The devil is in the details: Application of emulation-based data assimilation techniques to constrain a dynamic ecosystem model

By

Bailey A. Murphy

A thesis submitted in partial fulfillment of

the requirements for the degree of

MASTER OF SCIENCE

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2020

Abstract

Exchanges of carbon, water, and energy between the terrestrial biosphere and atmosphere are major drivers of the Earth's climate system, yet interactions between them remain the second-largest uncertainty in climate projections (Fisher et al., 2017). A key aspect of reducing this uncertainty lies in better parameterizing and calibrating terrestrial ecosystem models using data assimilation (DA) methods. In the atmospheric sciences, DA is often synonymous with addressing initial condition and forcing uncertainty in weather forecast models. However, for the land surface, and for its carbon, water, and energy exchange, uncertainty predominantly comes from the representation of biological processes and most importantly, the parameters that describe those processes (Dietze, 2017). Until recently, the application of Bayesian DA to complex ecosystem models has been computationally prohibitive, and thus was not a viable option for model calibration. With the recent application of statistical emulation techniques to traditional DA methodology, many of these barriers have been overcome (Fer et al., 2018).

In this study, the Ecosystem Demography model version 2.2 was used for the testing and application of an emulator approach to Bayesian parameter data assimilation, with model runs occurring at a study site in Northern Wisconsin's Chequamegon-Nicolet National Forest.

This study seeks to answer two primary questions:

1. Can model parameter data assimilation using a Bayesian emulator approach reduce individual parameter uncertainty?
2. Can parameter data assimilation using an emulator approach reduce overall model predictive uncertainty related to carbon cycle processes in a structurally complex ecosystem model?

Eight plant physiological and soil biogeochemical parameters contributing to the largest proportion of net ecosystem exchange (NEE) model predictive uncertainty were identified through sensitivity analysis and variance decomposition. Individual parameter uncertainty was reduced following PDA for the majority of targeted parameters, with remaining uncertainty primarily shifting to two parameters controlling respiration processes (growth respiration and respiration temperature increase). Overall model predictive uncertainty increased by 13.74% as a result of PDA, although distinct seasonality was observed. The confidence interval spread of model NEE predictions was successfully constrained during the growing season, but model performance was poor during winter and fall, due primarily to dramatically enhanced respiration predictions.

Acknowledgements

I would like to thank my fellow Desai lab mates for listening to me word vomit as I attempted to sort out problems, and Dr. Ankur Desai for being the most supportive advisor I could ask for and for always making space and time to help me. Every time we have a research meeting I leave with renewed excitement about the work I'm doing and all the cool 'idea tangents' that are out there to follow. I would like to thank my other committee members, Dr. Chris Kucharick and Dr. Jack Williams for taking the time to provide both feedback and support on my project, and reign things in when needed. Thank you to the PEcAn community for always being willing to help troubleshoot and offer suggestions, and to Stack Overflow, without which my programming skills would be garbage. Perhaps most importantly, thank you to my partner, family, and friends for putting up with me being super stressed out and boring and spending all my time working (sorry).

Lastly, I would like to acknowledge the financial support that allowed for this research to take place from the US National Science Foundation under The PEcAn Project: A Community Platform for Ecological Forecasting (NSF DBI #1457897), and the Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors (NSF AGS #1822420) as well as from the University of Wisconsin-Madison Center for Climatic Research Climate, People, Environment Program (CPEP).

Table of Contents

Abstract……i

Acknowledgements……..……..iii

Table of contents……iv

1. Introduction and background……....1

1.1 Models and the global carbon cycle…….....1

1.2 Model calibration and data assimilation……...6

a. Uncertainty in ecosystem modeling…….....9

b. Data assimilation in structurally simple models……..14

c. Data assimilation in structurally complex models……...16

2. Methods……..17

2.1 Experimental design……...17

2.2 Site description…….....19

2.3 Model description……...21

2.4 The Predictive Ecosystem Analyzer (PEcAn) as a workflow tool……..24

a. Integrated plant trait database…….....25

2.5 Bayesian approach to parameter data assimilation…….26

2.6 Basis of emulation for parameter data assimilation……28

b. The emulator method……..29

2.7 Process of model calibration……...33

c. Data preparation…….33

d. Uncertainty examination……35

e. Parameter data assimilation……37

f. Model performance assessment……..41

2.8 Statistical analysis……42

3. Results……47

3.1 Parameter selection……47
3.2 PDA success……...56
3.3 Parameter constraint through PDA……60
3.4 Model predictive uncertainty……..65
4. Discussion……..70
4.1 Parameter uncertainty……70
4.2 Seasonality in predicted NEE……74
5. Conclusions and Future Work……77
References……..80

1. Introduction and Background

1.1 Models and the global carbon cycle

Although they are far from perfect, models are one of the most common tools used by scientists to understand and predict changes in the earth system over time and space (Dietze et al., 2014). Earth system models simulate interactions among component land, atmosphere, ocean, and sea ice models to represent climate as the outcome of interacting physical, chemical, and biological processes (Bonan, 2019). Terrestrial models address processes happening at the earth's surface, in vegetated ecosystems and the interfaces between them. Over time terrestrial models have expanded in complexity as the questions we sought to answer became more comprehensive and computational ability improved.

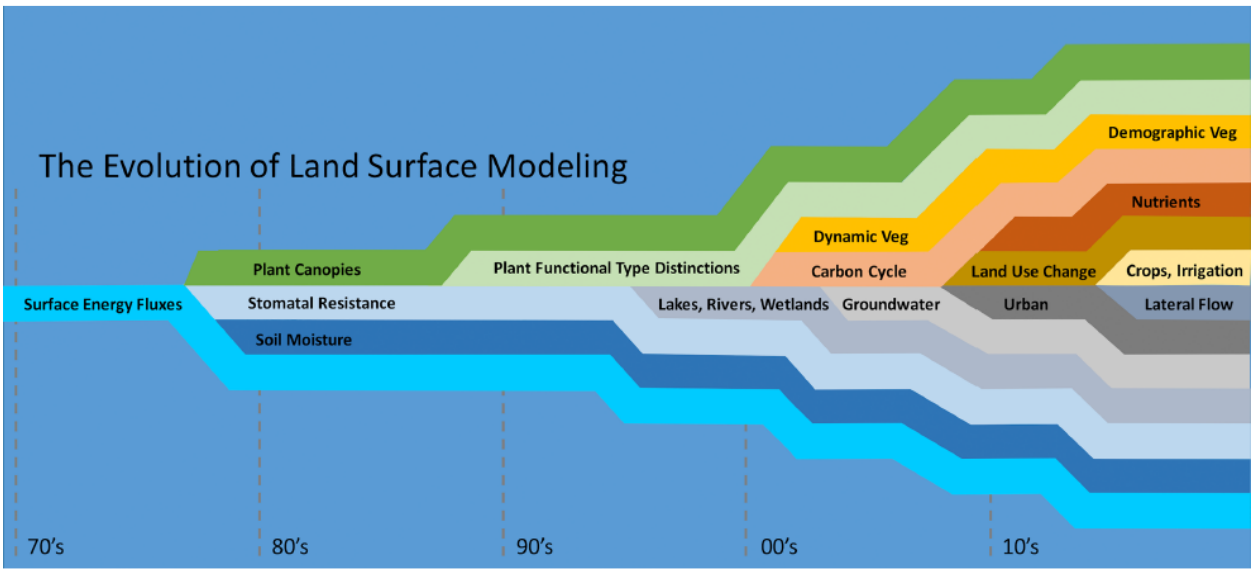

Figure 1: *Evolution of land surface models over time, demonstrating the increasing complexity of biological process representation (Fisher et al., 2020).*

Originating in the 1960's as simplified linear differential equations to represent material fluxes between 'compartments', by the late 1980's models were capable of representing plant canopies and soil hydrological properties, and by the early 2000's terrestrial models were able to simulate dynamic vegetation and capture complex carbon and water cycle interactions (Bonan, 2019). Results from some of the first fully coupled land surface model runs solidified the importance of the terrestrial carbon cycle in relation to global climate change. Model results showed that not only could the plant biophysical response to elevated levels of atmospheric CO_2 be significant enough to influence global climate, but that feedbacks between climate and the terrestrial carbon cycle could lead to an acceleration of global climate change significantly beyond what had originally been predicted by the scientific community (Cox et al., 2000; Sitch et al., 2003; Friedlingstein et al., 2013; Fisher et al., 2020).

A great deal of uncertainty was associated with these early predictions, a significant portion of which was due to terrestrial carbon dynamics and whether landscapes were classified as carbon sinks or carbon sources (Fisher et al., 2020). Tightly coupled to the water cycle, terrestrial ecosystems can act as either carbon sinks (through photosynthesis and primary production) or carbon sources (soil and plant respiration, organic matter decomposition, plant mortality and combustion), and provide climate feedbacks through latent heat fluxes, albedo, and water cycling (Fisher et al., 2017), although the magnitude of this forcing is not well constrained (Bonan, 2008). Globally, the terrestrial biosphere serves as a net carbon sink, and has taken up approximately 31% of the cumulative carbon emitted due to anthropogenic activities such as fossil fuel combustion and land-use change over the period of 1870-2017 (Le Qu*é*r*é* et al., 2018). The majority of this uptake occurs in forests, which are the largest terrestrial carbon sink on earth (Canadell and Schulze, 2014) and sequester extensive amounts of atmospheric carbon in both biomass and soils.

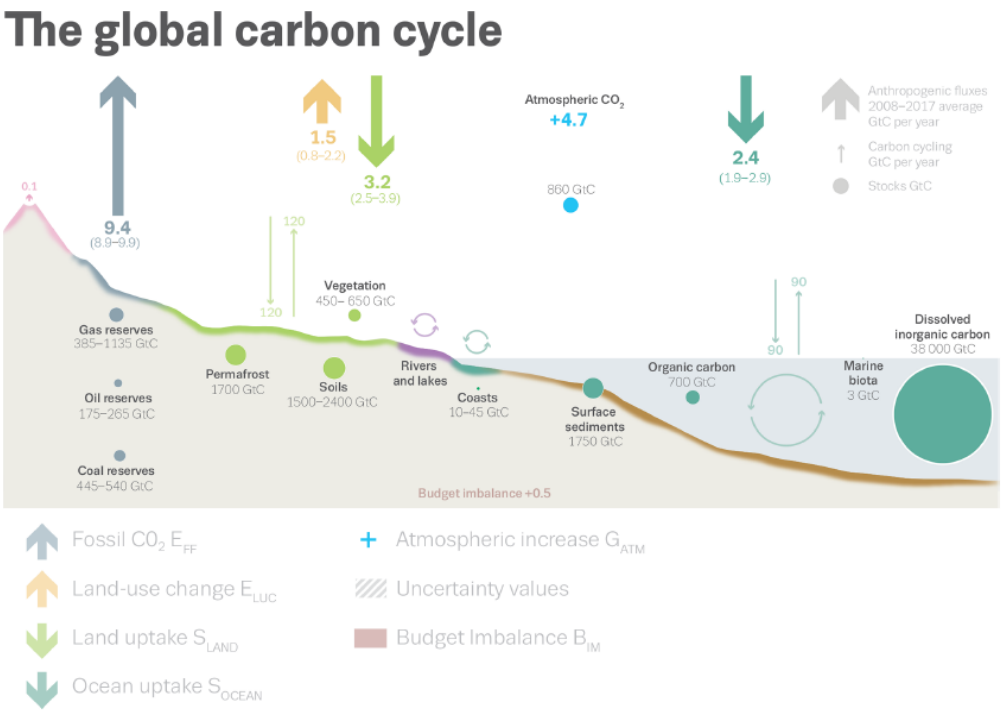

Figure 2: *Perturbation of the global carbon cycle caused by anthropogenic activities, represented as fluxes and stocks. Values shown are averaged over the period from 2008-2017, units are GtC/yr (Le Quéré et al., 2018).*

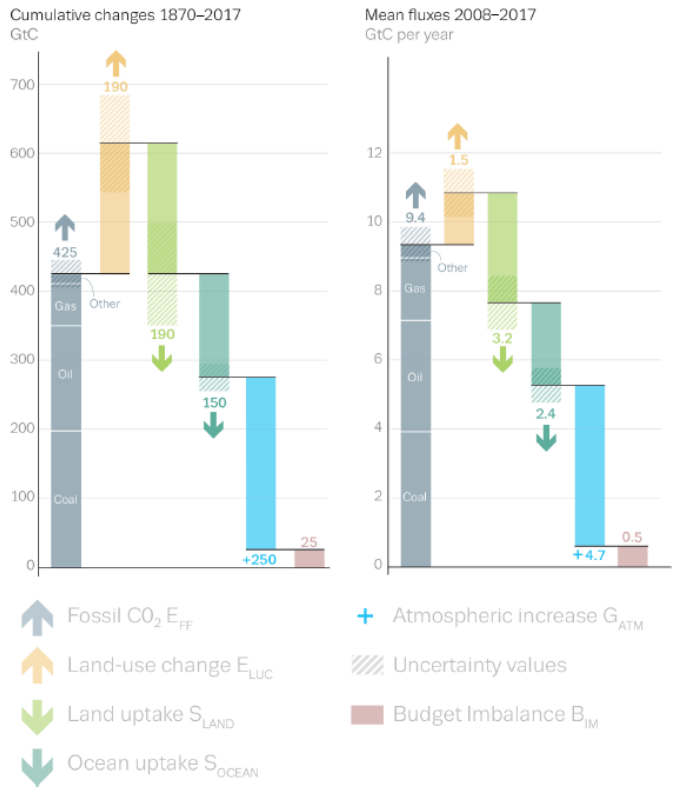

Figure 3: *Anthropogenic perturbation carbon flows displayed as cumulative changes during 1870-2017 (left) and mean fluxes per year from 2008-2017 (right), both in GtC (Le Quéré et al., 2018).*

Forests cover approximately 30% of the land surface (Bonan, 2008), existing in a wide range of climatic zones and with a high degree of structural and functional diversity. In addition to their influence on climate, forests provide essential ecological, economic, social, and aesthetic services to both humans and natural systems (Bonan, 2008). In pursuit of these ecosystem services, humans have used forest management techniques to alter the natural forest state, such as through supporting accelerated tree growth for logging purposes. Human management of terrestrial ecosystems drastically impacts the already complex land-atmosphere relationships, altering the magnitude and seasonality of carbon source and sink dynamics and further driving interannual

variability in atmospheric greenhouse gas concentrations (Desai et al., 2007). Predicting how forest carbon dynamics will shift in response to climate change is a challenge made more complex by the high degree of variability across environmental gradients and the broad range of spatial and temporal scales at which ecosystem processes operate. The ability to simulate carbon cycle dynamics in terrestrial systems through the use of models is one solution to this problem, and has been a major development in understanding the influence of forests on large-scale climate. Model simulations are especially useful tools to examine the carbon cycle, as many aspects of the terrestrial carbon cycle are difficult to measure directly (LeBauer et al., 2013; Bastin et al., 2019).

There are three general approaches to quantifying carbon dynamics in terrestrial systems: (1) direct measurement of carbon fluxes, (2) measuring changes in carbon stocks over time, and (3) the use of models to simulate the key processes involved in moving carbon through an ecosystem (Williams et al., 2005). Carbon fluxes can be directly measured using eddy covariance flux towers and leaf-level gas exchange measurements (Baldocchi, 2003). These tools can provide local-scale insight into fast carbon cycle processes and forest-atmosphere interactions at a nearly continuous timescale, but are not without limitations. Flux data frequently have gaps in measurement, night-time observations are often unreliable due to insufficient air mixing, the spatial scale of measurements are continuously changing with shifting wind speed and direction, and complex terrain can violate the basic assumptions of the eddy covariance approach (Williams et al., 2005). Carbon stocks can be determined by measuring the mass of carbon stored in various pools and calculating changes in stock size over time. Carbon pools include carbon stored in living biomass (plant material such as leaves, stems, and roots), litter, and soil. The soil carbon pool is often further partitioned based on biochemical properties and turnover times (Raupach et al., 2005). However, carbon stock measurement is labor intensive, spatially sparse and discontinuous,

and belowground pools are difficult to systematically measure in a reliable way (Williams et al., 2005). The third approach, modeling, integrates data from the previous two approaches to simulate carbon fluxes and pool storage based on established relationships and foundational laws of physics to predict changes in carbon dynamics over time or in response to perturbation.

1.2 Model calibration and data assimilation

Due to the large global variation in plant physiology and soil biogeochemistry, coupled with the strong dependence of many ecosystem models on localized ecosystem interactions, models must be calibrated at a given geographic location prior to running in order to produce representative output (Longo et al., 2019). Calibration is the process of choosing parameter or state variable values to optimize the fit of the model to test data from the location of interest (Van Oijen et al., 2017). The process of calibration can incorporate data from a variety of sources, including eddy covariance flux towers, ground collected vegetation and soils data, and plant physiology and soil biogeochemistry parameter information from existing literature.

The values that flexible parameters within a model are set to can have a substantial impact on model predictions. Models are more or less sensitive to different parameters, such that not all parameters carry equal weight in causing variance or uncertainty in model predictions (Van Oijen et al., 2017). Influential parameters also vary depending on plant functional types (PFT) present or the chosen model run duration. For example, leaf width, a parameter that controls leaf boundary layer conductance and thus affects carbon and moisture fluxes and the leaf-level energy budget, would likely be a low impact parameter for a deciduous PFT during the winter when leaves have been shed, but might become an influential parameter come spring when leaves emerge. Traditionally parameters have been represented with single point estimates, which are then subjectively ‘tuned’ to best match observations. However, the use of single point estimates

effectively treats each flexible parameter as certain instead of allowing it to fluctuate within a reasonable range based on what is known about the parameter from observations (LeBauer et al., 2013), and the subjectivity of the tuning process introduces additional unquantifiable uncertainty. A Bayesian DA approach to model calibration is an attractive solution to this problem, as it allows parameters to vary across a range of values derived from observational data and avoids the introduction of bias from manual tuning. Additionally, classical model calibration techniques are often plagued with issues of equifinality when applied to models with a large number of flexible parameters, as there can be multiple combinations of parameter values that produce results to best match the data, regardless of whether the values are biologically feasible, often resulting in the model getting the 'right' answer for the wrong reasons (Williams et al., 2005).

Models overcome barriers such as spatial limitations and measurement difficulties associated with the other terrestrial carbon quantification methods presented, but predictions can carry a great deal of uncertainty depending on how the model represents a system and the parameterization techniques used. In order to make model predictions reliable as a tool to both understand and respond to the challenge of climate change, uncertainty needs to not only be quantified, but steps must be systematically taken to reduce uncertainty in projections. A critical aspect of this is understanding the contributions to overall model error from all sources of uncertainty (data, parameter, model structure) and focusing efforts on constraining the most influential sources. DA is a technique that capitalizes on all three of the carbon quantification approaches presented to combine observational data with models and provide improved estimates of ecosystem carbon stocks and fluxes, along with rigorously tracked uncertainty estimates (Dietze et al., 2013).

The overarching goal of data assimilation in an ecosystem modeling framework is to improve the ability to predict ecosystem response to a perturbation or scenario of interest. The scenario of interest might be net ecosystem exchange under elevated atmospheric CO_2 concentrations, the effect of nitrogen fertilization on above-ground biomass production, or the impact of water limitation on evapotranspiration. There are multiple methods of model calibration, but utilizing a data assimilation approach is of interest for three key reasons. (1) Data assimilation (often called 'model-data-synthesis') refers to the combination of information contained within both observational data and models through statistical techniques (Raupach, 2005), and can be applied to both parameters and state variables. (2) DA allows for the interpolation of data where observations might be temporally or spatially sparse, or where observations cannot be directly made with current technology, such as for carbon fluxes between pools over large study areas (Raupach, 2005). (3) DA facilitates model testing and quality control through the incorporation of a statistically rigorous framework of accounting for uncertainty that allows for the quantification of how uncertainty around parameters or model predictions shifts in response to changes made to the model.

Although data assimilation can target both parameters and state variables, and traditionally has been applied more often to state variables (Dietze et al., 2013), this study utilizes DA to constrain parameters for two key reasons. (1) In structurally complex deterministic models with a large number of flexible parameters (such as most ecosystem models), uncertainty related to parameter values is a primary driver of overall model predictive uncertainty (LeBauer et al., 2013). (2) Biological systems tend to have stabilizing feedbacks that make state variables more predictable at certain spatial and temporal scales, but in ecological models process equations and their associated parameters are statistical approximations of diverse interactions that are not always

fully understood (Dietze et al., 2013), and thus their mathematical representation in a model is not exact. This is in contrast to physical science applications such as numerical weather prediction (where DA techniques originated from) where the equations and parameters governed by physical laws are established, but state variables are highly divergent with time (Dietze et al., 2013).

a. Uncertainty in ecosystem modeling

The importance of characterizing and reducing uncertainty related to model predictions has now been mentioned several times, but what is uncertainty and why is it so important? Simply put, uncertainty is the degree to which knowledge about a quantity is incomplete (Van Oijen, 2017). Since models are a tool to predict the future state of a system and complete knowledge about the future is impossible, uncertainty is a central pillar of any modeling endeavor. Uncertainty is such a core element of modeling that many seminal papers in the field of model-data synthesis argue that without proper accounting and communication of uncertainty, model results are essentially meaningless (Raupach et al., 2005). For reliable model calibration results using data assimilation methods, uncertainty from all sources must be tracked at each step in the assimilation process. Uncertainty itself can be represented by a probability distribution, which outlines the error values that are deemed statistically possible and their probability of occurrence (Van Oijen, 2017). This study utilized a Bayesian approach, so the probability distributions used to describe uncertainty are conditional probability distributions, and the degree of uncertainty related to a parameter is the standard deviation of a model parameter's posterior probability distribution (The PEcAn Project). Predictive variance, or the degree of uncertainty about a prediction made by the model, increases with time and a prediction is shown to no longer be useful when the model fails to predict a future state any better than the outcome of random chance could. The degree of variance associated with a model prediction is described by the following equation from Dietze 2017:

$$\mathrm{Var}[Y_{t+1}] \approx \underbrace{\left(\frac{\partial f}{\partial Y}\right)^2}_{\text{stability}} \underbrace{\mathrm{Var}[Y_t]}_{\text{IC uncert}} + \underbrace{\left(\frac{\partial f}{\partial X}\right)^2}_{\text{driversens}} \underbrace{\mathrm{Var}[X]}_{\text{driveruncert}} + \underbrace{\left(\frac{\partial f}{\partial \theta}\right)^2}_{\text{paramsens}} \left(\underbrace{\mathrm{Var}[\bar{\theta}]}_{\text{paramuncert}} + \underbrace{\mathrm{Var}[\alpha]}_{\text{paramvariablility}} \right) + \underbrace{\mathrm{Var}[\varepsilon]}_{\text{processerror}}$$

Equation 1: *equation describing the degree of variance associated with a prediction made about the future state of a given system*

Where Y_{t+1} is the future state of the system, Y_t is the current state of the system, and X is the internal dynamics and external driver/covariates that describe the relationship between the two system states over time. θ represents the set of parameters required by the function describing the effect of Y_t and X on Y_{t+1}, and is split into two components: $\bar{\theta}$ which describes the average value of the parameter and α, which represents the random effect, or the deviation from the parameter mean experienced at a given point in time or space. ε is the process error, which essentially represents the error arising over time from an imperfect representation of processes by the model (Van Oijen, 2017). With this equation predictive variance can be described as the sum of variances contributed by internal dynamics, external drivers/covariates, parameter, and random effects, as well as general process error. The contribution of each factor to overall predictive uncertainty is contingent upon the sensitivity of the model to that factor in addition to the degree of uncertainty associated with the factor. Unlike in numerical weather prediction, parameters in most ecological models are empirically estimated coefficients, not physical constants, which makes them significantly more uncertain. This combined with the fact that process-based ecological models

are generally parameter rich results in a disproportionately large contribution to overall predictive uncertainty from model parameters (LeBauer et al., 2013). In light of this, data assimilation efforts to calibrate ED2.2 (a model with a large number of flexible parameters) were targeted towards parameters instead of state variables in an attempt to reduce predictive uncertainty to the largest degree possible. The following is a more in-depth discussion of each uncertainty source term shown in equation 1: (1) internal dynamics, (2) external drivers or covariates, (3) parameters, (4) random effects, and (5) general process error.

(1) Uncertainty related to internal dynamics and initial conditions are encapsulated in the first term of equation 1. $\frac{\partial f}{\partial Y}$ describes the stability of the system in question, with a value greater than one meaning the system is actively changing and is regarded as unstable. If the system is unstable, initial condition uncertainty VAR[Y_t] will rapidly increase with time and will constitute the majority of predictive uncertainty. This situation is observed in weather forecasting, where parameters are physical constants but initial conditions are based on an inherently unstable atmosphere and can evolve chaotically over time. Ecosystems have a variety of self-stabilizing feedbacks that reduce the contribution of initial conditions to overall uncertainty, so this term can be regarded as small when applying equation 1.

(2) The second term of equation 1 describes model sensitivity to external forcing. The degree of sensitivity can vary depending on spatial or temporal scale (ex: system could be highly sensitive to some factor ‘B’ when making short term prediction, but the influence of ‘B’ could diminish over longer timescales). For models with multiple drivers, interactions between drivers can amplify or diminish uncertainty, depending on wether covariance between drivers is positive or negative.

(3) The third term addresses contributions from parameter sensitivity and uncertainty. Parameter uncertainty [Var($\bar{\theta}$)] refers to the uncertainty about the true values of the model parameters due to data deficiency and model simplification (McMahon et al., 2009; van Oijen, 2017). Parameter sensitivity describes the degree of influence a parameter has on model predictions.

(4) And (5) Capture the remaining unexplained variability, but since these terms quantify error in how ecological processes themselves are represented in a model, they cannot be constrained through data assimilation. These error terms can come from the equations chosen to represent each biological process represented within the model as well as from approximations employed to reduce computational expense (Diezte 2017). Equation choice is particularly influential because as stated previously, most equations in ecological models are empirical calibrations, not physical laws. In light of this, when comparing predictions made by a suite of different ecological models a significant portion of the predictive variability can be attributed to differences in equations used by the models to represent the same biological processes, but the influence of these differences can be especially difficult to isolate and quantify (Keenan et al., 2011).

In addition to the sources of uncertainty captured in equation 1, uncertainty can also be associated with the observational data used to drive a model. Uncertainty in observational data can result from gaps in the measured data, systematic bias, or general noise due to instrument quality. A large portion of the data used to drive ED2.2 in this study comes from eddy covariance towers, which have substantial uncertainty associated with measurements due to time-varying flux footprints and the influence of heterogeneous surface conditions. Surface heterogeneity often contributes significantly to uncertainty because the eddy covariance calculation method itself

assumes relatively homogenous surface characteristics (Baldocchi, 2008). Often times the range of random measurement uncertainty from flux tower instrumentation can be of a significant magnitude in comparison to the measurements themselves (Richardson et al., 2008). Systemic bias in flux tower data incorporates error due to insufficient mixing at night, and issues with energy balance closure. In an attempt to reduce systemic bias, most nocturnal measurements were removed from the Park Falls US-PFa tower dataset used in this study by applying a friction velocity (u^*) filter of 0.4 ms^{-1}. Flux values observed when values of u^* are below 0.4 ms^{-1} were screened out because stable conditions observed during the night complicate NEE calculations due to drainage flows and a lack of turbulent transport, and make NEE calculations unreliable (Davis et al., 2003). Although process and parameter uncertainties both need to be propagated into model predictions, observation error due to imperfect instrumentation isn't of interest to model, so although it should be quantified it doesn't need to be propagated forward (Fer et al., 2018).

b. Data assimilation with respect to NEE

DA is applied to improve model output by optimizing model fit based on a predictive variable of interest, which is often net ecosystem exchange (NEE) in studies related to representing carbon flux through a system.

$$\mathbf{NEE = -NEP = R_e - GPP}$$

Equation 2: *expression for net ecosystem exchange (NEE)*

NEE between terrestrial systems and the atmosphere is the balance of the movement of carbon into and out of an ecosystem, and is defined as the opposite sign of net ecosystem production (NEP), so that as a convention fluxes into an ecosystem (uptake of carbon by the system) are expressed as negative (Lovett et, al, 2006). NEP is the difference between the amount of CO_2 fixed by plants

via photosynthetic uptake (gross primary production, GPP) and the total ecosystem respiration (R_e). R_e is the sum of respiratory efflux of CO_2 from both heterotrophic (R_h) and autotrophic (R_a) sources (Law et al., 2002).

c. Data assimilation in structurally simple models

Brief overviews of two case studies are presented to highlight the success of DA in reducing model uncertainty and improving predictions for structurally simple models, and demonstrate the potential implications for application to complex ecosystem models. The studies presented are (1) the application of a carbon cycle mass balance to predict NEE by Williams et. al and (2) the dual assimilation of a 'green slime' ecosystem model with respect to both NEE and LE to predict gross primary productivity and water cycle dynamics by Moore et. al.

A 2005 study by Williams et. al used a simple carbon cycle mass balance box model to predict NEE in an Oregon ponderosa pine forest. The model had five carbon pools and nine flexible parameters. The use of a DA approach to calibrate the model with respect to NEE improved model predictions of NEE while reducing uncertainty in predictions. Average NEE estimates over the course of three years from the model alone were -251 + 197 g C m-2 compared to -419 + 29 g C m-2 predicted by the calibrated model. In generating carbon budgets, data assimilation was also shown to produce statistically unbiased estimates of NEE, compared to predictions made using the data or model alone. The study noted that a significant portion of the predictive error came from the model's oversimplification of processes (seasonal dynamics in litter turnover and foliar phenology, etc.), and that using a more detailed model in place of the simple model would provide 'enhanced analysis' and further reduction in error (Williams et al., 2005).

The dual parameterization of NEE and ET approach was tested for a simple model in a 2008 study by Moore et. al., comparing observational NEE and ET fluxes to those predicted by a model parameterized using data assimilation with respect to only NEE as well as both NEE and ET. The model used in this study was SIPNET (Simplified PnET model), a simplified ecosystem model that consists of three carbon pools (two vegetation pools and one soil pool), and three carbon fluxes to simulate carbon exchange between pools and the atmosphere. SIPNET has a total of 25 flexible parameters and initial conditions, but is much simpler in structure than ED2.2, as it represents a select few processes and drastically simplifies the processes included (Braswell et al., 2005).

Estimating transpiration using SIPNET with combined assimilation of NEE and ET resulted in a better match to observed data, and accounted for 67% of observed seasonal variation in transpiration (T) (Moore et al., 2008). Four key parameters related to resolving ecosystem water balance differed significantly depending on which optimized version of the model was used. These parameters included initial soil moisture content, water-use efficiency constant, soil water holding capacity, and aerodynamic resistance to wind speed (Moore et al., 2008). The improved representation of initial soil moisture content and water holding capacity is promising when considering applying the same approach to ED2.2 optimization, as ED2.2 has been shown to be highly sensitive to soil moisture properties (Longo et al., 2019). The two model versions also diverged in response to perturbation of meteorological inputs, such as precipitation. The NEE only version of SIPNET was non-responsive to changes in precipitation up to a reduction of 50%, whereas the NEE/ET version of the model was sensitive to even small changes in precipitation (Moore et al., 2008). As available precipitation is predicted to experience drastic fluctuations in a changing climate, and is closely tied to both ET and available soil moisture for ecosystem process

as well as overall ecosystem resilience, a realistic model sensitivity to fluctuations is desirable (Fisher et al., 2016). When translating precipitation and GPP estimates to water-use efficiencies (the ratio of carbon gain to water lost, the photosynthesis/transpiration tradeoff mentioned previously), the model version with parameters assimilated with respect to NEE alone predicted a ratio of GPP:T an order of magnitude higher than produced by the model conditioned with respect to both NEE and ET (Moore et al., 2008).

Although DA has demonstrated success in reducing uncertainty in simple models, it can be argued that the application of DA to constrain simple ecosystem models is subject to the same issues of equifinality previously discussed with respect to traditional model calibration methods, and thus is fundamentally flawed. Without incorporating sufficient processes to describe ecosystem interactions, actual process error is underestimated and parameter values could potentially be outside the realm of what is physically realistic, thus the calibrated model is overfit.

d. Data assimilation in structurally complex models

Data assimilation involves the use of complex statistical techniques such as Markov chain Monte Carlo to resolve probability density functions of parameters and the model state as they evolve over time, a process that requires large ensembles of model runs (Zobitz et al., 2011). Depending on the parameters of interest and the existing data available, this can mean 10^4-10^7 model runs to isolate ideal parameter or state variable ranges (Fer et al., 2018). For simple models with few flexible parameters this is not typically prohibitive, but for complex models data assimilation becomes computationally unrealistic, both because number of parameters increases by an order of magnitude or more and computational run time for a single realization is much longer. In order to apply DA techniques to models that are capable of representing complex ecosystem interactions, techniques must be adapted to reduce computational expense. A solution

to this problem is the use of model emulation, or an 'emulator' during DA. An emulator is a simplified statistical model ran in place of a full model for calibration purposes (Wang, 2014). The use of an emulator significantly reduces the computational demands of calibration, and allows DA techniques to be applied to complex simulation models. For example, a study by Fer et. al. in 2018 comparing data assimilation-based model calibration methods showed that calibrating ED2.2 using a class of emulator generated from the difference between model and observation error space as opposed to emulating the model alone, led to a decrease in computational time and improved model prediction. For example, the emulator calibration of ED2.2 using 100K iterations took roughly 27 hours to complete, whereas a traditional DA approach for the same number of model runs would have taken approximately 74 years of wall time. The integration of statistical emulators within traditional DA methods is a recent development in the field of ecosystem modeling, and although studies have demonstrated the ability of DA to reduce uncertainty around model predictions and improve our capacity to predict ecosystem dynamics, further testing is required to explore the potential of emulator based DA for structurally complex models and determine if the same enhancements can be achieved.

2. Methods

2.1 Experimental design

In this study, the Ecosystem Demography model version 2.2 (ED2.2) was used for the testing and application of a model calibration process using an emulator approach to Bayesian PDA. The use of a structurally complex model such as ED2.2 facilitated an investigation into the utility of emulator-based PDA to constrain individual parameter contributions to model predictive uncertainty, the primary source of uncertainty in dynamic ecosystem models (Dietze et al., 2013).

This study employed a unique approach to emulation, outlined in Fer et al., 2018. Instead of constructing an emulator from the direct model output, an emulator was constructed using sufficient statistics of the likelihoods (probability of getting '*X*' model output given vector '*Y*' model parameters), reflecting the interest in the magnitude of disagreement between the model and data and allowing for the retention of statistical parameters of the likelihood related to predictive error.

In order to address whether model predictions of NEE can be improved by reducing parameter uncertainty through PDA, a series of modeling experiments were developed. PDA was conducted with respect to NEE, the predictive variable of interest. Model ensembles were used to ensure that the full range of predictions were captured, with each ensemble consisting of 100 individual runs. Two separate ensembles were used to assess baseline model predictive performance and improvement as a result of PDA. Sensitivity analysis and variance decomposition were performed to assess individual parameter contribution to model predictive uncertainty, and were conducted twice, once prior to PDA to establish initial distributions of uncertainty and once post-PDA to assess changes in parameter uncertainty as a result of PDA. Model runs occurred at a study site in Northern Wisconsin's Chequamegon-Nicolet National Forest, with meteorological and gas flux data supplied by the WLEF tall tower (Ameriflux site US-PFa), a 447m eddy covariance flux tower with an observation record dating back to 1996. The model calibration process employed in this study is outlined in Figure 4 and described in greater detail below.

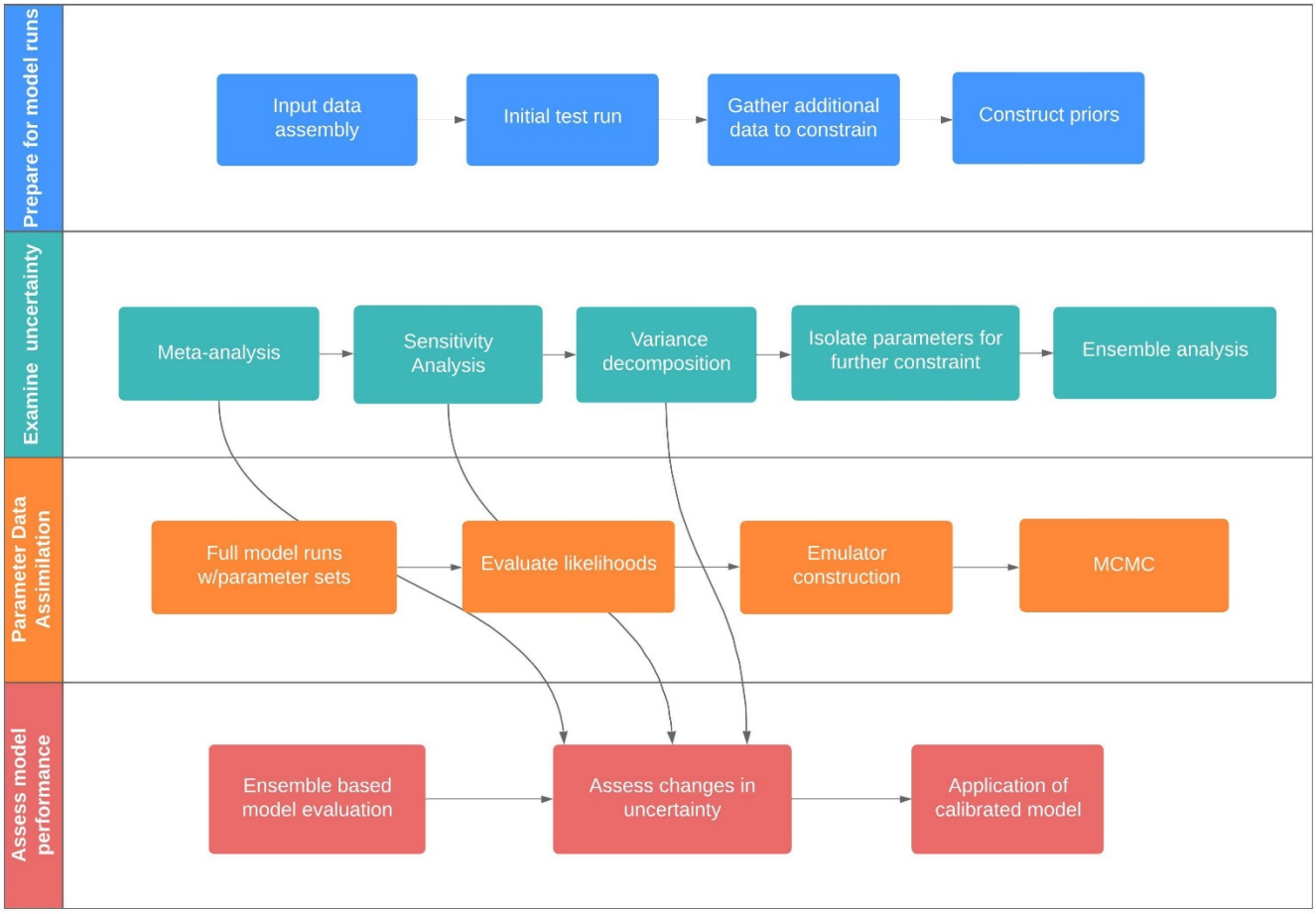

Figure 4: *Schematic of model calibration workflow*

2.2 Site description

The study area is located in the Chequamegon-Nicolet National Forest in Northern Wisconsin, near the town of Park Falls. Most of the region is heavily forested and trees are primarily deciduous but a significant conifer presence exists as well. There is a high degree of heterogeneity representative of a typical mid-latitude forest, displaying a diverse array of wetlands, meadows, streams, and lakes in addition to forest cover (USDA Forest Service, 2011). Heterogeneity is further accentuated by a long history of non-uniform forest management practices including thinning and clear-cuts, resulting in increased variability in stand age and structure. The

majority of the old growth in the forest was logged in the mid-19th to early 20th century (Desai et al., 2007), and a portion of the area was replanted by the Civilian Conservation Corps in the 1930's. Typical homogenous patches of landcover are generally around 20 hectares or less (Desai et al., 2015).

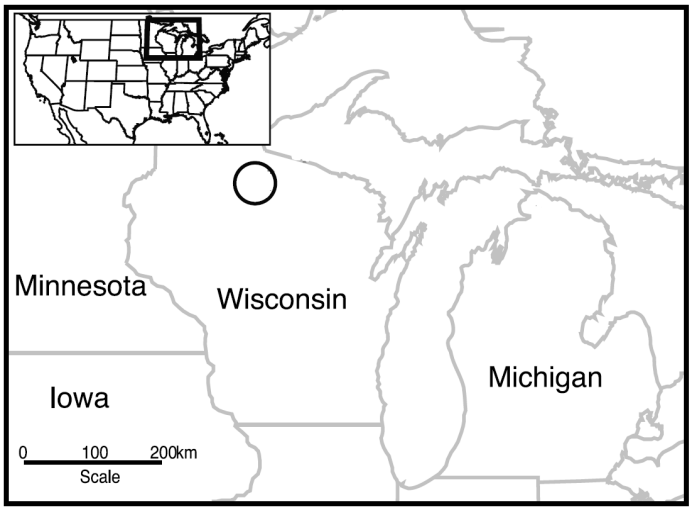

Figure 5: *Map depicting the location of the study site within a regional context. The black circle depicts a 60km radius around the location of the Park Falls, Wisconsin WLEF tall tower (adapted from Desai et al., 2010).*

The area is of relatively consistent low-grade elevation and has a low density of human population. Slight variations in terrain elevation in combination with significant precipitation in all seasons results in a mix of saturated (wetland) and unsaturated (upland) sandy loam soils (Davis et al., 2003). Upland forests comprise roughly 65% of the landscape (CHEESEHEAD, 2018) and deciduous tree types include aspen, sugar maple, red maple, basswood, paper, yellow, and black birch, as well as beech and several varieties of oak. Coniferous tree varieties include balsam fir, red, white, and jack pine, and white spruce. Wetlands are both forested and unforested and account for approximately 35% of the landcover. Wetland tree species include alder, cedar, tamarack, and

black spruce (Davis et al., 2003). The area experiences a humid continental climate with warm humid summers and cold snowy winters.

2.3 Model description

The Ecosystem Demography model version 2.2 (Moorcroft et al., 2001, Medvigy et al., 2009, Longo et al., 2019) is the model of choice for this study. The Ecosystem Demography model (ED2.2) is a size and age structured approximation of an individual-based vegetation model, often referred to as forest gap models. In forest gap models, ecosystem descriptors such as carbon pool magnitudes and net ecosystem productivity are emergent properties that result from simulated competition for resources between plant types with differing abilities to survive (Longo et al., 2019). This allows for the incorporation of heterogeneity in plant communities, which supports direct comparison with field measurements, and realistic spatial and temporal scaling. ED2.2 divides plant communities by plant functional type (PFT), with each group having distinct physiological parameters describing how they function within the larger ecosystem.

ED2.2 requires input drivers that are specific to the geographic location where the model will be run. Required drivers include meteorological data, soil characteristics, and detailed plant community information (PFT's, species diversity, individual tree diameters, etc.). Due to its hierarchical spatial structure, ED2.2 is a highly scalable model, and can be run at every level from the footprint of a single ECF tower to the entire continental United States. ED2.2 is spatially organized into grids, polygons, sites, patches, and cohorts. Grids are the regional or largescale areas of geographic interest. Polygons are domains of interest within which the large-scale meteorological forcing above the canopy can be assumed uniform (Longo et al., 2019). The physical size of the polygon depends on the resolution of the meteorological drivers, but is roughly considered to be equivalent to one grid cell in an atmospheric model (1^o x 1^o or ~100 km). Each

polygon is subdivided into one or more sites, with each site representing a region of shared abiotic physical properties (soil texture, elevation, etc.). Within each site, disturbance related heterogeneity is represented by a series of patches, which are defined by the time since the last disturbance and the type of disturbance that caused them. Each patch represents a collection of tree canopy gap sized areas (~10m) within a given site. Within patches, cohorts are groupings of similar size, age, and plant functional type characteristics (Longo et al., 2019). To account for fine-scale variability within the landscape, the energy, water, and carbon cycles are solved for each patch separately, and fluxes and storage terms are solved for each cohort within a given patch. Patches are not physically contiguous and don't exchange information or material with each other, although they can exchange heat and mass with the shared atmosphere above the patch-specific canopy airspace.

Within ED2.2, thermodynamic properties (including eddy fluxes) are scalable with mass, and the model is constructed such that when biomass changes due to plant growth or mortality, thermodynamic properties are also updated to reflect changes in heat and water holding capacity. Water and energy budget equations for vegetation are solved at the individual level and the corresponding equations for environments shared by plants such as soils and canopy air space are solved for each micro-environment in the landscape (Longo et al., 2019). ED2.2 treats the movement of carbon dioxide (CO_2) as a subset of the full carbon cycle, and CO_2 storage in the canopy air space is independently resolved. The remainder of the carbon cycle is captured as a virtual pool of accumulated carbon representing the net carbon balance for a patch. The accumulated carbon pool links short term carbon cycle components such as photosynthesis and respiration (both heterotrophic and autotrophic) with long-term carbon cycle components related to plant growth, reproduction, and mortality (Longo et al., 2019).

ED2.2 was the model chosen for this study due to its realistic representation of ecosystem processes, its ability to capture successional changes in ecosystem structure and function over time, ease of spatial scaling, and model outputs that aligned with primary variables of scientific interest. ED2.2 also resolves many fluxes at a sub-daily timestep, allowing for the examination of short-term physiological responses to environmental conditions in addition to long-term trends (LeBauer et al., 2013). The primary output variable of interest in this study was net ecosystem exchange of CO_2, a key variable related to the terrestrial carbon cycle. Additional ED2.2 outputs are shown in Table 1.

Table 1: *ED2.2 final output predictive variables*

Above ground woody biomass	Near surface CO2 concentration	Subsurface runoff
Absorbed fraction incoming PAR	Near surface module of wind	Surface incident longwave radiation
Active layer thickness	Near surface specific humidity	Surface incident shortwave radiation
Albedo	Net ecosystem exchange	Surface pressure
Autotrophic respiration	Net longwave radiation	Surface runoff
Average layer soil moisture	Net primary productivity	Time
Average layer soil temperature	Net shortwave radiation	Total evaporation
CO2CAS	Rainfall rate	Total living biomass
Crop yield **if crop PFT's chosen*	Root moisture	Total respiration
Frozen thickness	Sensible heat	Total snow depth
Gross primary productivity	Size of respective carbon pools	Total soil carbon
Ground heat	SMFrozenFrac	Total soil wetness
Heterotrophic respiration	SMLiqFrac	Transpiration
History time interval endpoints	Snow fraction	Vegetation temperature
Latent heat	Snow temperature	Water table depth
Leaf area index	Snow water equivalent	
Near surface air temperature	Soil respiration	

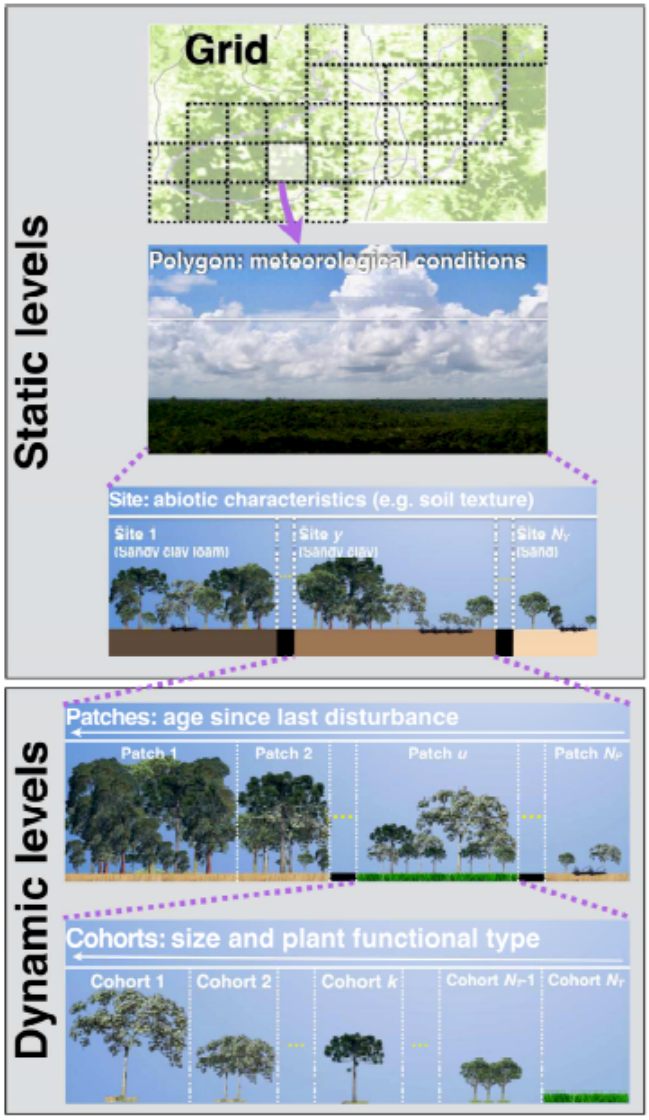

Figure 6: *Schematic representation of ED2.2's hierarchical spatial structure (Longo et al., 2019)*

2.4 The Predictive Ecosystem Analyzer as a workflow tool

Model runs and calibration were performed using the Predictive Ecosystem Analyzer (PEcAn). PEcAn is an open-source ecoinformatics workflow designed to make modeling and the characterization of uncertainty associated with models accessible to the broader scientific community (Lebauer et al., 2013). A foundational goal of the project is the acceleration of scientific progress through serving as a tool for the identification of knowledge gaps and

supporting targeted future research (Lebauer et al., 2013). PEcAn contains a variety of models in addition to the Ecosystem Demography model version 2.2, the model chosen for this study. Other models supported by Pecan at present include Community Land Model (CLM), Simplified Photosynthesis-EvapoTranspiration model (SIPNET), Data Assimilation Linked Ecosystem Carbon Model (DALEC), PREdict with Large Eddy Simulations (PRELES), Multiple-array Assimilation Evaporation Soil Plant Atmosphere model (MAESPA), Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS), LINKAGES, Functionally-Assembled Terrestrial Ecosystem Simulator (FATES), and BioCro.

In addition to providing a platform to run models, PEcAn contains modules to assist in data processing, model-data synthesis, and error propagation and analysis. A primary focus of PEcAn is error characterization, and in support of this the framework accomplishes two goals: (1) data synthesis and error propagation through an ecosystem model and (2) the attachment of information value to subsequent data collection to describe its ability to reduce uncertainty (Lebauer et al., 2013). PEcAn modules used in this study include model-specific input data formatting, meta-analysis, sensitivity, uncertainty, and ensemble analysis, and parameter data assimilation.

a. Integrated plant trait database

Pecan interacts with a plant trait database called the Biofuel Ecophysiological Traits and Yields Database (BETYdb) to query prior distributions and other trait data for parameters of interest. Trait data are organized by PFT in BETYdb, where each PFT has a suite of parameters and their associated distributions, citations, and experimental details as well as a list of plant species that fall within that PFT category. PFT's are broad groupings of plant species based on shared structural, phenological, and physiological traits, and are often further sub-divided by climate zone (Bonan et al., 2002). An example of a PFT used in this study is 'temperate late

hardwood', which describes a late successional hardwood tree species in an eastern temperate forest, with a default archetype of poplar or birch. PFT's are used to represent variety in vegetation types in land surface models, and break a landscape down into a mosaic of patches with common physiological properties. Representing a landscape in this way retains a degree of species diversity while allowing for climate and ecosystem processes to be successfully linked in models (Bonan et al., 2002). The motivation for associating a plant trait database with Pecan is that in order to implement a Bayesian approach to parameter estimation and uncertainty characterization, the Pecan workflow must have access to informative trait priors. 'Priors' are information known about a parameter from previous studies condensed into a representative probability distribution (Spiegelhalter and Rice, 2009), a concept that will be explained at greater depth in section 2.6.

2.5 Bayesian approach to parameter data assimilation

As stated previously, in structurally complex deterministic models with a large number of flexible parameters, uncertainty related to parameter values is a primary driver of overall model predictive uncertainty (LeBauer et al., 2013). Parameter uncertainty can be reduced by collecting additional data or through the application of PDA. PDA is a way to partition uncertainty by source and is used to reduce parameter uncertainty by updating parameter probability distributions to yield both best estimates for model parameters as well as their associated uncertainties, and propagate those uncertainties into model predictions.

The importance of incorporating uncertainty statements in model predictions is what makes a Bayesian approach ideal for DA. A Bayesian approach allows for uncertainty from different sources to be propagated through the PDA process and attached to final predictions, for parameter distributions to be iteratively updated as new information becomes available, and it facilitates model calibration with multiple data streams. The use of multiple data streams allows for the

combination of data of different types and timescales (eddy covariance, satellite-derived, direct biomass measurements, etc.), which has been shown to better constrain parameters and result in more accurate predictions (MacBean et al., 2016). The term 'Bayesian approach' refers to the application of Bayesian statistical methods (centered around Bayes theorem) to make predictions and represent the uncertainty associated with those predictions. A key element of Bayesian statistical methods is that 'known' information can continually be updated as new information is uncovered. Bayesian methods start with existing prior beliefs, referred to as 'priors'; in the context of a model parameter this could be the value range it is believed that a parameter could fall within, drawn from existing literature. Priors for a given parameter are then updated as new data is made available through experimental trials or PDA. This updated version of the priors from PDA are referred to as the 'posteriors', which are then used as priors in subsequent model runs (Van Oijen et al., 2017).

Bayes theorem itself represents a conditional probability, and provides the posterior distribution mentioned above. Given two random quantities y and θ, where *y* represents data and θ represents parameters in a statistical model, Bayes theorem can be stated as shown in Equation 3.

Equation 3: *Bayes theorem, a theorem from probability theory and statistics describing the probability of an occurrence, based on prior knowledge related to the occurrence*

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$p(\theta)$ represents the prior distribution for a parameter, or what is thought about the parameter before looking at the dataset in question. Priors and posteriors are represented as probability distributions not as single values, so they each represent a continuous range of possible values a parameter could

take, situated between a minimum and a maximum. The probability distribution is described by a probability density function, which provides the probabilities of occurrence for different potential outcomes. The value of the function at any given sample in the distribution space represents the relative likelihood that the value of the random variable would equal that sample. The prior distribution is combined with $p(y|\theta)$, the likelihood, to generate $p(\theta|y)$, the posterior distribution. The likelihood function is maximized and used to estimate the parameters of a probability distribution so that the data that's actually observed is represented as the most probable values by the statistical model (Spiegelhalter and Rice, 2009). The posterior distribution is created by using information about a parameter from observational data (the likelihood function), to update a prior state of beliefs about a parameter to become a posterior state of beliefs about a parameter or set of parameters (Ravenzwaaij et al., 2018), so it can be thought of as a compromise between the likelihood and the prior distribution. In Bayesian model calibration, the initial parameter values to be used to initiate Markov Chain Monte Carlo (MCMC) model runs are sampled from the generated posterior distribution (Dietze et al., 2014). An important distinction to make when applying a Bayesian framework in PEcAn is that each parameter of interest is required to have a prior distribution of its own, but the final output is the joint posterior probability of all parameters, so the posterior probabilities that are generated are not fully independent of each other (The PEcAn Project).

2.6 Basis of emulation for parameter data assimilation

PDA is the application of data assimilation methods to model parameters specifically, as opposed to state variables, as is commonly done in numerical weather prediction. The goal of PDA is to improve the ability to predict ecosystem response, and reduce the degree of uncertainty

surrounding those predictions by reducing the contributions of individual parameter uncertainty to overall predictive uncertainty.

b. The emulator method

Optimizing a model often requires tens of thousands of model runs to land on ideal parameter ranges, which can result in computational constraints if the model of choice requires a large amount of Central Processing Units (CPUs) to run. An emulator is a simplified statistical model used in place of a full model for calibration purposes (Sacks et al., 1989), and is one solution to the computational limitations of model optimization. The emulator used in this study is unique in that it isn't a simplified version of the full model itself, but instead represents the response surface of the full model (Wang et al., 2014), specifically the difference between the model and observation error space (Fer et al., 2018). An emulator is used when the full model is too complex to facilitate the use of data assimilation methods for calibration, as it would be computationally unrealistic.

The more traditional data assimilation method is referred to as 'brute-force', where after a parameter value is proposed the full model is run once, likelihoods and posteriors are evaluated, the DA algorithm of choice proceeds until a probability density function is generated, and then the process repeats itself many times. This technique is significantly slower, as the computationally costly step of running the full model needs to be accomplished separately for each algorithm iteration, whereas in the emulator approach model runs are parallelized (Fer et al., 2018). A visual comparison of brute-force and emulator calibration methods are shown in Figure 7.

For dynamic ecosystem models such as ED2.2, the model's computational demands make a traditional 'brute-force' calibration method impossible (Fer et al., 2018), as it involves running

the full model a large number of times (upwards of 50,000) and evaluating output. Such a large number of runs is computationally expensive for even simple models but prohibitive for complex models, where such a high number of runs could realistically take years of wall time.

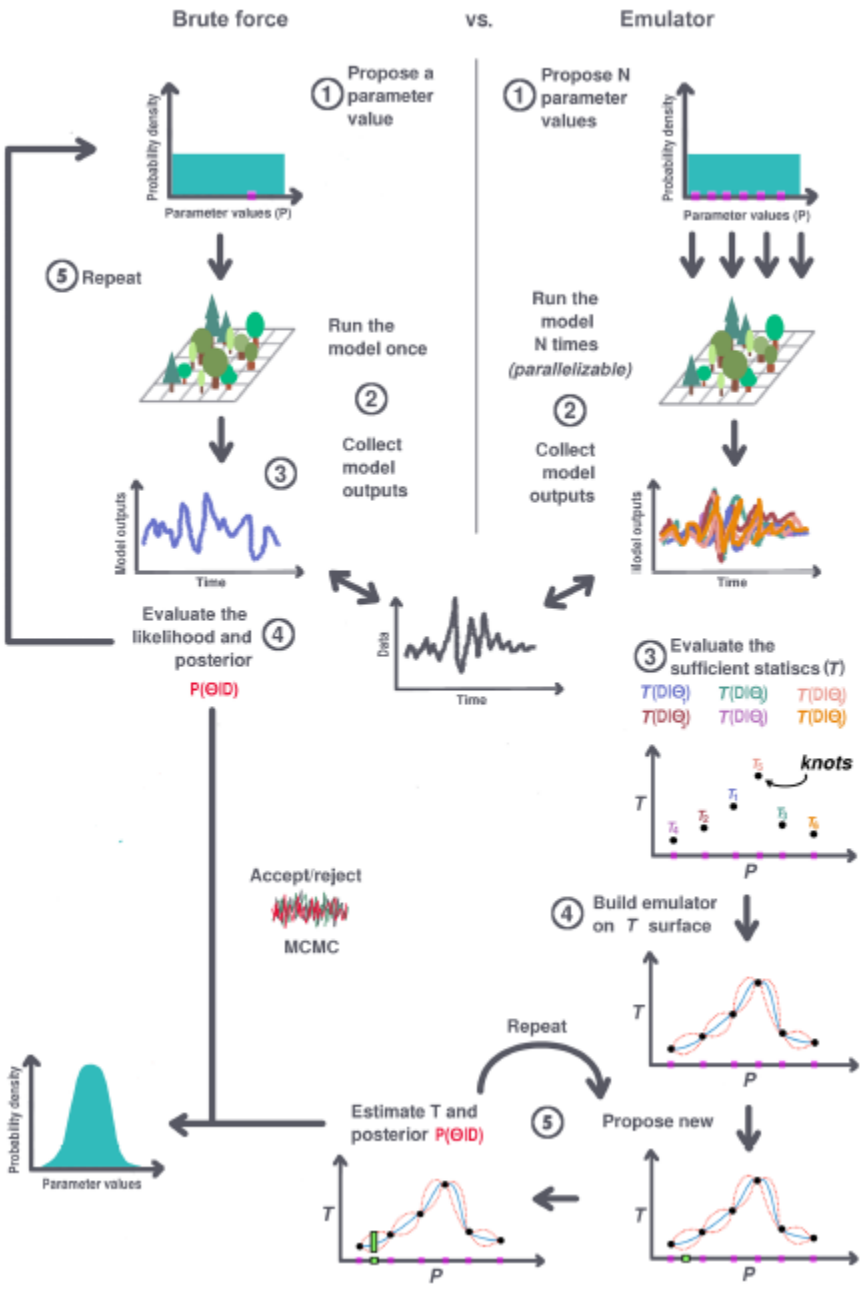

Figure 7: *Comparison of brute-force and emulator approaches from Fer, 2018. The axis T represents sufficient statistics values on the y-axis, and P represents parameter values on the x-axis.*

PEcAn can support both brute-force and emulator approaches to model calibration, but due to the computational complexity of ED2.2, an emulator approach was used in this study. In the emulator workflow, a set of 'N' parameter vectors are chosen from a given parameter's distribution, and the full model is run with this set of initial parameter values. Model output from each run is compared to observational data for a variable of interest (such as NEE), and a sufficient statistic representing deviations of predictions from observations is calculated (shown as 'T' in Figure 7). Note that model error (T) is calculated instead of the likelihood itself. Next, a simplified statistical model is fit through the sufficient statistic points (referred to as knots) that are generated through evaluating the full model. In this study, a Gaussian process model was used, as it has been shown to be the ideal model form for emulator construction (Wang et al., 2014). The statistical model acts as a response surface, describing how model error varies across a given parameter distribution, and interpolating model output in the parameter space between where the full model was actually run. The statistical approximation itself is known as the emulator, and the process of fitting it to the generated knots is known as 'building' the emulator on the statistical surface. The emulator building step includes error estimates around the constructed emulator.

Once the emulator is built, the full model no longer needs to be run and the emulator can instead be used in place of the full model in subsequent PDA steps, such as MCMC. New parameter values can be proposed, and the emulator estimates possible values of the response variable in a fraction of the time that would be required for the full model to do so. The actual construction of the emulator can be time consuming, as the full model must be run many times,

but since model runs are independent of each other computational time can also be saved here by parallelizing these construction runs, which is the standard in PEcAn. Since the emulator is being used here in the Bayesian sense, the statistical approximation being fit is really a likelihood function. In PEcAn, the actual likelihood itself isn't calculated, instead the emulator calculates the statistics of the likelihood, which contain all the information to calculate the likelihood without actually doing so. This allows for the statistical data behind a given likelihood to be retained, so important information about a likelihood such as its error statistics can still be calculated (Fer et al., 2018). One downside of this approach is that because a sufficient statistic is calculated instead of an overall likelihood, if you wish to calibrate with respect to more than one variable, an emulator must be constructed for each variable of interest.

The number of parameter vectors and thus knots that are proposed is an important choice when setting up the emulator. Using too few initial parameter vectors will slow convergence to optimal function values. Opting for more knots means the parameter space will be more thoroughly sampled, resulting in less approximation error. However, more knots also mean more computation time and expense, potentially reducing efficiency (Wang, 2014). Choosing where to place knots is another complex problem, as picking random locations in parameter space could result in uninformative points, for example points that are very close together, or that lie in the tails far from the mean of a parameter's distribution. PEcAn utilizes a Latin hypercube (LHC) design to choose knot locations, where the number of desired knots is user prescribed, but the location of knots is generated from a sequence of values for each parameter. LHC is an extension of stratified sampling, and ensures that the full range of a parameter's distribution is sampled (Sacks, 1989). LHC's are an attractive option because they are very computationally cheap to generate, and the sampling design is more suited to the objective of understanding how a known distribution of

initial parameter values propagates through to an output distribution (Sacks, 1989). Each sequence represents values across the quantiles of the prior distribution for a given parameter, and is randomly permuted independent of the other sequences to construct the overall design matrix (Fer et al., 2018).

In this study, emulator-based calibrations were run with three MCMC chains constituting 25,000 iterations. 230 total knots were generated, corresponding to roughly 16 knots for each parameter chosen, following the advice of PEcAn developers to use at least 10 knots per parameter. More information about chosen parameters is contained in section 3. For NEE focused PDA the full model runs required to inform emulator construction took approximately seven days, Gaussian process model fitting took 7.9 minutes, and the emulator took 109.84 (~4.5 days) hours to complete 25,000 iterations.

2.7 Process of model calibration

c. Data preparation

As presented in Figure 4, the first step of model calibration is input data assembly. Data required by the model as drivers must be assembled, quality checked, gap-filled if needed, and restructured into the format required by the model. For ED2.2, this includes meteorological driver data from eddy covariance flux towers, vegetation initial state files, and soil characteristics data. Meteorological drivers include temperature, specific humidity, CO_2 molar fraction, air pressure above the patch canopy, precipitation, wind speed, and shortwave and longwave radiation (Longo et al., 2019). Meteorological data are collected at a variety of timescales (in accordance with AmerifluxLBL standards), but all are sub-daily to facilitate accurate simulation of the diurnal cycle.

Vegetation initial state files describe forest composition and structure during a snapshot in time, generally the start year for a model simulation if data is available. These files come from the Forest Inventory and Analysis (FIA) dataset, a 'forest census' product compiled by the USDA Forest Service to document the health, structure, and distribution of forests across the United States, with inventories taking place as early as 1930. These are separated into files corresponding to the cohort and patch levels. Cohort files contain information about the year, patch number, cohort number, tree diameter at breast height, and PFT. Patch files contain site number, year, patch number, age, the percentage of the cohort area that the patch occupies, in addition to information about soil carbon and nitrogen stocks (both fast and slow). Site level files include information about elevation and soil type. Soil characteristics drivers include soil texture, number of soil layers, bottom boundary conditions (bedrock, free or reduced drainage, water table), and soil color. For ED2.2 in PEcAn soil texture is user-prescribed, provided as percentage of clay and percentage of sand, but other soil-related drivers come from the linked Food and Agriculture Organization of the United Nations (FAO) database. Once data is assembled an initial test run of the model is done to ensure settings are functional and there are no major discrepancies between the modeled output and observed data. Since model calibration in PEcAn utilizes a Bayesian approach, informative priors must be constructed for all flexible model parameters.

Before the process of PDA can begin, influential parameters are chosen from the suite of available flexible model parameters to target for PDA. ED2.2 has a large number of flexible parameters that vary depending on plant functional type, and a single PFT usually has ~20 parameters. PDA is computationally expensive for a model as complex as ED2.2, so only a handful of parameters can be chosen to target. Given this constraint, it is ideal to choose parameters that will be most impactful in reducing overall uncertainty and improving model fit (Fer et al., 2018).

A parameter can be impactful if it is either poorly constrained (uncertain) or if the model is highly sensitive to it, and both parameter uncertainty and sensitivity contribute equally to model predictive uncertainty (Dietze, 2017). Generally speaking, a parameter must be both sensitive and uncertain to some degree to be considered a good target for PDA. To assist in systematically analyzing and representing model component uncertainty as well as in the selection of high-impact parameters for PDA, PEcAn has an automated workflow consisting of three steps: meta-analysis, parameter sensitivity analysis, and variance decomposition analysis.

d. Uncertainty examination

During meta-analysis observational trait data such as prior probability distributions, sample means, error statistics, and study metadata are extracted from PEcAn's companion trait database BETYdb for each parameter within each plant functional type (Dietze et al., 2014). With this data, a Bayesian met-analytic model is fit for each trait within each PFT, and is used to generate a posterior probability distribution for the mean of each flexible model parameter. Using a hierarchical meta-analysis approach to construct the posterior probability distribution allows the distribution to come from the synthesis of information across multiple laboratory and observational studies, as opposed to a single study estimate (Dietze et al., 2014). As research proceeds and more data about parameter properties are made available, the meta-analytic model can be re-fit, and the posterior distributions updated to reflect the current scientific knowledge base. The posteriors generated through meta-analysis return later in the workflow as priors to be used in data assimilation.

The process of sensitivity analysis is used to determine which parameters the model is most sensitive to, meaning which parameters result in the largest change in model output when they are represented with slightly different values. Parameter sensitivity analysis involves repeating a

prescribed number of model runs (ensemble of size *n*) with different parameter perturbations in order to assess how slight changes in parameter values will affect specified model outputs. In PEcAn, parameter values are varied according to the quantiles of the posterior parameter distributions generated through meta-analysis, with a user-prescribed standard deviation from the mean value of their respective density functions (LeBauer et al., 2013). For this study, each parameter was varied $\pm$ 1 and 2 standard deviations (σ) from the mean value, resulting in five variations on a given parameter value for each run: -2σ, -1σ, 0, 1σ, 2σ, where 0 represents the mean value. A response function describing the given model output variable as a function of each parameter value is then generated (Dietze et al., 2014).

Variance decomposition combines information about both the uncertainty and sensitivity of parameters to estimate how much each input parameter contributes to overall uncertainty in model predictions (LeBauer et al., 2013). Variance decomposition is done by transforming the posterior parameter distribution generated through meta-analysis with the sensitivity function from parameter sensitivity analysis (Dietze et al., 2014). Variance decomposition output in PEcAn involves two variables: the coefficient of variation (CV), and the elasticity. The coefficient of variation (also known as the relative standard deviation) is a measure of the dispersion of a probability or frequency distribution. In other words, it's the uncertainty associated with each parameter, where a higher CV represents a higher level of uncertainty. The elasticity of a parameter is the sensitivity of the model output to each parameter normalized by both the parameter and variable output means. For example, an elasticity value of one means that a one percent change in the input variable results in a one percent change in the output variable, so higher elasticity values mean a parameter has a disproportionate impact on the model. Elasticity and CV are both taken into account to generate the partial variance for a given parameter. Partial variance is used to look

at the overall predictive uncertainty because it's normalized across sensitivity and parameter variance so units can be disregarded and uncertainty is expressed as a percentage, making it easier to compare across parameters with a wide range of units.

Once target parameters have been chosen through variance decomposition, an ensemble of size *n* is run, with each run sampling values from a given parameter's uncertainty distribution to create a probability distribution of model projections. Ensemble analysis is used to place a confidence interval on the model, which is important when interpreting model output, and ensuring that the model is making predictions that are within the range of reality. As a final feasibility check before proceeding to PDA, the degree of uncertainty in model predictions is compared to what is observed in reality, and it's ensured that variables have been constrained with additional data to reduce uncertainty where possible.

e. Parameter data assimilation

Once PDA officially begins, the first step is to propose initial parameter value sets. The values come from prior information about the chosen parameters drawn from meta-analysis. Once initial parameter values have been proposed, the full model is run for each parameter set, using initial vegetation files and meteorological data as drivers. This step in the process if often time consuming, as with a complex model like ED2.2 the full model runs step can take several days to a week or more depending on computing power. Once full model runs are complete, likelihoods are evaluated for each parameter set. The full model output is compared to observational data and the likelihood is calculated using a heteroskedastic Laplacian, which allows variance to change with flux magnitude.

Following likelihood calculation, the next step is emulator construction, described previously. The emulator is constructed on the likelihood surface and interpolates model output in parameter space between the points where the model has actually been run. After the emulator has been built, it's passed along to the sampling strategy and assessment algorithm in place of the full model, and posterior probability density functions are estimated for each parameter in question. There are a variety of sampling and assessment algorithms that can be used in PDA, but the approach used in this study was MCMC sampling with an adaptive Metropolis-Hastings algorithm. MCMC is a fundamental element of the PDA approach applied in this study and is a specific type of Monte Carlo method, which are often used in Bayesian statistics to make inferences using data. Monte Carlo methods take a given probability distribution and use a computer-generated sample set from that distribution to calculate a 'plug-in estimate' for a feature of interest of the given distribution, where the feature of interest might be the mean or the variability of the distribution. The calculated plug-in estimate is called the 'Monte Carlo approximation' of the true distribution function (Taboga, 2017). Plug-in estimates are applied in probability theory related statistics to estimate features of the probability distribution when it's difficult or impossible to precisely compute the feature itself. Essentially the plug-in principle proposes that a difficult to compute feature of a given distribution can be approximated using the same feature of an empirical distribution created from samples drawn from the given distribution (Taboga, 2017). An empirical distribution describes a sample of observations of a variable of interest. To actually use this method to approximate the value of a given variable, multiple computer-generated sample sets would be pulled from the distribution in question, plug-in estimates would be generated for each generated sample, and the final value for the feature in question would be the mean value of that feature from the computer-generated samples of observations.

The key distinction between MCMC and classic Monte Carlo is that MCMC uses these techniques to generate sample sequences of dependent observations, whereas classic Monte Carlo relies on samples made up of independent observations. The sample sequences of dependent observations that are generated are known as Markov Chains. The chain of values represents estimates from each step in a quasi-random walk through a given parameter's target distribution. Each chain is initialized with starting values for each of the parameters isolated for PDA. Eventually the walk-through parameter space converges to the target distribution, so the accepted parameters approximate the true parameter distribution and the model is said to have 'converged' and is a good representation of reality (Taboga, 2017).

MCMC begins with a single guess for the starting value of a parameter. This value is chosen at random from the posterior distribution of the parameter of interest. In PEcAn, the posterior distribution values are chosen from is the distribution generated for a given parameter during meta-analysis. From the starting value, a jump is proposed to move from that position to another location in the chosen parameter's posterior distribution. In PEcAn the jump and landing location is proposed by an adaptive Metropolis-Hastings algorithm, which pulls a sample from a normal distribution centered around the current parameter position, with a prescribed standard deviation that sets the distance of the jump (Fer et al., 2018). Before actually making the leap, the suitability of the proposed jump location must be assessed. If the resulting distribution from the proposed position explains the data better than the old position, then it's a good jump to make. How well the new position explains the data is calculated by computing the probability of seeing that value given the observed data, and computing the likelihood. The likelihood corresponding to each new jump location is approximated by the emulator and the new value is accepted or rejected based on its posterior probability relative to the current parameter value. If the proposed value is

accepted, the jump takes place and it becomes the next value in the chain, otherwise the jump to that location doesn't take place and the next sample in the chain is just a copy of the current sample (Sherlock et al., 2010).

One way to interpret good versus bad jump locations is through the idea of "uphill" and "downhill" proposals. Uphill proposals are proposed jump locations that take the chain closer to a local mode, and are always accepted, whereas downhill proposals are accepted or rejected with probability exactly equal to the relative "heights" of the posterior at the proposed and current values. The more frequent rejection of downhill proposals is what keeps the Markov chain generally in the main posterior of the distribution (Ravenzwaaij et al., 2018). This process represents one iteration of MCMC, and is repeated until the posterior distribution has been thoroughly explored and enough samples are generated for the walk to converge to the target distribution.

A key issue with MCMC is that since the starting value for a parameter is chosen at random, even though it is pulled from the posterior distribution it might not necessarily be representative of that distribution. For example, if the starting parameter values for a chain are located near the mode of the true distribution, then they're representative of the majority of the data, MCMC output can be trusted, and the chain is representative. However, if the starting parameter values are from the tails or represent more extreme values far from the true distribution of the data, then more iterations are required before the chain approaches the sample mean. The portion of the chain that represents the iterations prior to approaching the true sample mean is known as the 'burn-in' period, and is generally removed from the final output to prevent an incorrect reflection of the updated posterior distribution. Using multiple chains is one way to ensure that MCMC convergence output is reliable. Each chain is initialized with different starting parameter values

sampled from the distribution, and runs through *n* iterations. In this study, three chains were used to ensure convergence, running through approximately 25,000 MCMC iterations each.

So how can you tell if chains have actually converged? MCMC convergence is initially assessed visually, followed by statistical diagnostics if a more objective method for assessing burn-in is desired. A common statistic to determine whether or not chains have converged is the Gelman-Brooks-Rubin statistic, $\widehat{R}$, which compares the variance within each chain to the variance across all chains (Gelman and Rubin, 1992). In PEcAn, this is how burn-in periods are determined. If the chains have converged perfectly then $\widehat{R}$ should equal 1. Values less than 1.1 are typically considered sufficient, but anything larger indicates a lack of chain convergence. When running a Gelman diagnostic on the MCMC chains produced in this study, a threshold value of 1.1 was used.

f. Model performance assessment

If the emulator is being run in rounds, the PDA process cycles back to the beginning and re-proposes a percentage of the parameter value sets from the posterior of the previous round and a percentage from the initial prior distribution before running through the whole process again. At the end of PDA, a final posterior distribution is generated for the parameter set that was initially proposed. This updated posterior distribution represents the combination of prior knowledge and observational data to provide a better estimate of a parameters true value range. This updated posterior distribution is then used as the parameter's prior distribution when the model is subsequently run, and variance decomposition is repeated to asses changes in parameter uncertainty and overall model fit as a result of PDA. The calibrated version of the model can then be run as an ensemble against the data used to do the calibrating, additional data at the same location (such as switching the variable you're interested in predicting), and new data from a different location. Running the model against data not used for calibration ensures that the

calibrated model is actually representative, and not overly fit to the training dataset (Richardson et al., 2010). Model performance is assessed through statistical analysis, and it's determined if any of the previous calibration steps need to be revisited.

2.8 Statistical analysis

Statistical analysis performed in this study can be segregated as it relates to three primary goals:

(1) Determine if the PDA process was successful
(2) Quantify the degree of uncertainty reduction in individual parameters
(3) Gauge how overall model predictive uncertainty changed as a result of reducing parameter uncertainty through PDA

(1) Determining PDA success

PDA success is first determined visually using trace diagrams, marginal density plots, and Gelman plots. Visual assessment is then confirmed using the Gelman-Brooks-Rubin diagnostic statistic to determine chain convergence. Finally, correlations between parameters are assessed. Following the completion of the MCMC portion of PDA, four plots are generated for each targeted parameter within each PFT: trace, marginal density, Gelman, and correlation plots. Trace plots show the sampled values of a given parameter over time, with time represented as model iterations. Trace plots help assess mixing of chains and how quickly the MCMC process converges in its distribution. An 'ideal' trace plot exhibits rapid up-and-down fluctuations with no discernable long-term trends or drifts, with all plotted chains showing similar patterns. Long-term trends or drifts as well as large differences between chains indicate slower convergence or a lack of convergence (Robert and Casella, 1999). Figure 8 depicts successful MCMC chain convergence for the parameter 'respiration optimal water', a soil parameter targeted for PDA. Three chains were

used in this study for each parameter; the trace plot shows three different plotted elements corresponding to each chain represented by the colors green, red, and black.

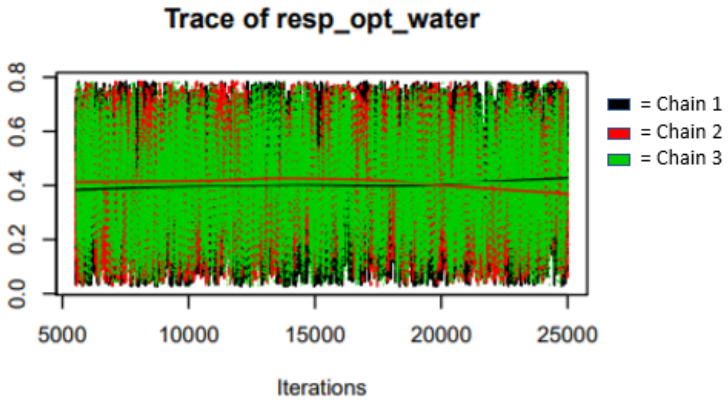

Figure 8: *MCMC trace plot for the soil parameter 'respiration optimal water'. The number of model iterations (a stand in for time) is displayed on the x-axis, and the sampled parameter values are displayed on the y-axis. The wide spread across the y-axis indicates a sufficient number of sampled were used for MCMC, and the consistency in the plots over time as well as similarity between chain patterns indicates convergence.*

Marginal density plots are essentially a smoothed histogram of the values shown in the trace plot, the distribution of the values of a given parameter in the chain. Values are averaged over the values a parameter takes across all model iterations with other parameters marginalized to remove correlations between parameters. A marginal density plot indicates PDA success for a given parameter when the plot is void of any sharp peaks or valleys, as shown in Figure 9. Density plots will often have an approximately normal distribution, while others may be skewed, have clearly defined boundaries, or even on occasion distributions can be multimodal. There is no requirement in Bayesian statistics that the distributions be normal, as long as MCMC has

converged. However, it is worth noting that the most common cause of multi-modal posteriors is a lack of convergence (The PEcAn Project).

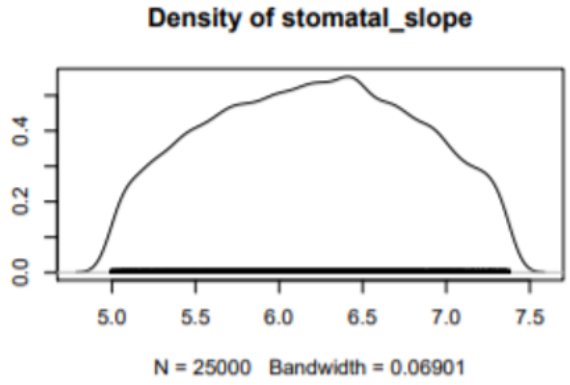

Figure 9: *Marginal density plot for the parameter stomatal slope belonging to PFT Early Hardwood. On the x-axis, 'N' corresponds to the number of MCMC model iterations, and 'bandwidth' is the value used for smoothing (calculated using the minimum of the standard deviation, the interquartile range, and the sample size). What appears as a solid black line on the x-axis is actually a high number of closely placed ticks represented the actual samples used. Average parameter values are displayed on the y-axis.*

The Gelman-Brooks-Rubin diagnostic statistic ($\widehat{R}$) can be calculated to compare the variance within each chain to the variance across all chains and confirm chain convergence (Gelman and Rubin, 1992). A diagnostic statistic (also called a 'shrink factor') of one means between chain variance is approximately equal to within chain variance and chains have converged perfectly, while larger values signify a notable difference between chains. A general rule of thumb is that shrink factor values below ~1.1 signify acceptably low variation between chains, this was the threshold value used in this study. Gelman plots are used to assess how similar MCMC chains

are, as a way to detect if they've collectively approached the target distribution. The Gelman plot shows the development of shrink factors over time, which helps illuminate whether a chain reduction is stable or if the factors fluctuate with time.

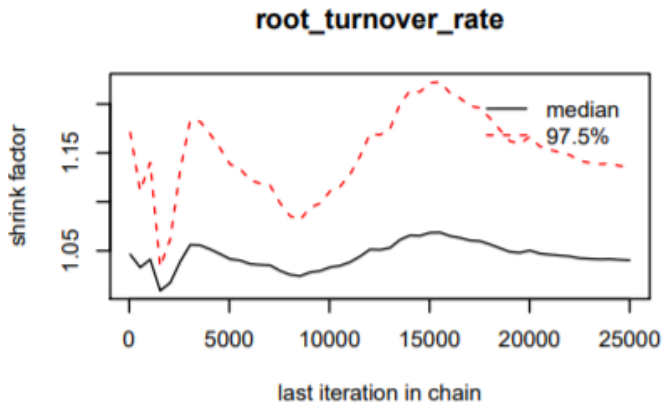

Figure 10: *Gelman plot for the parameter 'root turnover rate' belonging to PFT Late Hardwood. The plot displays a shrink factor roughly around 1.05, indicating acceptably low variation between the three chains used in MCMC. It can be seen that what little fluctuation exists in the shrink factor is primarily near the start of PDA, with fluctuations leveling out over time.*

Correlation matrices are used to show pairwise correlations between parameters targeted for PDA. Evaluating correlation is an important check when assessing PDA results, as strong correlations between parameters can signify that optimization of one parameter could dramatically impact another, skewing PDA results and complicating analysis.

(2) Individual parameter uncertainty

To assess how well uncertainty surrounding individual parameters was reduced through PDA, sensitivity and uncertainty analysis were conducted both before and after PDA, and the change in percent contribution to variance was determined for each parameter on a PFT basis.

(3) *Model predictive uncertainty*

Overall model performance is assessed by comparing metrics from 100-member ensembles run with both pre-and-post PDA settings. Each run is of a single year duration (2008), with parameter values randomly sampled from their respective distributions, such that each ensemble run used slightly different parameter values. Reduction in model uncertainty and thus improvement in model performance is assessed using root mean square error (RMSE) and variance decomposition. Bias in model predictions is often assessed by fitting a linear regression and calculating R^2 and residual error, the difference between a data point and the regression line. However, these metrics are not appropriate to determine model bias in this study, as neither the model nor the predictions are linear. Instead, RMSE should be used in place of residual error for calculating R^2, as it is based on deviations from the 1:1 line obtained by comparing model predictions against observed values and not on deviations from a linear regression line. RMSE can be visualized through a predicted versus observed plot. In a perfect model the data would fall exactly along the 1:1 line. The deviations away from this line are the model residuals. If observations lie along a line other than the 1:1 this indicates that the model is biased in some way (Clayton and Zhu, 2017). Completing sensitivity analysis and variance decomposition for model runs using post-PDA settings allows for comparison of how contributions to model predictive uncertainty from individual parameters were altered as a result of PDA.

3. Results

3.1 Parameter selection

Our first step is to identify suitable parameters for data assimilation, based on how well constrained they are and how sensitive the model is to them. Although sensitivity analysis and variance decomposition are helpful analyses to use in choosing parameters, they should not be the only determining factor. Model structure must also be considered, primarily how relationships between parameters are represented, and if parameters have competing or overly identical roles (Fer et al., 2018). The representation of parameter relationships in a model can also give important insight to understand variance decomposition output. For example, ED2.2 simulates stomatal conductance, an important variable in determining photosynthetic rate, using the Ball-Berry model. In this study, the parameter stomatal slope had the third highest average partial variance (23.506%) of all the flexible parameters, indicating that the model is highly sensitive to stomatal slope and that it causes a significant portion of the overall variance in model output. Stomatal slope is a primary driver of stomatal conductance (Dietze et al., 2014). Cuticular conductance also had a high average partial variance, and was one of the parameters chosen for PDA. Cuticular conductance again plays a key role in the Ball-Berry model (Duursma et al., 2018). Both of these parameters ranking as ideal targets for PDA demonstrates not only the importance of stomatal conductance and photosynthetic rate in the structure of the model, which is intuitive when looking at carbon or water flux through an ecosystem, but also highlights the relatively high degree of uncertainty contributed to the overall model by this photosynthesis-related sub-model.

In this study uncertainty analysis was conducted twice: once to assess uncertainty and determine which parameters to target for PDA then again post-PDA to assess changes in parameter

uncertainty and determine if PDA was able to reduce overall model predictive uncertainty through the reduction of uncertainty contributions from individual parameters. Uncertainty analysis runs will now be referenced as 'pre-PDA' or 'post-PDA' runs.

Table 2: *Calibrated ED2.2 Parameters (predictive variable = NEE)*

Parameter	Category	Description	Units
Growth respiration factor	Plant phys.	Proportion of daily carbon gain lost to plant growth respiration	fraction
Stomatal slope	Plant phys.	Slope of relation between stomatal conductance and photosynthetic rate term (A)	ratio
Water conductance	Plant phys.	"Water availability factor", sets a plant's supply of water	m-2 a-1 (kgC root)-1
Cuticular conductance	Plant phys.	Leaf (cuticular) conductance when stomata fully closed	µmolH2Om-2 s-1
Root turnover rate	Plant phys.	Rate of fine root loss (temperature dependent)	1/yr
V_{cmax}	Plant phys.	Maximum rubisco carboxylation capacity	µmolCO2m-2 s-1
Respiration temp. increase	Soil biogeochemical	how rapidly heterotrophic respiration increases with increasing temperature	1/K
Respiration optimal water	Soil biogeochemical	Optimal soil water for heterotrophic respiration as fraction of max possible soil water	dimensionless, range 0-1

Initial uncertainty analysis identified eight ideal parameters to target for PDA, with six related to plant physiological properties and two soil biogeochemical parameters related to decomposition. Parameters chosen for PDA are common across all five plant PFT's with the exception of cuticular conductance, which is unique to the PFT 'temperate.Late_Hardwood' in this study. Plant physiological parameters were growth respiration factor, stomatal slope, water conductance, cuticular conductance, root turnover rate, and V_{cmax}. Soil parameters were respiration temperature increase, and respiration optimal water (see Table 2). Contribution to overall

uncertainty from the listed parameters varied according to PFT, and not all parameters chosen were shown to be influential across all PFT's (Figure 11). In fact, the only parameter shown to contribute to a high degree of uncertainty across all plant PFT's was water conductance. Partial variance was calculated for a given parameter's contribution to uncertainty within each PFT, and then averaged across all PFT's, representing the overall contribution to variance. Parameters were filtered to weed out those with low contributions to model uncertainty. Only parameters shown to contribute greater than 1.1% of overall model uncertainty (averaged partial variance) were chosen for PDA. As shown in Figure 11, growth respiration factor was the most influential vegetation parameter, with a pre-PDA average partial variance of 34.51%, followed by water conductance (26.41%), stomatal slope (23.51%), cuticular conductance (20.40%), root turnover rate (4.02%), and lastly V_{cmax} (3.74%). Descriptions of each parameter chosen for PDA are provided below.

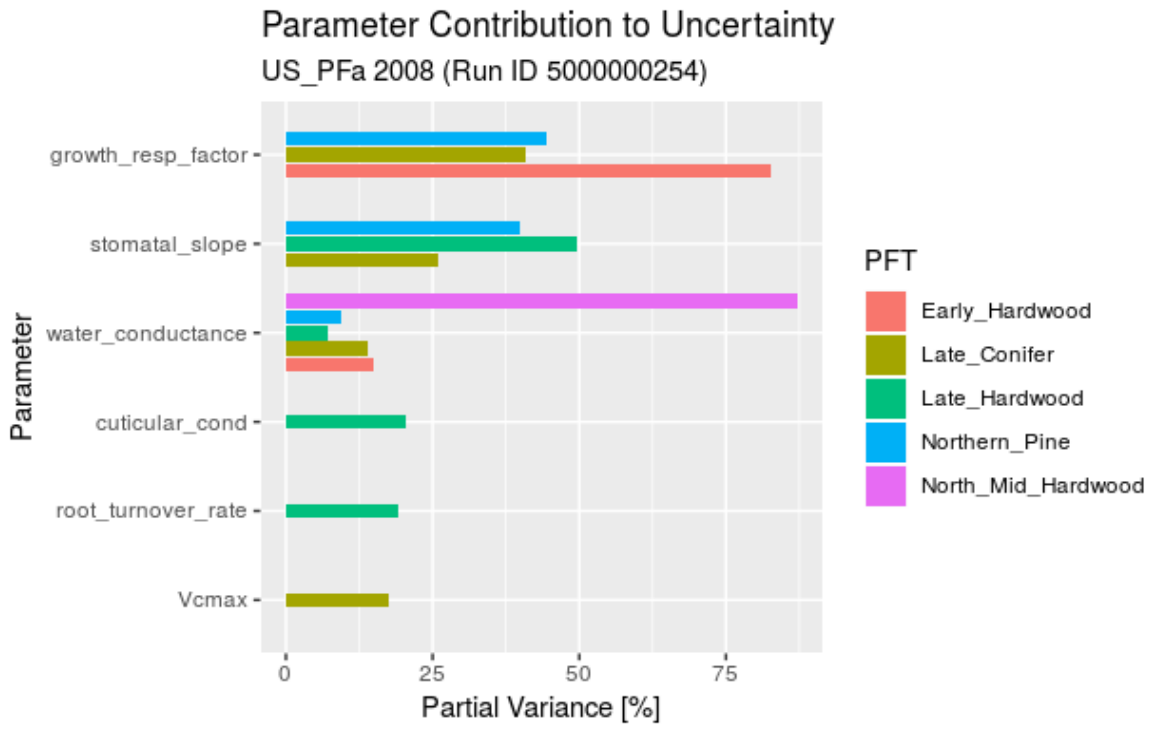

Figure 11: *Results of uncertainty analysis for plant physiological parameters in NEE focused ED2.2 model runs, displayed as partial variance. Impact is displayed by plant functional type, with a longer bar corresponding to a larger contribution to overall model predictive uncertainty.*

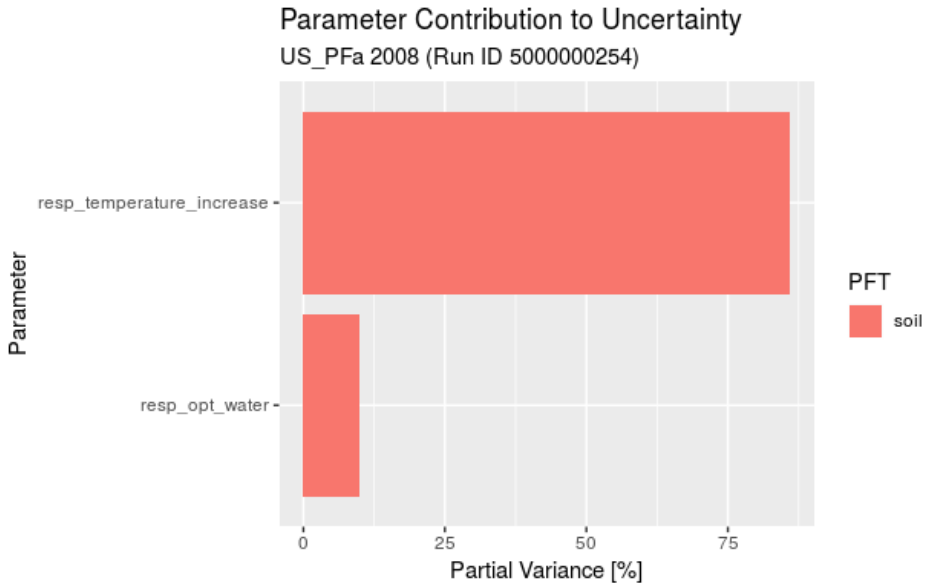

Figure 12: *Results of uncertainty analysis for soil biogeochemical parameters in NEE focused ED2.2 model runs, displayed as partial variance. A longer bar corresponds to a larger contribution to overall model predictive uncertainty.*

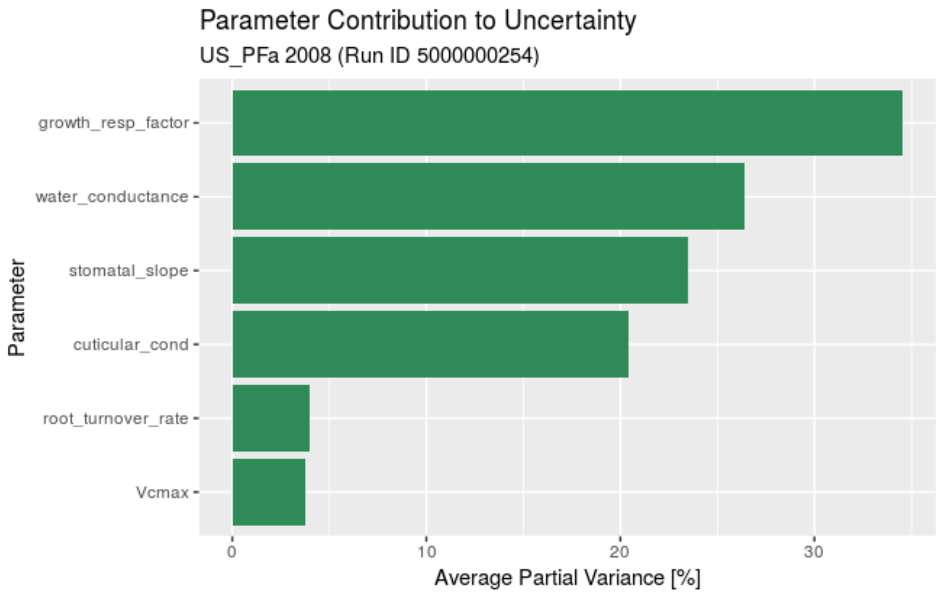

Figure 13: *Results of uncertainty analysis for plant physiological parameters in NEE focused ED2.2 model runs, with partial variance averaged across all plant functional types present.*

Table 3: *Pre-PDA prior distributions used in meta-analysis and model parameterization (variable = NEE). Format is distribution type(a, b). EH = temperate.Early_Hardwood, LC = temperate.Late_Conifer, LH = temperate.Late_Hardwood, NH = temperate.North_Mid_Hardwood, NP = temperate.Northern_Pine*

Plant physiological parameters					
	EH	LC	LH	NH	NP
Growth respiration factor	beta(4.06, 7.22)	beta(2.63, 6.520)	beta(4.06, 7.2000)	beta(2.63, 6.5200)	beta(2.63, 6.5200)
Stomatal slope	uniform(2.00, 16.000)	uniform(2.00, 16.000)	lnorm(1.76, 0.3800)	uniform(2.00, 16.000)	uniform(2.00, 16.000)
Water conductance	lnorm(-5.4, 3.000)	lnorm(-5.40, 3.000)	lnorm(-5.40, 3.000)	lnorm(-5.40, 3.000)	lnorm(-5.40, 3.000)
Root turnover rate	uniform(0.00, 10.000)	uniform(0.00, 10.000)	weibull(1.55, 0.8620)	uniform(0.00, 10.000)	uniform(0.00, 10.000)
Vcmax	uniform(0.00, 500.000)	uniform(0.00, 500.000)	weibull(1.70, 800.0000)	uniform(0.00, 500.000)	uniform(0.00, 500.000)
Cuticular conductance	*NA*	*NA*	lnorm(9.40, 0.70000)	*NA*	*NA*
Soil biogeochemical parameters					
Respiration temp. increase	uniform(0.05, 0.20)				
Respiration optimal water	beta(1.000, 1.00)				

Growth Respiration Factor

Plant cellular respiration is often broken down into contributions from both growth and maintenance respiration. Growth respiration is cellular respiration associated with plant growth processes, such as synthesis of new structures and nutrient uptake. Maintenance respiration is associated with protein and membrane turnover, and maintaining ion concentration gradients (Lotscher et al., 2004). The majority of total plant cellular respiration is associated with

maintenance, with a small fraction remaining for growth and growth respiration (Iersel et al., 2003). The growth respiration factor is a constant fraction of the net assimilation rate associated with biosynthesis of metabolites for plant growth (Dietze et al., 2014). Main controls on growth respiration factor are the associated costs of biosynthesis. Due to the fact that growth respiration factor in ED2.2 integrates a number of respiratory processes, it is challenging to directly measure and few quality observational data exist.

Stomatal Slope

ED2.2 models stomatal conductance using a variant of the classic Ball-Berry model outlined in Leuning et. al 1995, where stomatal conductance is primarily driven by stomatal slope. Stomatal conductance is the rate of CO_2 uptake or water vapor release through the stomata on a leaf's surface. Stomatal conductance is directly related to the boundary layer resistance of a leaf's surface, and thus the concentration gradient of water vapor between the leaf and surrounding atmosphere, making it an integral variable in leaf-level calculations of transpiration (Taiz et al., 1991). Stomatal slope represents the relationship between stomatal conductance and an aggregated term representing photosynthetic rate, CO_2 concentration, and vapor pressure deficit (Dietze et al., 2014). Stomatal slope can be estimated using stomatal response curves from leaf-level gas exchange (Dietze et al., 2014).

Water Conductance

Water conductance controls the supply of water from soil moisture that is available to be taken up by a plant. In ED2.2 water conductance isn't a trait that can be experimentally measured, it's an upper bound value placed on transpiration. ED2.2 calculates water conductance as a function of available soil moisture and root biomass. It can be estimated from eddy covariance

carbon and water flux data, or through sap-flow measurements (Dietze et al., 2014). Water conductance supplies additional controls on stomata opening and closing.

Cuticular Conductance

Cuticular conductance represents the movement of water across the plant cuticle, and is one of two main pathways of minimum leaf conductance (the other being movement through incompletely closed stomata) (Duursma et al., 2018). Minimum leaf conductance is the rate of water loss a plant experiences when stomata are closed, but water loss continues at a diminished rate through the plant cuticle, the wax-like protective film covering the surface of leaves. Cuticular conductance plays an important role in the Ball-Berry model of stomatal conductance, and can be directly measured using leaf-level gas exchange techniques (Duursma et al., 2018).

Root turnover rate

Root turnover rate is the rate at which the continuous cycle of fine root growth and dieback proceeds. Fine roots serve as the interface between plants and soil, and play a crucial role in the cycling of carbon, water, and nutrients within a vegetated ecosystem. Root turnover rate constitutes a significant flux of carbon from aboveground biomass into soils, transitioning carbon into the belowground carbon cycle. Root turnover rate is primarily controlled by the amount of root litter present, which can be used to estimate root turnover, but there is currently no direct and reliable way to measure this parameter (Lukac et al., 2011).

V_{cmax}

V_{cmax} is the maximum rate of carboxylation, and serves as a measure of photosynthetic capacity (Wilson et al., 2000). ED2.2 employs the Farquhar photosynthesis model, an enzyme kinetic model of leaf photosynthesis, to mechanistically reflect plant physiological responses to

atmospheric CO_2 levels (Walker et al., 2014). V_{cmax} is calculated from leaf level gas exchange measurements.

Respiration temperature increase

Respiration temperature increase is a parameter related to soil decomposition, and determines how quickly heterotrophic respiration will increase in response to increasing temperature (Ecosystem Demography Model 2.2 Wiki). Microbial decomposition of soil organic matter has been shown to increase with increases in temperature, but the rate at which it increases is important when simulating ecosystem response to temperature changes. This parameter is heavily influenced by soil moisture, as well as composition of microbial communities (Suseela et al., 2012). The observed correlation between soil respiration, temperature, and moisture is especially prominent in temperate ecosystems, where root processes and aboveground productivity are also closely linked to temperature and moisture levels (Ryan and Law, 2005). Respiration temperature increase can be measured using soil respiration chambers under laboratory or field conditions where temperature can be manipulated.

Respiration optimal water

Respiration optimal water is the optimal soil water content for heterotrophic respiration, as a fraction of the maximum possible available soil water. Heterotrophic respiration rates are strongly linked to soil moisture; decreases in water content below field capacity result in diminished microbial activity, and thus decreased CO_2 efflux. The optimal water content for respiration is often found at intermediate levels, as soil water content can also exceed optimal conditions for soil respiration (Davidson et al., 2000). Respiration optimal water content is

determined using measurements of temperature, water content, and CO_2 production for each soil horizon (Davidson et al., 2000).

3.2 PDA Success

PDA success was evaluated by PFT for each target parameter. PDA was deemed successful for a given parameter if MCMC chains were well mixed, the Gelman-Brooks-Rubin diagnostic statistic showed that variation between the three chains used was below the 1.1 threshold, correlation between parameters was low, and shrink factor values were shown to be stable over time. Both parameters related to soil biogeochemical process had a successful PDA process. MCMC chains were well mixed, marginal density plots were of a normal distribution, correlation between parameters was low (0.0067), and the diagnostic statistics were between 1.00-1.01, indicating a low degree of variance between MCMC chains. MCMC chains converged within the first 7,000 iterations and remained consistent throughout the remainder of MCMC with very little fluctuation in the shrink factor value. The sample size per chain was 19,460, slightly lower than the target of 25,000, likely due to fewer jumps being accepted when moving through the parameter distribution space.

The majority of the parameters related to plant physiological processes resulted in a successful PDA process, with three exceptions. The parameter ‘water conductance’ was targeted for PDA in all five PFT’s, and MCMC chains failed to completely converge within the designated 25,000 iterations in all five cases. Water conductance chains consistently had a narrower sampling value spread than the other parameters, although the overall chain patterns were consistent, indicating very little correlation between successive draws. Marginal density plots for water conductance weren’t multi-modal but they did display a right skew, suggesting the mean value is greater than the median. Water conductance displayed significant correlation to two other

parameters: 'V_{cmax}' for the PFT Late Conifer (correlation coefficient = -0.4728) and 'cuticular conductance' for the PFT Late Hardwood (correlation coefficient = 0.4811), shown in Figures 14 and 15, although the correlation coefficients for both are below $\pm$ 0.5, indicating a low enough correlation between variables that analysis didn't require modification. Gelman-Rubin diagnostic values ranged between 1.33-2.18, suggesting a significant degree of variation between chains. The second exception was V_{cmax}, which didn't converge for the PFT Late Conifer (diagnostic value = 1.68), although convergence was successful for the remaining four PFT's. The Gelman plot for Late Conifer V_{cmax} shows very little fluctuation in the shrink factor value after about 2,500 iterations, but a high value (> 1.1) is consistently maintained over time (Figure 16). For all other PFT's V_{cmax} chains were well mixed, and marginal density plots showed an approximately normal distribution with no sharp peaks or valleys. The final exception is for the parameter 'cuticular conductance', which is uniquely associated with PFT Late Hardwood. Cuticular conductance had a diagnostic value of 1.91. The significant physiological relationship between cuticular conductance and water conductance could explain the poor convergence for this parameter. The sample size per chain for all of the parameters related to plant physiological processes was 25,000.

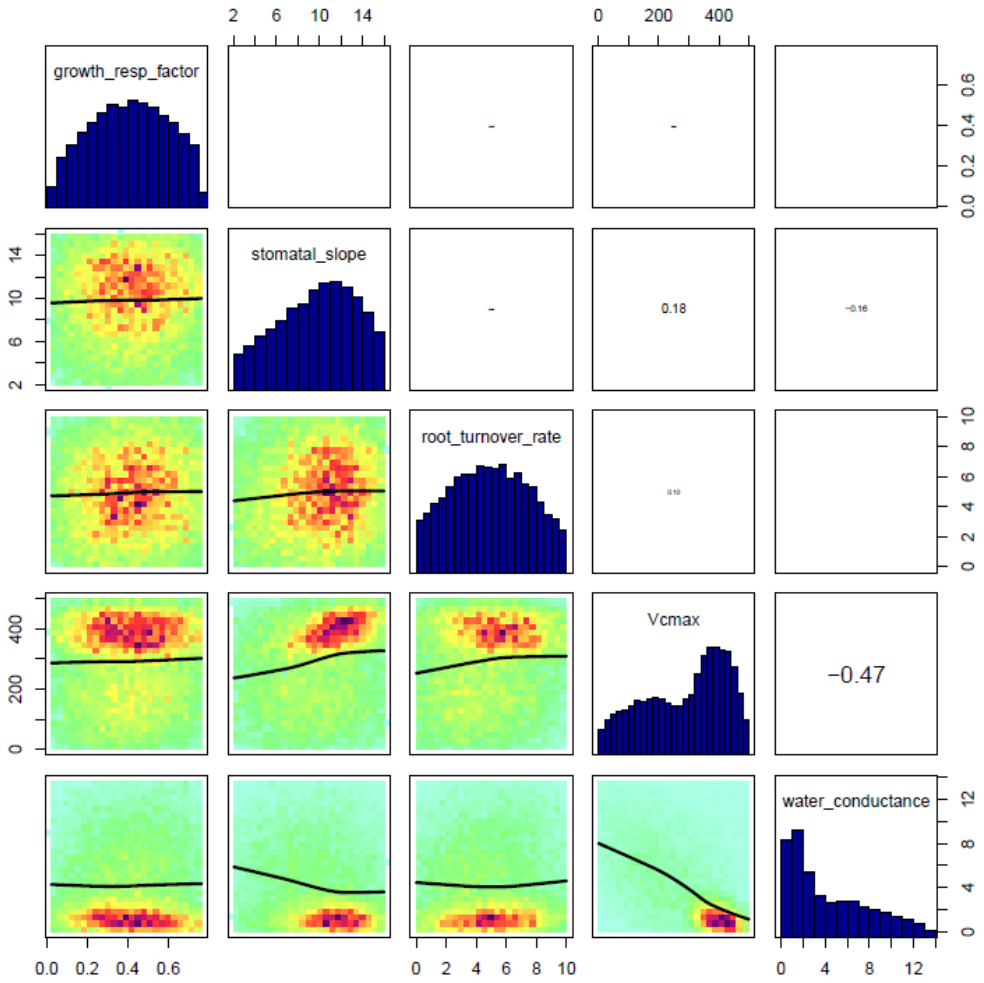

Figure 14: *Plot showing the degree of correlation between parameters targeted by PDA for PFT Late Conifer. Oval to round shaped point clouds indicate a lack of correlation, while tightly clustered point clouds appearing along a diagonal indicate linear correlations. Blue histograms are the marginal density plots for each parameter. The font size of correlation coefficients is in proportion to the degree of influence, where an empty white box or dashed line corresponds to a negligible correlation coefficient.*

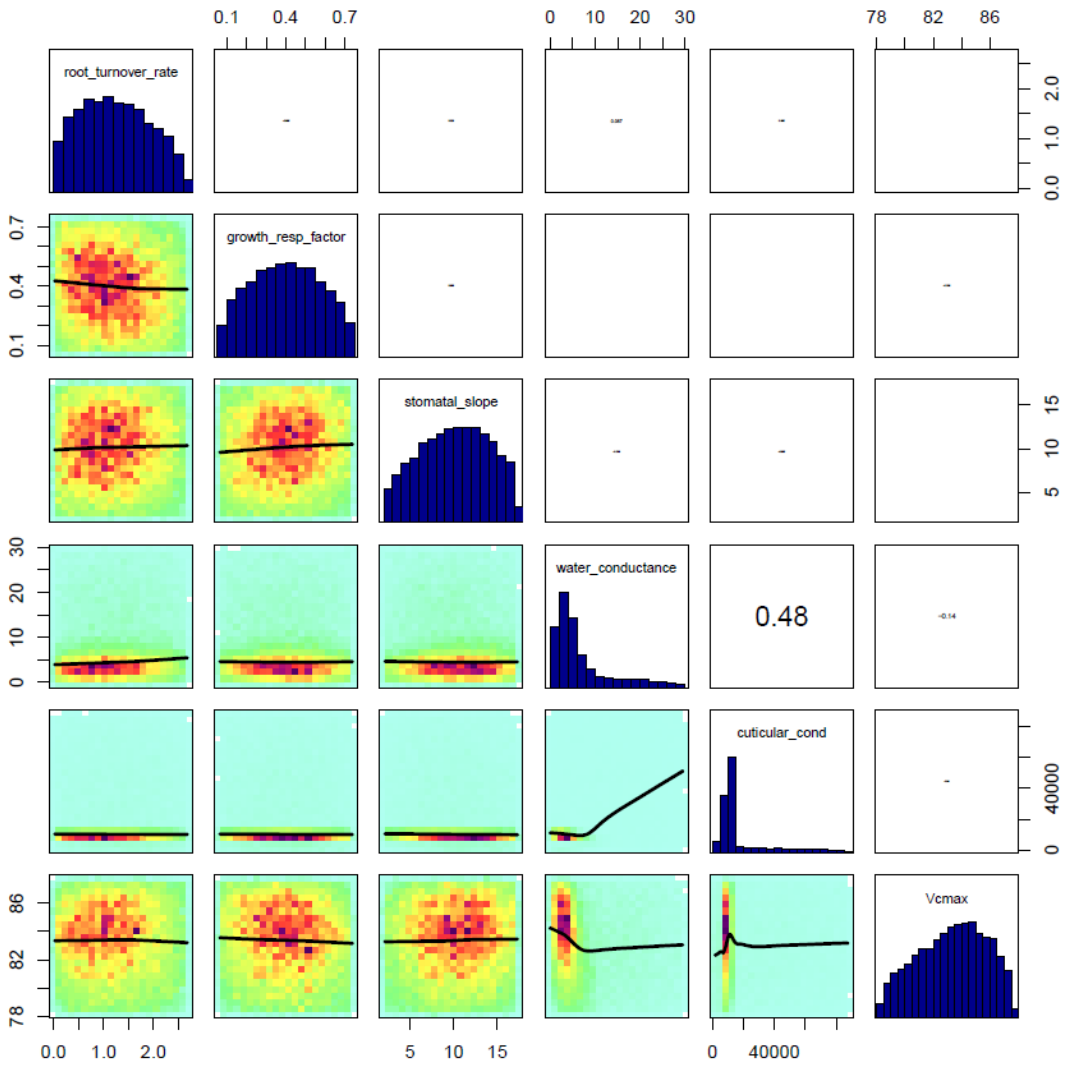

Figure 15: *Plot showing the degree of correlation between parameters targeted by PDA for PFT Late Hardwood. Blue histograms are the marginal density plots for each parameter. The font size of correlation coefficients is in proportion to the degree of influence, where an empty white box or dashed line corresponds to a negligible correlation coefficient.*

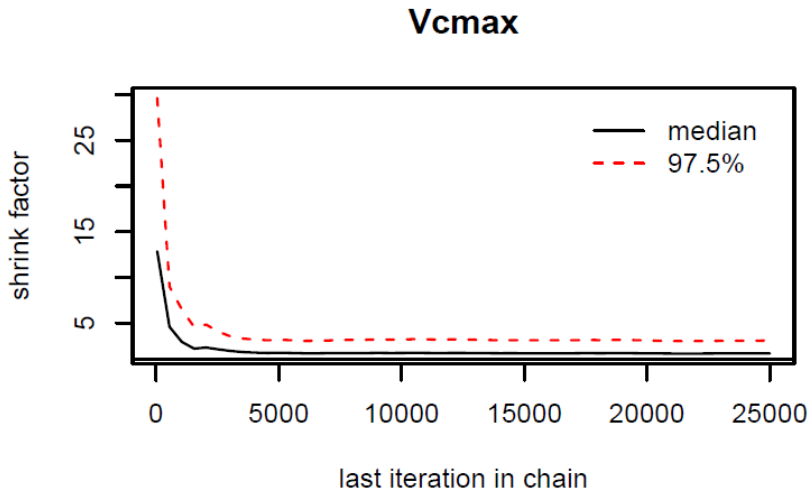

Figure 16: *Gelman plot for parameter V_{cmax} of PFT Late Hardwood*

3.3 Parameter constraint through PDA

Analysis of parameter constraint through PDA was assessed by PFT, although partial variances averaged across PFT's are presented as well. This approach to characterizing changes in individual parameter contributions to overall model predictive uncertainty was used because partial variance values are percentages that represent the normalized uncertainty contributions from all flexible parameters within a given PFT, with the sum of all parameter partial variances for a given PFT equaling one hundred percent. As such, if a parameter is well constrained its individual contribution to the total variance decreases, but that removed uncertainty must shift elsewhere. In practice this means when a set of parameters are targeted for PDA, partial variances are reduced for the majority of targeted parameters as they are able to be constrained, but will actually increase for one or two of the targeted parameters that were not well constrained through PDA, which in essence 'absorb' the uncertainty from the well constrained parameters.

Following PDA, growth respiration factor was the dominant parameter contributing to predictive uncertainty for three out of five vegetation PFT's, with partial variances ranging

between 73.3% and 99.1% (Table 4). Uncertainty contributions from growth respiration factor were less significant for the remaining two vegetation PFT's Late Hardwood and Northern Mid Hardwood, where partial variances were less than 3%. Water conductance was substantially constrained for all five vegetation PFT's, with variance reduction rates all above 96.65%. Stomatal slope was successfully constrained for four out of five vegetation PFT's (reductions ranging 7.64% to 79.12%), with the exception being North Mid Hardwood, where stomatal slope contributions to model predictive uncertainty increased from 0.56% to 7.3%. Root turnover rate uncertainty decreased for PFT's Late Conifer and Northern Pine, but increases were observed for Early Hardwood, Late Hardwood, and North Mid Hardwood. PDA was not able to successfully constrain V_{cmax}, as increases in partial variance were seen for all PFT's with the exception of Late Conifer, which showed a decrease of 67.6%. Of the two soil decomposition parameters targeted for PDA, respiration temperature increase and respiration optimal water, partial variance of respiration temperature increase dominated both before and after PDA. PDA shifted uncertainty from respiration optimal water to respiration temperature increase, which accounted for 99.5% of the soil PFT variance after PDA.

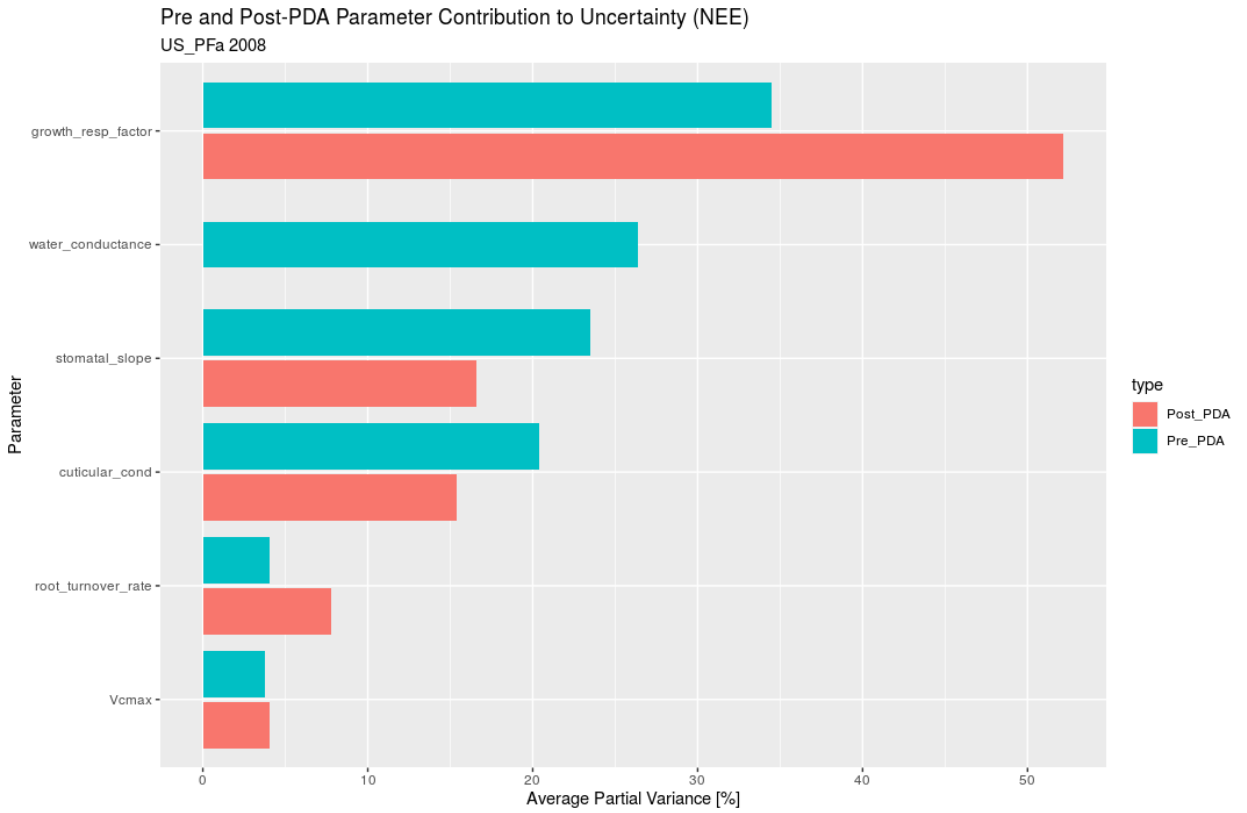

Figure 17: *Vegetation parameter contributions to overall model predictive uncertainty. Both pre-PDA and post-PDA values are averaged across the five different vegetation PFT's used.*

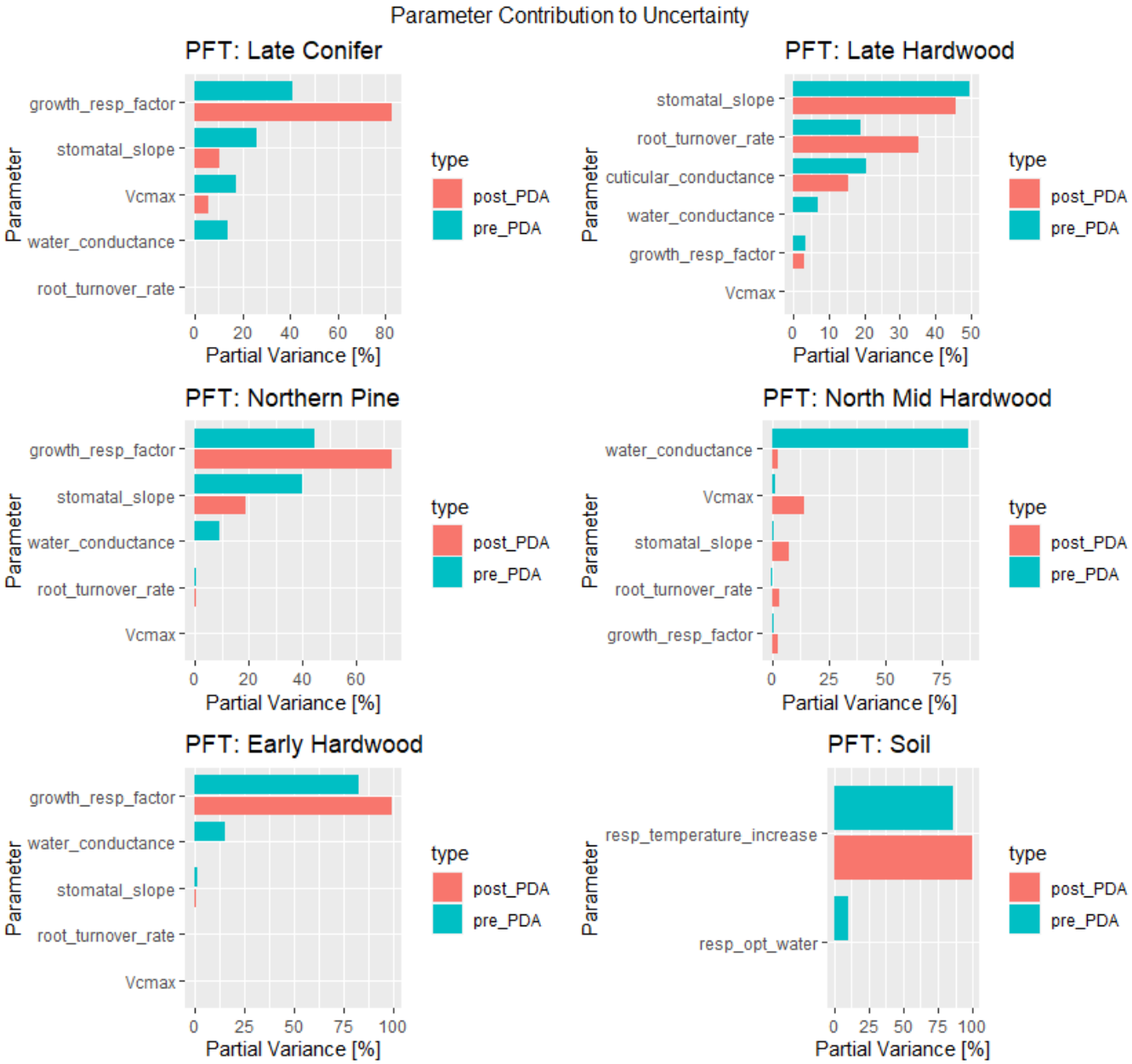

Figure 18: *Distribution of partial variance of parameters targeted for PDA, separated by PFT. Partial variance is presented for both pre and post-PDA.*

Table 4: *Parameter contributions to model predictive uncertainty, sorted by parameter to display trends in partial variance shifts as a result of PDA.*

PFT	Parameter	Pre-PDA	Post-PDA	Percent Change
Soil	resp_temperature_increase	86	99.5	15.69767442
Soil	resp_opt_water	9.95	0.235	-97.63819095
Early_Hardwood	growth_resp_factor	82.8	99.1	19.68599034
Late_Conifer	growth_resp_factor	41	82.7	101.7073171
Late_Hardwood	growth_resp_factor	3.53	2.95	-16.4305949
Northern_Pine	growth_resp_factor	44.5	73.3	64.71910112
North_Mid_Hardwood	growth_resp_factor	0.714	2.8	292.1568627
Early_Hardwood	water_conductance	14.8	0.00293	-99.9802027
Late_Conifer	water_conductance	13.8	0.0105	-99.92391304
Late_Hardwood	water_conductance	6.99	0.00802	-99.88526466
Northern_Pine	water_conductance	9.34	0.0188	-99.7987152
North_Mid_Hardwood	water_conductance	87.1	2.92	-96.64753157
Early_Hardwood	stomatal_slope	1.37	0.286	-79.12408759
Late_Conifer	stomatal_slope	26.1	10.2	-60.91954023
Late_Hardwood	stomatal_slope	49.7	45.9	-7.645875252
Northern_Pine	stomatal_slope	39.8	19.1	-52.01005025
North_Mid_Hardwood	stomatal_slope	0.559	7.3	1205.903399
Early_Hardwood	root_turnover_rate	0.0202	0.0226	11.88118812
Late_Conifer	root_turnover_rate	0.00473	0.00157	-66.80761099
Late_Hardwood	root_turnover_rate	19.2	35.1	82.8125
Northern_Pine	root_turnover_rate	0.455	0.44	-3.296703297
North_Mid_Hardwood	root_turnover_rate	0.41	3.37	721.9512195
Early_Hardwood	Vcmax	0.00513	0.0236	360.0389864
Late_Conifer	Vcmax	17.5	5.67	-67.6
Late_Hardwood	Vcmax	0.0239	0.0486	103.3472803
Northern_Pine	Vcmax	0.00684	0.0271	296.1988304
North_Mid_Hardwood	Vcmax	1.15	14.5	1160.869565
Late_Hardwood	cuticular_conductance	20.4	15.4	-24.50980392

3.4 Model predictive uncertainty

RMSE of model predicted NEE was calculated both before and after PDA using monthly averages, and was examined separately for both the entire year and the growing season. RMSE prior to PDA was 0.0303 kgC m-2 mo-1, compared to 0.0344 kgC m-2 mo-1 after PDA, indicating an increase of 13.74% in RMSE as a result of PDA. R^2 calculated using RMSE for the model prior to PDA was 0.5136, meaning that 51.36% of the variation around the mean of NEE is explained by the model. Following PDA R^2 decreased to 0.3707, suggesting the model's ability to explain variance in predicted NEE decreased as a result of PDA. However, a time-series examination of the data shows that while overall model predictive ability may have decreased, changes in predicative ability varied by season.

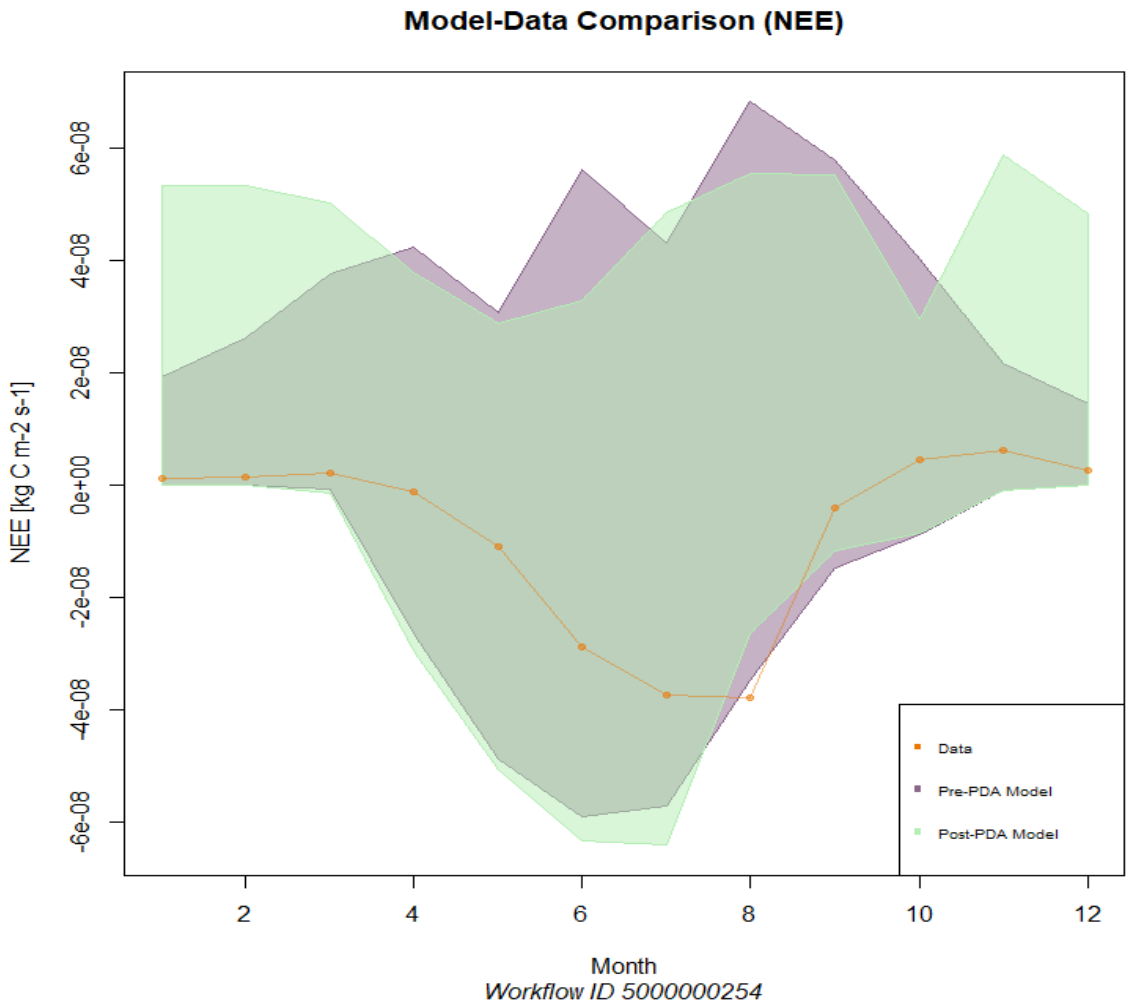

Figure 19: *Monthly smoothed plot of NEE 97.5% confidence interval spread before and after PDA, with observed average monthly flux values overlaid.*

Figure 19 depicts the seasonal variation in pre and post PDA narrowing of the confidence interval (CI) surrounding predicted NEE. The confidence interval can be thought of as the range of plausible values that are observed and captures the uncertainty in predictions, so ideally the CI spread would be narrowed through the application of PDA. It can be seen that during the growing season (defined as May 1-August 31) this is achieved, but during winter and fall the post-PDA CI

spread actually increases compared to the pre-PDA spread. This increase in spread is observed in the positive direction, indicating a predicted increase in the amount of carbon released by the ecosystem back into the atmosphere through enhanced respiration. There was no significant increase in predicted autotrophic respiration as a result of PDA (average difference = 8.10×10^{-10} kgC m-2 s-1), but comparison of pre and post PDA estimates of soil respiration indicates that simulation of enhanced soil respiration during winter and fall is responsible for the increase in CI spread. As shown in Figure 20, model predicted soil respiration rates are relatively consistent pre and post PDA during the growing season (average difference = 5.92×10^{-11}) but diverge substantially during winter and fall, with respiration rates increasing by up to 5.08×10^{-9} kgC m-2 s-1.

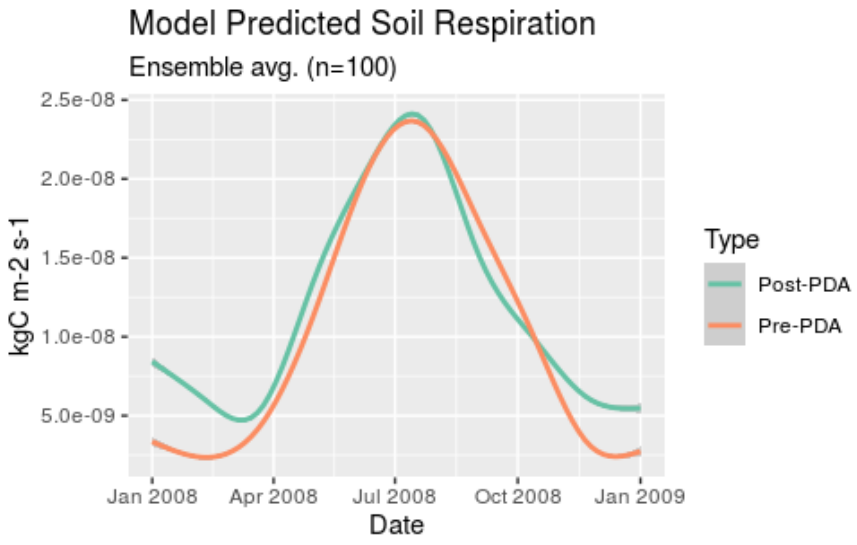

Figure 20: *Comparison of pre and post PDA model predicted soil respiration rates, where values are averaged across ensembles of size 100.*

During the growing season RMSE for monthly smoothed NEE predictions was 0.0501 kgC m-2 mo-1 prior to PDA and 0.0554 kgC m-2 mo-1 following PDA, an increase in error of 10.56% as a result of PDA, but 3.18% less than the increase observed when examining the full year at once.

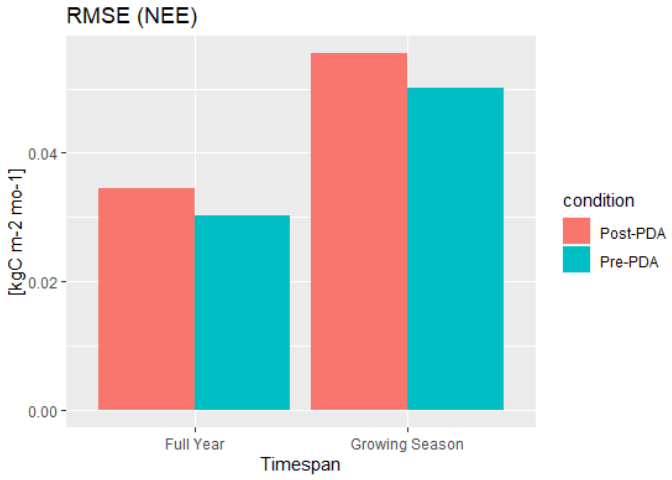

Figure 21: *Comparison of root mean squared error (RMSE) for pre and post PDA, presented both over the course of a full year and within the growing season.*

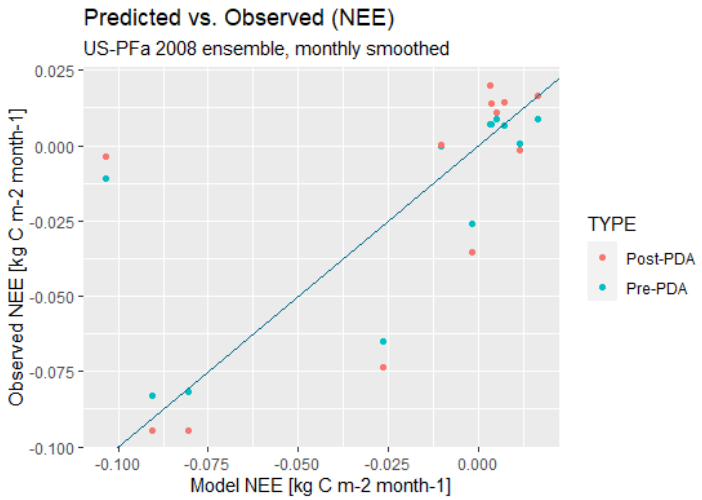

Figure 22: *Model predicted versus observed plot for monthly smoothed values of NEE both pre and post PDA, closer proximity of datapoints to the blue 1:1 line indicates a higher degree of agreement between model predicted values and observational data.*

When examining ecosystem carbon dynamics, model predictions of cumulative NEE are of particular interest, as they describe whether an ecosystem is a carbon source or a carbon sink as well as the magnitude of that source or sink. Cumulative NEE calculated from ECF tower observations classifies the site as a net carbon sink, with an uptake of 0.2672 kgC yr-1. The model also classifies the site as a net carbon sink both pre-and-post PDA at the annual scale, but the sink magnitude differs slightly, with the pre-PDA model predicting an uptake of 0.2285 kgC yr-1 and the post-PDA model predicting an uptake of 0.2268 kgC yr-1, as shown in Figure 23. Thus, when examining cumulative NEE at the annual scale, there is very little difference in carbon sink magnitudes at the chosen study site before and after PDA.

Narrowing in on the growing season, a greater difference between post-and pre-PDA predictions of carbon sink magnitudes is observed. The calibrated model predicts a net carbon uptake by the ecosystem of 0.2648 kgC, compared to an uptake of 0.2399 kgC predicted before calibration, and an uptake of 0.301 kgC observed by the ECF tower. This disagreement between growing season and annual cumulative NEE is likely due to enhanced soil respiration predicted by the model during winter and fall, reducing the net CO_2 uptake observed over the course of a full year.

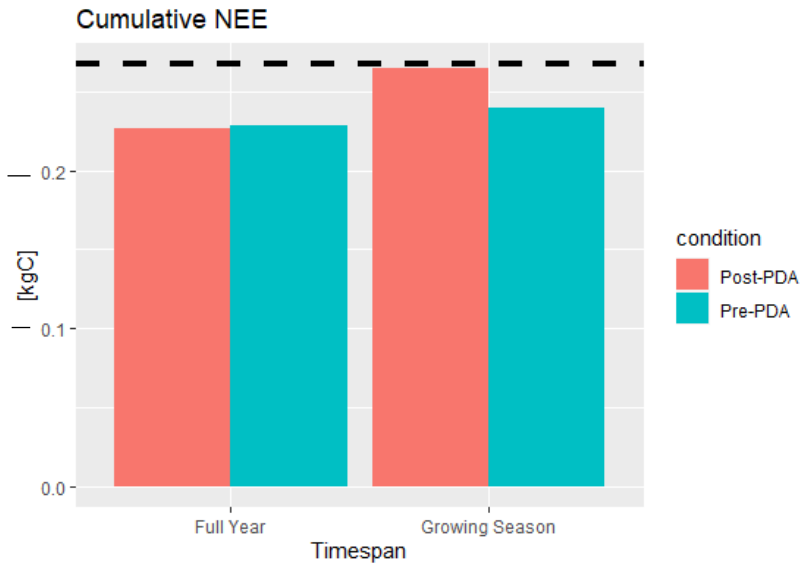

Figure 23: *Comparison of cumulative NEE for pre and post PDA, presented both over the course of a full year and within the growing season. The dotted black line represents the carbon sink magnitude calculated from observations.*

4. Discussion

4.1 Parameter uncertainty

Through sensitivity analysis and variance decomposition prior to the start of PDA, the flexible model parameters responsible for the greatest degree of model predictive uncertainty regarding NEE were identified. As NEE is frequently the variable used to describe the movement of carbon through an ecosystem, these isolated parameters can be thought of as the primary drivers of parameter uncertainty related to the terrestrial carbon cycle. Of the six plant physiological and two soil biogeochemical parameters chosen to target for PDA, all were related to either respiration (growth respiration, soil respiration temperature increase), water movement (water conductance, cuticular conductance, respiration optimal water content), or photosynthesis (stomatal slope,

V_{cmax}), with one parameter related to a turnover rate (root turnover rate). These influential parameters were consistent with what has been found in similar studies examining parameter uncertainty in complex models (Fer et al., 2018, Riccuito et al., 2018, Dietze et al., 2014). Fer et al. ran ED2.2 large ensemble simulations in a temperate northern hardwood forest as well (Bartlett Experimental Forest), and of the nine plant physiological and soil biogeochemical parameters targeted for PDA, five are consistent with parameters targeted in this study. Parameters related to respiration were found to be highly influential, with uncertainty related to soil respiration temperature increase dominating overall parameter uncertainty, and stomatal slope resulting in the largest degree of uncertainty in vegetation parameters. Ricciuto et al., 2018 demonstrated that although the same handful of parameters contributed significantly to NEE predictive uncertainty across multiple climate and vegetation types, patterns of sensitivity and uncertainty varied by PFT, a result which was observed in this study as well. However, several studies also highlighted mortality related parameters as contributing a large percentage of the collective parameter uncertainty, a result not observed in this study, although this too could be due to analysis only spanning a single year (Riccuito et al., 2018, Dietze et al., 2014).

The error-space emulation-based PDA process first introduced in Fer et al., 2018 and replicated here was shown to be a viable option for the application of data assimilation to constrain a dynamic ecosystem model. Although increases in individual parameter contributions to uncertainty were observed for a few of the flexible parameters, overall, the approach utilized in this study successfully reduced individual parameter uncertainty. A lower degree of correlation was observed between flexible parameters than was initially anticipated, with the only significant correlation observed between water conductance and either V_{cmax} or cuticular conductance, with correlation coefficients of -0.47 and 0.48. As both water and cuticular conductance are key

parameters that tightly couple the water and carbon cycles, this result could be due in part to assimilating only with respect to the carbon cycle, as opposed to a dual assimilation approach incorporating both LE and NEE, as was done in Fer et al., 2018. For other flexible vegetation parameters correlation was on average below 0.1, and was 0.0067 for the two examined soil parameters. The consistently low degree of correlation between parameters indicates that parameters are well constrained and that equifinality is likely not a substantial concern (Keenan et al., 2013). This is an important finding considering equifinality, or getting the 'right' answers for the 'wrong' reasons is often noted as one of the cardinal sins of model calibration (Franks et al., 1997).

Looking specifically at vegetation parameters; water conductance, stomatal slope, and cuticular conductance were all well constrained through PDA, with the contribution of water conductance to model predictive uncertainty becoming insignificant (<1%) post-PDA. The importance of both stomatal slope and water conductance illustrate the tight coupling between the carbon and water cycles. The ability to constrain both of these parameters is substantial; sensitivity studies have highlighted them as primary targets for improving through PDA, as they are responsible for a significant portion of model predictive uncertainty across a range of landcover and climate types (Ricciuto et al., 2018, Dietze et al., 2014).

In an ecosystem model such as ED2.2 where the carbon and water cycles are closely linked, quantifying the response of ecosystem carbon uptake to fluctuations in water availability through data assimilation could lead to more accurate predictions of changes in the carbon cycle in response to perturbation (Moore et al., 2008). The improvement is likely due to the importance of water related processes in modulating other terrestrial greenhouse gas fluxes (Raupach et al., 2005). Evapotranspiration and GPP (see equation 2) are closely linked by the processes that control their

magnitudes. For example, plants make constant tradeoffs between opening stomata for carbon uptake and losing water as a consequence, a process mediated by stomatal conductance, hydraulic properties of water uptake, and rate of carbon fixation. These properties are subsequently controlled by temperature, humidity, light availability, soil moisture, and CO_2 and nitrogen concentrations. Additionally, the upper Midwestern United States is a patchwork of forests interspersed with numerous streams, lakes, and wetlands, with over 15,000 lakes and 84,000 miles of rivers and streams in Wisconsin alone (WDNR, 2009). In light of this unique landscape, water cycle processes likely have even more influence on carbon dynamics and spatial variation in the chosen study region.

However, increases in parameter uncertainty were observed for three vegetation parameters: growth respiration factor, root turnover rate, and V_{cmax}. Increases in uncertainty were minimal for root turnover rate, and V_{cmax}, ranging between 0.3-3.5%, the vast majority of remaining uncertainty shifted to growth respiration factor, which experienced an average increase in predictive uncertainty across PFTs of 17.6%. So why did we observe such a substantial increase in uncertainty related to growth respiration factor, and is that finding unique to this study? In fact, growth respiration factor has been singled out as a highly uncertain carbon cycle parameter in a number of other modeling studies (Dietze et al., 2014), and has been described by Ricciuto et al., 2018 as the parameter which "dominates the variability in carbon cycle variables". Compared to the other flexible parameters targeted in this study, growth respiration factor is challenging to obtain direct measurements of, and existing priors are generally broad estimates that follow a uniform distribution. With the implementation of a Bayesian approach, the quality of priors, which are informed by measurements, dictate the success of PDA to some extent and broad uniform distributions are relatively uninformative (LeBauer et al., 2013). Considering the other targeted

parameters are well represented by observations, growth respiration factor essentially absorbs all of the residual variability in order to close the carbon budget. As stated in Dietze et al., 2014, reducing uncertainty related to this parameter is more than can be expected of PDA alone and hinges on improving additional direct constraints or using different model formulations to ensure biological realism.

Narrowing in on parameters related to soil biogeochemical processes, only two parameters were shown to be influential, respiration optimal water and respiration temperature increase. Respiration optimal water was fully constrained through PDA, with remaining uncertainty shifting to respiration temperature increase. This parameter describes how quickly heterotrophic respiration increases in response to increased temperatures, and is highly influenced by soil moisture and microbial community composition. The increase in parameter uncertainty related to soil respiration is interesting, considering this is where we saw a large degree of divergence between pre- and post-PDA model predictions, with soil respiration rates significantly increasing during the fall and winter months following PDA.

4.2 Seasonality in predicted NEE

PDA was able to successfully constrain the confidence interval surrounding model NEE predictions during the growing season, but the same performance enhancement was not observed during winter and fall. The amplification of soil respiration following PDA is an interesting result, and could be attributed in part to a number of contributing factors, several of which will now be explored. Considering the large seasonal differences in carbon sink and source dynamics typically observed in a mid latitude terrestrial ecosystem (Desai et al., 2007), it's likely that the model is overcompensating in favor of improved prediction alignment with observations during the growing

season, as most of the parameters targeted for PDA in this study are active primarily during the summer months.

The post-PDA predicted NEE disagreed with observations predominantly with regard to soil respiration rates, predicting higher rates of respiration during the winter and fall months than what was observed. In order to understand this result, we need to critically evaluate not only the model predictions, but the observations they're being compared against as well. The observations used in PDA as well as for the comparison of calibrated model predictions are obtained solely through an ECF tower, no direct soil respiration chamber data were used. Studies have shown that ECF towers systematically underestimate respiration rates compared to chamber-based measurements (Phillips et al., 2017). Often when comparing soil respiration (R_s) field data to total ecosystem respiration (R_{eco}) calculated by ECF towers, it is found that $R_s > R_{eco}$, a biological impossibility as R_{eco} includes not only R_s but also aboveground autotrophic respiration such as stem and leaf respiration (Bolstad et al., 2004). This discrepancy is particularly influential in forests and is thought to emerge from difficulties in how R_{eco} is calculated by ECF tower instrumentation (Phillips et al., 2017), highlighting potential biases in tower estimates of ecosystem respiration. Figure 24 shows that this is frequently the observed situation at Willow Creek, a research site in Northern Wisconsin located near the Park Falls WLEF tall tower site used in this study, with a similar landscape.

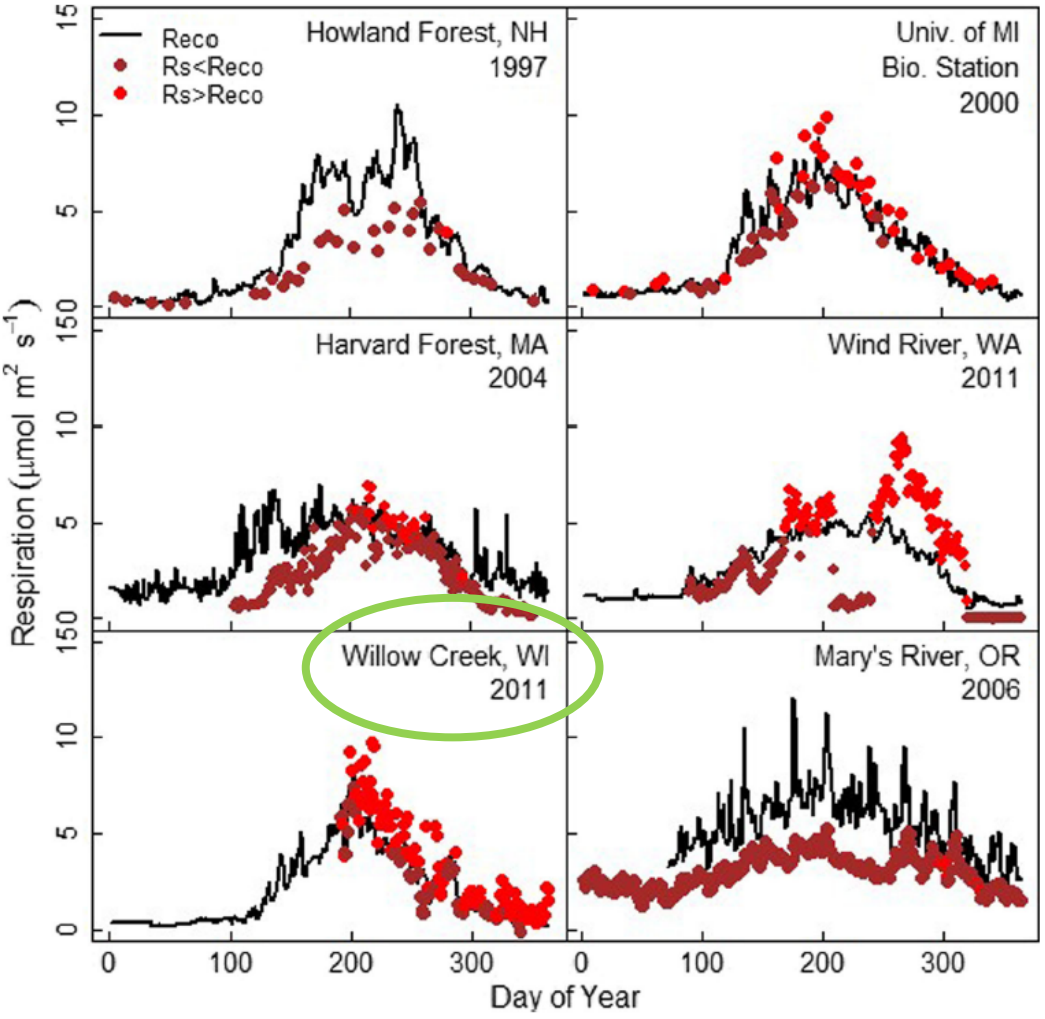

Figure 24: *Comparison of ecosystem respiration (R_{eco}) and soil respiration (R_s) at Willow Creek (lower left panel), an ECF tower equipped research site near the WLEF tall tower site used in this study, figure from Phillips et al., 2017.*

Perhaps the calibrated model is in fact overpredicting soil respiration rates, but with the assumed under-prediction of observed respiration rates, the magnitude of disagreement between the model and reality may not be as dramatic as it seems. It is worth noting that the two parameters that were not well constrained through PDA, growth respiration factor and respiration temperature

increase, are directly related to establishing ecosystem respiration rates. The incorporation of an additional soil respiration chamber data stream would likely further constrain the ecosystem respiration response and provide supplemental reductions of parameter contributions to uncertainty (MacBean et al., 2016, Phillips et al., 2017)). Viewing these results through a holistic lens, the calibrated model's poor performance in predicting carbon sink magnitudes during the winter and fall may not be of high importance, as carbon fluctuations during the growing season are often orders of magnitude larger than during the winter and fall when plants are not photosynthesizing, and play a more dominant role in the determination of the overall ecosystem carbon budget. Running model simulations for multiple years would likely help ascertain if these observed seasonal variations persist over a longer timescale.

5. Conclusions and Future Work

This study demonstrated that the approach outlined by Fer et al., 2018 is a viable methodology for the application of PDA to constrain a dynamic ecosystem model. It was shown that the success of Bayesian PDA depends largely on the quality and abundance of observational data available to construct informative prior distributions for model parameters. Ecosystem respiration rates (predominantly soil respiration) represent the largest remaining parameter uncertainty, and are the most significant point of disagreement between the model predictions and observations. Overall model predictive error, assessed using RMSE, increased as a result of PDA. The increase in RMSE shows that although individual carbon cycle parameters were well constrained through PDA, additional observations need to be incorporated to constrain growth respiration factor, and hydrological cycle dynamics should be integrated to see which other parameters might be highly influential for water cycle processes and potentially coupled with carbon cycle processes that were overlooked in this study. Examining the influence of closely

related hydrological parameters could illuminate whether carbon cycle parameters were optimized at the expense of influential hydrological parameters, resulting in lower overall model predictive power.

The reallocation of remaining uncertainty to parameters driving ecosystem respiration rates poses the question of whether poor constraint of respiration is responsible for the decrease in model predictive ability following PDA. Incorporating soil respiration chamber data as an additional observational constraint would likely improve model parameters related to respiration, which could then refine predictions of carbon dynamics (Phillips et al., 2017), as NEE has been shown to be particularly sensitive to how respiration is parameterized (Fox et al., 2009). The large spike in model predicted respiration during the winter and fall months observed in this study was not commonly seen in other similar studies where PDA was constrained with soil respiration data in addition to ECF tower observations (Fer et al., 2018). Completion of multi-year model simulations would help discern whether this is a persistent seasonal trend or an abnormality observed in this single study year. Additionally, other studies have indicated that optimizing ED2.2 parameters results in substantial improvements in the model's ability to predict seasonal and long-term patterns of CO_2 uptake (Medvigy et al., 2007).

Building upon the improvements mentioned above, future work also includes running the calibrated model with ECF tower data not used in the calibration process. The true test of model performance following PDA is to determine how well the model performs with independent data sets (Richardson et al., 2010). Fortunately, there is an abundance of ECF data within the same study region available through the Chequamegon Heterogenous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors (CHEESEHEAD) intensive field campaign. CHEESEHEAD was launched summer 2019 in Northern Wisconsin centered around

the WLEF tall tower utilized in this study. During the CHEESEHEAD project one of the world's highest density networks of ECF towers were deployed for four months spanning summer to fall to examine how the atmospheric boundary layer responds to spatial heterogeneity in surface energy fluxes (Butterworth et al., 2020). Seventeen stand-level ECF towers were deployed within a 10X10 km domain, in addition to two pre-existing stand-level towers and the WLEF tall tower, for a total of twenty ECF towers in the study domain, with tower sites varying in vegetation type as well as management history. Utilizing these rich datasets with flux data from the same study location experiencing the same climate forcing will allow for a true test of optimal model performance, and further exploration of the growing season trends observed in this study.

References

Baldocchi, D. (2008). Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany, 56*(1), 1-26.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., . . . Wofsy, S. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society, 82*(11), 2415-2434. doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Becknell, J. M., Desai, A. R., Dietze, M. C., Schultz, C. A., Starr, G., Duffy, P. A., . . . Staudhammer, C. L. (2015). Assessing Interactions Among Changing Climate, Management, and Disturbance in Forests: A Macrosystems Approach. *BioScience, 65*(3), 263-274. doi:10.1093/biosci/biu234

Bolstad, P. V., Davis, K. J., Martin, J., Cook, B. D., & Wang, W. (2004). Component and whole-system respiration fluxes in northern deciduous forests. *Tree Physiology, 24*(5), 493-504. doi:10.1093/treephys/24.5.493

Bonan, G. (2019). *Climate Change and Terrestrial Ecosystem Modeling*.

Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science, 320*(5882), 1444. doi:10.1126/science.1155121

Braswell, B. H., Sacks, W. J., Linder, E., & Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology, 11*(2), 335-355. doi:10.1111/j.1365-2486.2005.00897.x

Butterworth, B. J., Desai, A. R., Metzger, S., Townsend, P. A., Schwartz, M. D., Petty, G. W., . . . et al. *Connecting Land-Atmosphere Interactions to Surface Heterogeneity in CHEESEHEAD 2019*.

Canadell, J. G., & Raupach, M. R. (2008). Managing Forests for Climate Change Mitigation. *Science, 320*(5882), 1456-1457. doi:10.1126/science.1155458

Canadell, J. G., & Schulze, E. D. (2014). Global potential of biospheric carbon management for climate mitigation. *5*(1), 5282.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature, 408*(6809), 184-187. doi:10.1038/35041539

Desai, A. R., Helliker, B. R., Moorcroft, P. R., Andrews, A. E., & Berry, J. A. (2010). Climatic controls of interannual variability in regional carbon fluxes from top-down and bottom-up perspectives. *Journal of Geophysical Research: Biogeosciences, 115*(G2). doi:10.1029/2009jg001122

Desai, A. R., Moorcroft, P. R., Bolstad, P. V., & Davis, K. J. (2007). Regional carbon fluxes from an observationally constrained dynamic ecosystem model: Impacts of disturbance, CO2 fertilization, and heterogeneous land cover. *Journal of Geophysical Research: Biogeosciences, 112*(G1). doi:doi:10.1029/2006JG000264

Desai, A. R., Noormets, A., Bolstad, P. V., Chen, J., Cook, B. D., Davis, K. J., . . . Wang, W. (2008). Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: Implications for regional scaling. *Chequamegon Ecosystem-Atmosphere Study Special Issue: Ecosystem-Atmosphere Carbon and Water Cycling in the Temperate Northern Forests of the Great Lakes Region, 148*(2), 288-308.

Dietze, M. C., Lebauer, D. S., & Kooper, R. O. B. (2013). On improving the communication between models and data. *Plant, Cell & Environment, 36*(9), 1575-1585. doi:10.1111/pce.12043

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., . . . Wang, D. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences, 119*(3), 286-300. doi:10.1002/2013jg002392

Duursma, R. A., Blackman, C. J., Lopéz, R., Martin-StPaul, N. K., Cochard, H., & Medlyn, B. E. (2019). On the minimum leaf conductance: its role in models of plant water use, and ecological and environmental controls. *New Phytologist, 221*(2), 693-705. doi:10.1111/nph.15395

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*(19), 5801-5830. doi:10.5194/bg-15-5801-2018

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., & Sitch, S. (2014). Modeling the Terrestrial Biosphere. *Annual Review of Environment and Resources, 39*(1), 91-123. doi:10.1146/annurev-environ-012913-093456

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., . . . Wood, E. F. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research, 53*(4), 2618-2626. doi:10.1002/2016wr020175

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*(4), e2018MS001453. doi:10.1029/2018ms001453

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., . . . Van Wijk, M. T. (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. doi:https://doi.org/10.1016/j.agrformet.2009.05.002

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *Journal of Climate, 27*(2), 511-526. doi:10.1175/jcli-d-12-00579.1

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *7*(4), 457-472.

Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W., & Richardson, A. D. (2012). Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biology, 18*(8), 2555-2569. doi:10.1111/j.1365-2486.2012.02684.x

Law, B. E., Falge, E., Gu, L., Baldocchi, D. D., Bakwin, P., Berbigier, P., . . . Wofsy, S. (2002). Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *FLUXNET 2000 Synthesis, 113*(1), 97-120.

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*(2), 133-154.

Longo, M., Knox, R. G., Medvigy, D. M., Levine, N. M., Dietze, M. C., Kim, Y., . . . Moorcroft, P. R. (2019). The biophysics, ecology, and biogeochemistry of functionally diverse, vertically- and horizontally-heterogeneous ecosystems: the Ecosystem Demography Model, version 2.2 — Part 1: Model description. *2019*, 1-53.

Lovett, G. M., Cole, J. J., & Pace, M. L. (2006). Is Net Ecosystem Production Equal to Ecosystem Carbon Accumulation? *, 9*(1), 152-155.

Lukac, M. (2012). Fine Root Turnover. In S. Mancuso (Ed.), *Measuring Roots: An Updated Approach* (pp. 363-373). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lötscher, M., Klumpp, K., & Schnyder, H. (2004). Growth and maintenance respiration for individual plants in hierarchically structured canopies of Medicago sativa and Helianthus annuus: the contribution of current and old assimilates. *New Phytologist, 164*(2), 305-316. doi:10.1111/j.1469-8137.2004.01170.x

M, R. R., P, J. R., D, J. B., R, S. D., M, H., D, S. O., . . . C, C. S. (2005). Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology, 11*(3), 378-397. doi:10.1111/j.1365-2486.2005.00917.x

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *9*(10), 3569-3588.

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A METHOD FOR SCALING VEGETATION DYNAMICS: THE ECOSYSTEM DEMOGRAPHY MODEL (ED). *Ecological Monographs, 71*(4), 557-586. doi:doi:10.1890/0012-9615(2001)071[0557:AMFSVD]2.0.CO;2

Phillips, C. L., Bond-Lamberty, B., Desai, A. R., Lavoie, M., Risk, D., Tang, J., . . . Vargas, R. (2017). The value of soil respiration measurements for interpreting and modeling terrestrial carbon cycling. *Plant and Soil, 413*(1), 1-25. doi:10.1007/s11104-016-3084-x

Resources, W. D. o. N. (2009). Wisconsin Lakes. In: Bureau of Fisheries and Habitat Management.

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., . . . Hollinger, D. Y. (2008). Statistical properties of random CO2 flux measurement uncertainty inferred from model residuals. *148*(1), 38-50.

Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J. P., Dail, D. B., Davidson, E. A., . . . Savage, K. (2010). Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia, 164*(1), 25-40.

Ryan, M., Law, B. Interpreting, measuring, and modeling soil respiration. *Biogeochemistry* **73,** 3–27 (2005). https://doi.org/10.1007/s10533-004-5167-7

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statist. Sci., 4*(4), 409-423. doi:10.1214/ss/1177012413

Service, U. S. D. o. A. F. (2011). Land Areas of the National Forest System. In: United States Department of Agriculture.

Sherlock, C., Fearnhead, P., & Roberts, G. O. (2010). The Random Walk Metropolis: Linking Theory and Practice Through a Case Study. *Statist. Sci., 25*(2), 172-190. doi:10.1214/10-STS327

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., . . . Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology, 9*(2), 161-185. doi:10.1046/j.1365-2486.2003.00569.x

Spiegelhalter, D., & Rice, K. (2009). Bayesian statistics. In (Vol. 4, pp. 5230). Scholarpedia.

Suseela, V., Conant, R. T., Wallenstein, M. D., & Dukes, J. S. (2012). Effects of soil moisture on the temperature sensitivity of heterotrophic respiration vary seasonally in an old-field climate change experiment. *Global Change Biology, 18*(1), 336-348. doi:10.1111/j.1365-2486.2011.02516.x

Taboga, M. (2017). *Lectures on Probability Theory and Mathematical Statistics - 3rd Edition*: CreateSpace Independent Publishing Platform.

Taiz, L. (1991). *Plant Physiology*. Redwood City, CA: Benjamin/Cummings Pub. Co.

Van Oijen, M. (2017). Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models. *Current Forestry Reports, 3*(4), 269-280. doi:10.1007/s40725-017-0069-9

Walker, A. P., Beckerman, A. P., Gu, L., Kattge, J., Cernusak, L. A., Domingues, T. F., . . . Woodward, F. I. (2014). The relationship of leaf photosynthetic traits – Vcmax and Jmax – to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. *Ecology and Evolution, 4*(16), 3218-3235. doi:10.1002/ece3.1173

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *60*, 167-179.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., . . . Wang, Y. P. (2009). Improving land surface models with FLUXNET data. *6*(7), 1341-1359.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., & Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology, 11*(1), 89-105. doi:10.1111/j.1365-2486.2004.00891.x