org/10.1111/ele.13728

Osterkamp, T. (2005). The recent warming of permafrost in Alaska. *Global and Planetary Change*, *49*, 187–202. https://doi.org/10.1016/j.gloplacha.2005.09.001

Osterkamp, T., Jorgenson, M., Schuur, E., Shur, Y., Kanevskiy, M., Vogel, J., & Tumskoy, V. (2009). Physical and ecological changes associated with warming permafrost and thermokarst in interior Alaska. *Permafrost and Periglacial Processes*, *20*, 235–256. https://doi.org/10.1002/ppp.656

Osterkamp, T., & Romanovsky, V. (1999). Evidence for warming and thawing of discontinuous permafrost in Alaska. *Permafrost and Periglacial Processes*, *10*, 17–37.

Perkins, T. A., Reiner, R. C., España, G., ten Bosch, Q. A., Verma, A., Liebman, K. A., Paz-Soldan, V. A., Elder, J. P., Morrison, A. C., Stoddard, S. T., Kitron, U., Vazquez-Prokopec, G. M., Scott, T. W., & Smith, D. L. (2019). An agent-based model of dengue virus transmission shows how uncertainty about breakthrough infections influences vaccination impact projections. *PLoS Computational Biology*, *15*, e1006710. https://doi.org/10.1371/journal.pcbi.1006710

Peters, D. P., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., & Villanueva-Rosales, N. (2014). Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. *Ecosphere*, *5*, 1–15. https://doi.org/10.1890/ES13-00359.1

Pietzsch, B., Fiedler, S., Mertens, K. G., Richter, M., Scherer, C., Widyastuti, K., Wimmler, M.-C., Zakharova, L., & Berger, U. (2020). Metamodels for evaluating, calibrating and applying agent-based models: A review. *Journal of Artificial Societies and Social Simulation*, *23*, 9.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raiho, A., Rollinson, C., Dawson, A., Tipton, J., McLachlan, J., & Dietze, M. (2020). Determinants of predictability in multi-decadal forest community and carbon dynamics. *bioRxiv*.

Raiho, A., Scharf, H., Roland, C., Swanson, D., Stehn, S., & Hooten, M. (In Revision). Searching for refuge: A framework for identifying site factors conferring resistance to climate-driven vegetation change. *Diversity and Distributions*.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Renwick, K. M., Curtis, C., Kleinhesselink, A. R., Schlaepfer, D., Bradley, B. A., Aldridge, C. L., Poulter, B., & Adler, P. B. (2018). Multi-model comparison highlights consistency in predicted effect of warming on a semi-arid shrub. *Global Change Biology*, *24*, 424–438. https://doi.org/10.1111/gcb.13900

Roland, C. A., Schmidt, J. H., & Nicklen, E. F. (2013). Landscape-scale patterns in tree occupancy and abundance in subarctic Alaska. *Ecological Monographs*, *83*, 19–48. https://doi.org/10.1890/11-2136.1

Roland, C. A., Schmidt, J. H., Winder, S. G., Stehn, S. E., & Nicklen, E. F. (2019). Regional variation in interior Alaskan boreal forests is driven by fire disturbance, topography, and climate. *Ecological Monographs*, *89*, e01369.

Santner, T. J., Williams, B. J., & Notz, W. I. (2018). *Empirical best linear unbiased prediction of computer simulator output*. Springer.

Scharf, H. R., Raiho, A. M., Pugh, S., Roland, C. A., Swanson, D. K., Stehn, S. E., & Hooten, M. B. (2021). Multivariate Bayesian clustering using covariate-informed components with application to boreal vegetation sensitivity. *Biometrics*, 1–10. Early View. https://doi.org/10.1111/biom.13507

Scheffer, M., Hirota, M., Holmgren, M., Van Nes, E. H., & Chapin, F. S. (2012). Thresholds for boreal biome transitions. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 21384–21389. https://doi.org/10.1073/pnas.1219844110

Shugart, H., Foster, A., Wang, B., Druckenbrod, D., Ma, J., Lerdau, M., Saatchi, S., Yang, X., & Yan, X. (2020). Gap models across micro-to mega-scales of time and space: Examples of Tansley's ecosystem concept. *Forest Ecosystems*, *7*, 1–18.

Shur, Y. L., & Jorgenson, M. (2007). Patterns of permafrost formation and degradation in relation to climate and ecosystems. *Permafrost and Periglacial Processes*, *18*, 7–19. https://doi.org/10.1002/ppp.582

Sniderhan, A. E., & Baltzer, J. L. (2016). Growth dynamics of black spruce (*Picea mariana*) in a rapidly thawing discontinuous permafrost peatland. *Journal of Geophysical Research: Biogeosciences*, *121*, 2988–3000.

Speer, J. H. (2010). *Fundamentals of tree-ring research*. University of Arizona Press.

Speich, M., Dormann, C. F., & Hartig, F. (2021). Sequential Monte-Carlo algorithms for Bayesian model calibration–a review and method comparison. *Ecological Modelling*, *455*, 109608. https://doi.org/10.1016/j.ecolmodel.2021.109608

Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (2020). Deep learning optimizes data-driven representation of soil organic carbon in Earth system model over the conterminous United States. *Frontiers in Big Data*, *3*, 17. https://doi.org/10.3389/fdata.2020.00017

Temperli, C., Zell, J., Bugmann, H., & Elkin, C. (2013). Sensitivity of ecosystem goods and services projections of a forest landscape model to initialization data. *Landscape Ecology*, *28*, 1337–1352. https://doi.org/10.1007/s10980-013-9882-0

Thomson, M. C., Doblas-Reyes, F., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A., & Palmer, T. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, *439*, 576–579. https://doi.org/10.1038/nature04503

Tuo, R., & Jeff Wu, C. (2016). A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, *4*, 767–795.

Tuo, R., & Wu, C. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics*, *43*, 2331–2352. https://doi.org/10.1214/15-AOS1314

Turnbull, L. A., Levine, J. M., Loreau, M., & Hector, A. (2013). Coexistence, niches and biodiversity effects on ecosystem functioning. *Ecology Letters*, *16*, 116–127.

Van Cleve, K., Dyrness, C., Viereck, L., Fox, J., Chapin, F., III, & Oechel, W. (1983). Taiga ecosystems in interior Alaska. *BioScience*, *33*, 39–44. https://doi.org/10.2307/1309243

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., & Marathe, M. (2018). Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*, *22*, 43–49. https://doi.org/10.1016/j.epidem.2017.02.010

Verrelst, J., Sabater, N., Rivera, J. P., Muñoz-Mari, J., Vicent, J., Camps-Valls, G., & Moreno, J. (2016). Emulation of leaf, canopy and atmosphere radiative transfer models for fast global sensitivity analysis. *Remote Sensing*, *8*, 673.

Wagner, R. G., & Robinson, A. P. (2006). Critical period of interspecific competition for four northern conifers: 10-year growth response and associated vegetation dynamics. *Canadian Journal of Forest Research*, *36*, 2474–2485. https://doi.org/10.1139/x06-058

Wikle, C. K., & Hooten, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST*, *19*, 417–451. https://doi.org/10.1007/s11749-010-0209-z

Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., & Bower, M. R. (2018). Monitoring dynamic spatio-temporal ecological processes optimally. *Ecology*, *99*, 524–535. https://doi.org/10.1002/ecy.2120

Wirth, C., Lichstein, J., Dushoff, J., Chen, A., & Chapin, F. (2008). White spruce meets black spruce: dispersal, postfire establishment, and growth in a warming climate. *Ecological Monographs*, *78*, 489–505.

Yan, X., & Shugart, H. (2005). Fareast: a forest gap model to simulate dynamics and patterns of eastern Eurasian forests. *Journal of Biogeography*, *32*, 1641–1658. https://doi.org/10.1111/j.1365-2699.2005.01293.x

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.