UNIVERSITY OF CALIFORNIA,

IRVINE

Predictive Modeling of Environmental Systems:

Applications of parameter estimation, data assimilation, sensitivity analysis, and model emulation

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Elias C. Massoud

Dissertation Committee:

Professor Jasper A. Vrugt, Chair

Professor Amir AghaKouchak

Professor Kristen Davis

2017

Dedication

" *Progress lies not in enhancing what is, but in advancing toward what will be.* "

- Gibran Khalil Gibran, Lebanese Poet

During the 2012-2013 school year there was lots going through my mind. I had completed a Master's degree at UCLA in Civil and Environmental Engineering, a talented and skilled young man with a large amount of potential, searching for what my future held next. I then met my advisor, and friend, Dr. Jasper Vrugt at a graduate welcome event at UCI, and suddenly the next 4 years were laid out; although uncertainty on what to do only began from there.

Research commenced immediately and I started to investigate a whole range of studies on numerical modeling of complex systems. By questioning new as well as old theories, applying novel tests on existing studies, and sharing fresh results with collaborating scientists and researchers from around the world, I was able to put together a series of chapters that are contributing to today's knowledge on the prediction of environmental systems. But this is only the beginning, as the responsibility of sharing everything I've learned now lays heavy with me and the excitement to move forward and continue contributing to this field of research urges me on.

Before presenting the chapters written through this journey, I would like to acknowledge all who have allowed me to get to this point. The journey starts with my family. My father Charbel and mother Najwa Massoud, I love you dearly, you always remind me to be bright and to achieve my full potential. My sister Dr. Theresa Massoud-Chater, you are a continuous source of support and motivation, especially since you also wrote a dissertation and understand what it takes to accomplish this achievement. My sister Maria Massoud-Nammour, you have been there for me since the beginning, and I know you will always be there. My brother and best friend Youssef Massoud, without you I would not be the person I am today, you are always someone I can reflect off and learn from.

The knowledge I gained during these four years are much in thanks to a group of researchers that selflessly dedicated time and energy to help me advance and succeed. Dr. Chonggang Xu at the Los Alamos National Laboratory, thank you for hosting me on two separate occasions and allowing me to stay a total of 6 months to conduct research with your group. Dr. Jef Huisman at the University of Amsterdam, thank you for allowing me to visit and expand on the work done at your research facility. Dr. Eric Laloy at the Belgium Nuclear Research Facility, thank you for reminding me that a collaboration with a person across the world is indeed possible, and for always responding in such a timely fashion as to allow progress and innovation to take place ever so smoothly. To the many professors and researchers at UCI, thank you for allowing me to be a part of the school and for sharing so many great ideas and conversations.

Last, and surely not least, Jasper my academic advisor, you taught me so many things about coding, statistical theory, writing papers, lecturing and teaching, and most importantly, about life. Thank you for everything. These past four years have been a roller coaster, a journey that I will always appreciate. I know that wherever I go next, and whatever I do, I will carry with me so much of what I learned from you. I hope this friendship continue for many years, and if we are lucky, many decades.

As a final note, I would like to give appreciation to my American culture as well as my Lebanese heritage. Both have instilled in me special values, shaping me into the person I am today. Born and raised in California to a Lebanese household always put me at a culture-clash. Now, these difficulties have proven to be blessings indeed, and my thesis is dedicated to all who have left their native countries in search of a brighter future. This is to you.

Irvine, California
August 2017
Elias C. Massoud

Contents

List of Figures

List of Tables

Acknowledgement

The UC-Lab Fees Research Program Award 237285
Primary PhD. funding 2013-2017
- Next Generation Carbon-Nitrogen Model , PI's: Chonggang Xu and Jasper Vrugt

United States Department of Energy (DOE)
DOE Office of Science Graduate Student Research Program (SCGSR) 2015
- Internship at Los Alamos National Laboratory (LANL) under guidance of scientist Chonggang Xu in the Vegetation Dynamics Research Lab

Curriculum Vitae

Elias C. Massoud

EDUCATION

Doctor of Philosophy, Civil & Environmental Engineering
University of California, Irvine (2013-2017)

Master's of Science, Civil & Environmental Engineering
University of California, Los Angeles (2011-2012)

Bachelor's of Science, Civil & Environmental Engineering
University of California, Los Angeles (2007-2011)

TEACHING EXPERIENCE

Civil & Environmental Engineering 20, Computer Programming for Engineers
University of California, Irvine.
Instructor: Summer 2014, 2015, 2017.
Teaching Assistant: Fall 2013, Spring 2014, Fall 2014, Fall 2016.

Civil & Environmental Engineering 180, Transportation Engineering and Planning
University of California, Los Angeles.
Teaching Assistant: Spring 2012.

GRANTS

United States Department of Energy (DOE)
DOE Office of Science Graduate Student Research Program (SCGSR) 2015
- 5-month Internship at Los Alamos National Laboratory (LANL) in the Vegetation Dynamics Research Lab

JOURNAL PUBLICATIONS

1. Ali, A. A., Xu, C., Rogers, A., Fisher, R. A., Wullschleger, S. D., McDowell, N. G., **Massoud, E. C.**, Vrugt, J. A., Muss, J. D., Fisher, J. B., Reich, P. B., and Wilson, C. J.: A global scale mechanistic model of the photosynthetic capacity, Geosci. Model Dev. Discuss., 8, 6217-6266, doi:10.5194/gmdd-8-6217-2015, 2015.

2. **Massoud, E. C.**, Purdy, A.J., Miro, M., Famiglietti, J.S., Vrugt, J.A., Groundwater Sustainability in the Central Valley: An empirical approach to estimate groundwater depletion and recharge. Journal of Water Resources Planning and Management, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017.

3. Vrugt, J. A., and **Massoud, E. C.**, eds. Book chapter: Bayesian Analysis of Complex Systems. XXX Press, Submitted to Springer 2016.

4. **Massoud, E. C.**, Huisman, J., Beninca, E., Heerklos, R., Vrugt, J. A.: Probing the limits of predictability: data assimilation improves forecasts of complex dynamics in ecosystems. Ecology Letters, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017.

5. **Massoud, E. C.**, Xu, C., Fisher, R. McDowell, N.G., et al.: Identification of key parameters controlling vegetation dynamics in a demographically structured land surface model. Journal of Advances in Modeling the Earth System, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017.

6. **Massoud, E. C.**, Christofferson B.O., Purdy, A.J., Xu, C.: Bayesian inference of hydraulic properties in and around a Douglas White Fir using a process based ecohydrologic model. Environmental Modeling & Software, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

7. **Massoud, E. C.**, E. Laloy, Xu, C., Vrugt, J.A.: Emulation of environmental models using polynomial chaos expansion. Environmental Modeling & Software, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

8. Johnson, D.J., Needham, J., Xu C., **Massoud, E. C.**, et al. Tree survival in the tropics reveals limited demographic diversity. Nature, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

Abstract of the Dissertation

Predictive Modeling of Environmental Systems: Applications of parameter estimation, data assimilation, sensitivity analysis, and model emulation

By: Elias C. Massoud

Doctor of Philosophy in Civil & Environmental Engineering
University of California, Irvine, 2017
Professor Jasper A. Vrugt, Academic Chair

In 2013, the World Meteorological Organization (WMO) urged the global community for coordinated international action against accelerating and potentially devastating climate change. Preliminary data indicated that CO_2 levels increased more between 2012 and 2013 than during any other year since 1984, and this was possibly related to reduced uptake by the Earth's biosphere in addition to the steadily increasing emissions from the Earth's surface. In the upcoming decades, it will be critical for scientists and policy makers to not only resolve the problem of carbon emissions by assessing human behavior, but also to understand as thoroughly as possible the underlying coupled processes of the Earth's atmosphere and biosphere in order to adequately measure and estimate the fluxes of carbon, water, and energy that are dictating the climatic trends we observe today. Fortunately, our ability to understand Earth's processes and predict climate change is improving.

This thesis covers a suite of environmental models and numerical methods to disentangle information found both in observed data as well as model simulations. Various methods are applied such as parameter estimation with Markov Chain Monte Carlo (MCMC), state estimation with data assimilation using the Ensemble Kalman Filter (EnKF), and sensitivity analysis of model parameters using the Fourier Amplitude Sensitivity Test (FAST), which all in one way or another offer treatments to predictive uncertainty. Furthermore, applying these methods on more sophisticated and complex models can be impossible sometimes due to their

high CPU costs; in this thesis model emulators are built using Polynomial Chaos Expansion (PCE) to reduce the computational burden for various environmental models. Overall, our goal in this dissertation is to present what tools are currently available for making predictions of environmental systems, with emphasis on maintaining accuracy of model simulations when compared to observed data, optimizing the efficiency of computationally heavy models to minimize their run time costs, and obtaining fidelity of model structures to properly represent the underlying hydrologic, biophysical, and biogeochemical processes occurring on our Earth.

Chapter 1

Uncertainty Quantification of Complex System Models: Bayesian Analysis

by Jasper A. Vrugt and Elias C. Massoud

Reference:

Vrugt, J. A., and Massoud, E. C., eds. Book chapter: Bayesian Analysis of Complex Systems. XXX Press, Submitted to Springer 2016

This book chapter summarizes the main elements of Bayesian probability theory to help reconcile dynamic environmental system models with observations, including prediction in space (interpolation), prediction in time (forecasting), assimilation of data, and inference of the model parameters. Special attention is given to the treatment of parameter uncertainty (first-order approximations and Bayesian intervals), the prior distribution, the formulation of the likelihood function (using first-principles), and sampling techniques used to estimate the posterior target distribution.

1.1 Introduction and Scope

The Earth is the densest planet in our solar system and the only astronomical object known to man-kind to harbor life. About 71% of the Earth's surface is covered with water, and the remaining 29% constitutes land mass made up of continents and islands dissected by rivers, lakes, and other sources of water that contribute to the hydrosphere. The large scale motion of the Earth's outermost shell (lithosphere), composed of several tectonic plates which float on a hotter, softer layer in the mantle (asthenosphere), has created mountain ranges and volcanic activity on plate boundaries. Co-evolution and juxtaposition of these topographic features with climatic and geologic variations have resulted in a highly diverse landscape with large variations in soils, vegetation, geomorphology, and biota (biosphere). These landscapes can be conceived as a series of large and small ecosystems, nested within one another in a hierarchy of spatial scales.

Ecosystems constitute a complex network of living organisms, which are interconnected and linked together with the abiotic environment through a myriad of interrelated physical, chemical and biological processes operating at or near the Earth's surface. Many of these processes are difficult, costly, labor intensive, and/or unethical to measure directly in the field, particularly at large spatial scales. This daunting complexity has stimulated researchers in many different fields of study to explore the use of mathematical modeling to mimic the behavior of complex systems. Computer models are particularly useful to gain (new) insights and understanding of system functioning and to predict behavior into the space (interpolation) and time (forecasting) domain. The capabilities of such models exceed by far traditional paper-and-pencil calculations and can involve simulations

on spatial scales of individual atoms to the entire ecosystem, and temporal scales of nanoseconds to many millions of years. Examples include numerical weather prediction models, astrophysical and cosmological simulations of dark matter, computational modeling of the brain, and spatially distributed simulation of environmental systems. The CPU-time of these simulations can vary from less than a second for simple dynamic models with fixed (integration) time step up to many hours of calculation for spatially explicit models involving multidimensional numerical solution of (systems of) differential-algebraic or ordinary/partial differential equations.

The model building process is strongly influenced by perception, intuition, and prior knowledge on system functioning and reality, and colored by mental concepts (state of mind). From countless processes and mechanisms, the modeler seeks to isolate, detect, and generalize into laws those key principles that explain the observed data. Their selection and translation to a mathematical model is the most critical, difficult, and subjective part of modeling. To guard against the use of an inadequate model, statisticians advise selecting the "best" model among a set of plausible candidate models chosen and/or construed by the researcher(s). This approach rules out model selection bias and recognizes explicitly the ambiguity in the interpretation and analysis of complex natural systems. The ensemble of models, or hypotheses, constitute a finite sample of possible explanations of the data deemed plausible a-priori from the extremely large, perhaps even incomprehensible, space of alternatives. This can include black-box, conceptual (empirical), and physically-based models and involve widely different mechanisms of the spatio-temporal processes that determine system behavior and response. Each of these models might be as justifiable as the other (Clark et al., 2011; Vrugt and Robinson, 2007b; Ye et al., 2008).

Figure 1.1 provides a schematic overview of most important sources of uncertainty that affect our ability to mimic perfectly complex dynamical systems. These sources of uncertainty have been discussed extensively in the literature, and many different (non)statistical methods have been developed to quantify parameter, calibration data, model output and state variable uncertainty. Model structural errors (4: epistemic error) have received relatively little attention, yet are key to learning and scientific discovery (Gupta et al., 2008; Vrugt et al., 2005; Vrugt and Sadegh, 2013).

1.2 Model Calibration

Consider a n-vector of measurements, $\widetilde{\mathbf{Y}} = \{\widetilde{y}_1, \ldots, \widetilde{y}_n\}$ observed at discrete times $t = \{1, \ldots, n\}$ that summarizes the response of an environmental system $\Im$ to k temporally-variant control inputs, $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, with column elements, $\mathbf{b}_t = \{b_{t1}, \ldots, b_{tk}\}$. We use a computer model, $\mathcal{M}(\cdot)$, to explain the observed data

$$\widetilde{\mathbf{Y}} \leftarrow \mathcal{M}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}}) + \mathbf{E}, \tag{1.1}$$

where $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_d\}$ is the $d \times 1$-vector of model parameters, $\mathbf{x}_0$ stores the values of the state variables at the start of simulation, $\widetilde{\mathbf{B}}$ signifies the $k \times n$ control matrix with temporal measurements of the forcing variables, and $\mathbf{E} = \{e_1, \ldots, e_n\}$ is a vector of residuals. The index t for time takes on strictly positive integer values in the remainder of this Chapter, $t \in \{1, \ldots, n\} \in \mathbb{N}_+$, yet may take on real values, $t \in (0, n] \in \mathbb{R}_+$ in the actual system model, $\mathcal{M}(\cdot)$, to resolve for continuous-time processes, wherein the simulated output at $t = 0$ is defined completely by $\mathbf{x}_0$.

The model in Equation (1.1) simplifies considerably the description of the spatially distributed real-world system, into a lumped topology consisting of much fewer, and discrete, entities. This simplification is computationally convenient in that it reduces to a finite dimension the state space of the system, and the partial differential equations of the continuous time and space domain of the physical system into ordinary differential equations with much fewer parameters. If deemed appropriate, a spatially explicit can be used instead with control vector, $\mathbf{b}$, formulated as a two- or three-dimensional matrix to account explicitly for spatially-varying boundary conditions. Without further loss of generality, we restrict the model parameters to a closed space, Θ, equivalent to a d-dimensional hypercube, $\boldsymbol{\theta} \in \Theta \in \mathbb{R}^d$, called the feasible parameter space. The n-vector of error residuals, $\mathbf{E}$, thus depends on the assumed model, $\mathcal{M}$, and its associated parameters, initial states and forcing data, hence

$$\mathbf{E}_{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}}) = \widetilde{\mathbf{Y}} - \mathbf{Y}_{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}}) \tag{1.2}$$

where $\mathbf{Y}_{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}})$ signifies the simulated output of the model, $\mathcal{M}$.

For the time being, lets make the convenient assumption that $\mathcal{M}$ mimics perfectly the underlying

system, $\Im$, it is intended to represent. Lets further assume that the forcing data are observed without measurement error, $\delta(\mathbf{B}, \widetilde{\mathbf{B}}) = 0$, and that errors in the initial states, $\mathbf{x}_0$, pose no harm as their impact on the simulated output, $\mathbf{Y}$, diminishes rapidly with advancing time. This latter assumption is certainly appropriate for real-world systems controlled by negative (or degenerative) feedback. A prime example of negative feedback is the thermostat. If room temperature drops below a reference value, the furnace will supply heat and restore the temperature to "normal". This cooling-warming cycle thus regulates room temperature. Similar degenerative interactions are found among processes that control the rainfall-runoff transformation in a watershed. Surface runoff, overland flow, evaporation, infiltration, transpiration, drainage, and recharge (among others), act together to remove excess precipitation, thereby promoting convergence of the soil moisture status to a stable state. The existence of such equilibrium state is easily verified in practice using repeated numerical simulation with a watershed model using different values of the initial states. This equilibrium state does not exist for systems whose behavior is regulated by positive feedback as small perturbations to the initial states can lead to widely different responses of the model via exponential growth, oscillation or chaotic behavior. For systems with negative feedback, a spin-up period of Q days therefore suffices to promote stability and ameliorate the effect of state initialization errors on the model output, $\lim_{t \to Q} \delta\big(y_t(\widetilde{\mathbf{x}}_0), y_t(\mathbf{x}_0)\big) \to 0$.

For systems with generative (negative) feedbacks, the error in the initial states poses no harm as its effect on system simulation rapidly diminishes when time advances. One can therefore take advantage of a spin-up period to remove sensitivity of the modeling results (and error residuals) to state value initialization.

The assumptions of perfect model, input data, and initial states (due to spin-up period) are common to environmental modeling. This ideal case leaves as our only "unknowns" the model parameters. The residual vector can thus be written as

$$\mathbf{E}(\boldsymbol{\theta}) = \widetilde{\mathbf{Y}} - \mathbf{Y}(\boldsymbol{\theta}) = \{e_1(\boldsymbol{\theta}), \ldots, e_n(\boldsymbol{\theta})\}, \tag{1.3}$$

and necessitates inference on $\boldsymbol{\theta}$ to minimize the n-vector of residuals, $\mathbf{E}(\boldsymbol{\theta})$.

Figure 1.2 shows an overview of the model calibration problem. The prior values of the

parameters provide a simulation (gray line) that mimics reasonably well the transient behavior of the system (blue dots), yet underestimates systematically the peaks and recession periods. A subsequent trial of the parameter values much better explains (green line) the observed data. Model calibration now involves the searching of the parameter values that mimic "best" the observed system response. The calibrated model can then serve a host of different purposes such as process analysis, evaluating different management strategies, and prediction of system behavior into the space and/or time domain. These tasks can only be completed with confidence if the physical system of interest and its control inputs (forcing variables) satisfy "constancy". This stationarity assumption is rather convenient and opens up the wide arsenal of (multivariate) statistical and nonlinear optimization methods for inference of the model parameters (Sadegh et al., 2015). Note, per scope of this book, we will focus primarily on hydrologic models with *output* calibration targets such as river discharge and soil moisture content.

The word "best" appears purposely quoted in the previous paragraph as much research has shown that there is no unambiguously correct way in which to determine unique model parameters. Indeed, the "optimal" parameter values are dependent critically on the assumptions that are made with respect to the different error sources of Figure 1.1. In the ideal case (perfect model, input data and initial states), the sum of squared residuals (SSR)

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\theta}} F(\boldsymbol{\theta}) = \sum_{i=1}^{n} e_i(\boldsymbol{\theta})^2, \tag{1.4}$$

provides unbiased and minimum-variance estimates of the parameters when the measurement errors of the calibration data, $\widetilde{\mathbf{Y}}$, are homoscedastic (constant variance) and serially (temporally) uncorrelated. This is also referred to as the least squares solution. Visually, this solution minimizes the sum of squared vertical distances between the n-vector of data points, $\widetilde{\mathbf{Y}}$ and the corresponding simulated values, $\mathbf{Y}(\boldsymbol{\theta})$, of the model. The lower the value of the SSR the better the model fits the data. When the measurement errors are believed to have a nonconstant variance, the heteroscedastic maximum likelihood estimator (HMLE) can be used (Sorooshian and Dracup, 1980). Other commonly used metrics in hydrologic model calibration include the coefficient of

determination, the index of agreement, and the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970), although these metrics are not rooted in statistical theory.

In Equation (1.4) the function $F(\boldsymbol{\theta})$ is also called the objective function. This function lumps the n-residuals of the vector $\mathbf{E}(\boldsymbol{\theta})$ into a single aggregate measure of model-data mismatch, and this measure is subsequently minimized (or maximized, if appropriate) with an optimization algorithm. This constitutes the field of constrained optimization, that is,

$$\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} F(\boldsymbol{\theta}), \tag{1.5}$$

where the goal is to find the optimum parameter values as rapidly as possible using the smallest number of model evaluations. The constraints are equivalent to the lower and upper bounds of the parameters, and ensure that the optimum parameter values reside in the feasible parameter space, $\boldsymbol{\theta}$. The choice of objective function, however remains a rather intricate and difficult task, fraud with subjective assumptions regarding model structural, control data, and calibration data errors.

Optimization algorithms provide an estimate of the "best" parameter values that minimize (maximize) some predefined objective function, $F(\boldsymbol{\theta})$. It would be naive, however to rely on such single unique estimate of the parameters in the presence of epistemic uncertainty and measurement errors of the control input and calibration data. Indeed, practical experience suggests that it is typically difficult to find a single "best" vector of parameter values, whose performance obviates consideration from other feasible solutions. It is therefore of paramount importance to investigate and delineate properly the space of feasible solutions. This is key to (among others) analysis of parameter identifiability and quantification of the uncertainty associated with the simulated model output.

1.3 Parameter Uncertainty: First-order Approximation

Per statistical theory, we can approximate the confidence intervals of the parameters by centering around the $d\times 1$-vector of optimum parameter values, $\boldsymbol{\theta}^*$, a d-variate normal distribution, $\mathcal{N}_d\big(\boldsymbol{\theta}^*, \mathbf{C}(\boldsymbol{\theta}^*)\big)$, with $d \times d$ covariance matrix $\mathbf{C}$ evaluated at $\boldsymbol{\theta}^*$. The probability density function, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, of this

multivariate Gaussian is given by

$$P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = (2\pi)^{-\frac{d}{2}}|\mathbf{C}(\boldsymbol{\theta}^*)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T\mathbf{C}(\boldsymbol{\theta}^*)^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\right), \tag{1.6}$$

where $|\cdot|$ signifies the determinant operator, and T denotes transpose. The symbol " | " in $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ characterizes conditional probability, and conveys that the distribution of Equation (1.6) is conditioned on the observed data, $\widetilde{\mathbf{Y}}$. Note that Equation (1.6) reduces to the univariate normal distribution if $\mathbf{C}(\boldsymbol{\theta}^*)$ is a 1×1 matrix (scalar). The $d\times d$ covariance matrix, $\mathbf{C}(\boldsymbol{\theta}^*)$, can be derived from the model parameter sensitivity matrix, $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$ as follows

$$\mathbf{C}(\boldsymbol{\theta}^*) = \sigma^2_{\mathbf{E}}\big(\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)^T\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)\big)^{-1}, \tag{1.7}$$

where $\sigma^2_{\mathbf{E}} = \mathrm{SSR}/(n-d)$ denotes the variance of the residuals, and the symbol $^{-1}$ signifies matrix inverse. The transpose operator acting on the first of two $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$'s (term between brackets) enforces equal inner and outer dimensions of n and d, respectively, so that the matrix product produces, after inversion and multiplication with $\sigma^2_{\mathbf{E}}$, the $d\times d$ covariance matrix, $\mathbf{C}(\boldsymbol{\theta}^*)$.

The $n\times d$ Jacobian matrix, $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$, stores the first-order partial derivatives of the model output with respect to each of the parameters

$$\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*) = \frac{\partial\mathcal{M}(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}^*} = \begin{bmatrix}\frac{\partial\mathbf{Y}(\boldsymbol{\theta}^*)}{\partial\theta_1} & \cdots & \frac{\partial\mathbf{Y}(\boldsymbol{\theta}^*)}{\partial\theta_d}\end{bmatrix} = \begin{bmatrix}\frac{\partial y_1(\boldsymbol{\theta}^*)}{\partial\theta_1} & \cdots & \frac{\partial y_1(\boldsymbol{\theta}^*)}{\partial\theta_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n(\boldsymbol{\theta}^*)}{\partial\theta_1} & \cdots & \frac{\partial y_n(\boldsymbol{\theta}^*)}{\partial\theta_d}\end{bmatrix}. \tag{1.8}$$

The jth column of $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$, thus stores the sensitivity of the n elements of the model output $\mathbf{Y} = \{y_1,\ldots,y_n\}$ to the jth parameter. These columns are also referred to as basis functions. For models whose output $\mathbf{Y}$ depends linearly on the values of the parameters, the d basis functions of $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$ can be derived analytically. In fact, these basis functions will be constant and independent of $\boldsymbol{\theta}$, hence the first-order approximation of $\mathbf{C}$ in Equation (1.7) will be exact. For all other models, the different basis functions cannot be derived by analytical means nor by analytical approximation, and numerical simulation is required to approximate the Jacobian matrix. For

instance, we can approximate the first column of the sensitivity matrix, $\partial \mathbf{Y}(\boldsymbol{\theta}^*)/\partial\theta_1$, by perturbing the first parameter with a quantity, $\Delta\theta_1$ and calculating the resulting change in model output from the default (optimal) simulation with $\boldsymbol{\theta}^*$

$$\frac{\partial \mathbf{Y}(\boldsymbol{\theta}^*)}{\partial\theta_1} = \frac{\Delta \mathbf{Y}(\boldsymbol{\theta}^*)}{\Delta\theta_1} \approx \frac{\mathbf{Y}(\{\theta_1^* + \Delta\theta_1, \ldots, \theta_d^*\}) - \mathbf{Y}(\{\theta_1^*, \ldots, \theta_d^*\})}{\Delta\theta_1}. \tag{1.9}$$

This recipe is repeated for the other $d-1$ parameters (columns) of $\mathbf{J}_{\mathcal{M}}(\boldsymbol{\theta}^*)$. Equation (1.9) uses a one-sided interval to approximate the partial model derivatives. More accurate results will be obtained if a two-sided interval is used with $-\Delta\theta_1$ and $\Delta\theta_1$, yet this doubles the number of simulations.

For models with valid basis functions, the first-order approximation of Equation (1.6) constitutes an exact description of the parameter uncertainty. For all other models with invalid basis functions, the first-order approximation is only an approximation of the actual parameter uncertainty. This approximation can be deficient if the covariance matrix varies considerably over the domain of $\boldsymbol{\theta}$ for which there is significant uncertainty. This is more the rule than the exception for nonlinear system models, particularly when the actual $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ distribution exhibits is poorly described by a multivariate normal distribution due to the presence of multimodality, local minima, and strong nonlinear parameter interactions. What is more, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ can be truncated by the prior distribution.

1.4 Bayesian Inference

Bayesian inference allows for an exact description of parameter uncertainty (and other sources of uncertainty) by treating the parameters (and nuisance variables) as probabilistic variables with joint posterior probability density function, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. This multivariate distribution, the so-called posterior parameter distribution, is the consequence of two antecedents, a prior distribution which captures our initial degree of beliefs in the values of the model parameters, and a likelihood function which quantifies by the rules of probability theory the level of confidence (= conditional belief) in the parameter values, $\boldsymbol{\theta}$, in light of the observed data, $\widetilde{\mathbf{Y}}$, alone. Bayes' theorem (also referred to

as Bayes' law or Bayes' rule) expresses mathematically, and in a simple formula, the fundamental relationship between the prior, conditional, and posterior (= updated) beliefs of the parameters. The theorem is named after the English statistician, philosopher, and Presbyterian minister Thomas Bayes (1701-1761), who formulated the solution in written notes. These ideas were not published until two years after Bayes' death, when his work on inverse probability emerged posthumously in the *Philosophical Transactions of the Royal Society of London* in the masterwork "*An Essay towards solving a Problem in the Doctrine of Chances*" (Bayes and Price, 1763). In this publication, Bayes' relates the "direct" probability of a hypothesis conditional on some body of data to the "inverse" probability of the data conditional on the hypothesis (nowadays referred to as likelihood).

Bayes' theorem can be derived from the basic axioms of probability, specifically conditional probability, and reads in our application

$$P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = \frac{P(\boldsymbol{\theta})P(\widetilde{\mathbf{Y}}|\boldsymbol{\theta})}{P(\widetilde{\mathbf{Y}})}, \tag{1.10}$$

where $P(\boldsymbol{\theta})$ and $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ signify the prior and posterior parameter distribution, respectively, and $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \equiv P(\widetilde{\mathbf{Y}}|\boldsymbol{\theta})$ denotes the likelihood function. The model evidence, $P(\widetilde{\mathbf{Y}})$ (or marginal likelihood) acts as a normalizing constant (scalar)

$$P(\widetilde{\mathbf{Y}}) = \int_{\boldsymbol{\Theta}} P(\boldsymbol{\theta})P(\widetilde{\mathbf{Y}}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})\mathrm{d}\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} P(\boldsymbol{\theta}, \widetilde{\mathbf{Y}})\mathrm{d}\boldsymbol{\theta}, \tag{1.11}$$

so that the posterior distribution integrates to unity over the prior (feasible) parameter space, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \in \mathbb{R}^d$. Knowledge of $p(\widetilde{\mathbf{Y}})$ is strictly necessary for hypothesis testing to select the most plausible model of the real-world system $\Im$ from a set of different models deemed valid a-priori. We will briefly discuss the topic of model selection in this chapter. For now, we suffice to say that the evidence of a model is largest, if its data likelihood is high relative to other models and distributed uniformly over the prior parameter space, $\boldsymbol{\theta}$. The evidence estimates can also serve as weights to average the simulations of the different models, as in Bayesian model averaging.

If we rely on a single hypothesis, $\mathcal{M}(\cdot)$, of the system $\Im$ of interest, then the denominator, $P(\widetilde{\mathbf{Y}})$ in Equation (2.4), is of no particular interest as all statistical inferences about the parameters of

$\mathcal{M}(\cdot)$ can be made from the unnormalized posterior distribution

$$P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \tag{1.12}$$

The proportionality sign conveys that the posterior density of the unnormalized distribution is an unknown multiple (scalar) of the normalized density. This might be confusing, yet consider a empirical histogram of $M = 1000$ independent observations of some variable of interest. Whether we use for each bin units of frequency, relative frequency (= frequency/M) or probability density (histogram integrates to one), the 95% confidence intervals of the variable are unaffected. Thus, our inferences of the parameters are protected against linear transformations of the density.

Numerical implementation of the Bayesian paradigm in Equation (1.12) requires the user to specify a prior parameter distribution, $P(\boldsymbol{\theta})$ and the likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. The next two sections review these two antecedents.

1.4.1 The Prior Distribution

The prior distribution should encode all the "subjective" knowledge about the parameters, $\boldsymbol{\theta}$, before collection of the data, $\widetilde{\mathbf{Y}}$. This distribution, often simply called the prior, expresses one's beliefs about the parameters before the data (also referred to as evidence) is taken into account. The work by Berger (2013) describes up to ten different techniques to construct the prior distribution.

In general, a prior distribution can be construed on the basis of findings reported in the literature or other publications, past experimental data collected in the laboratory or field, or other direct and indirect information. This can include "soft" data based on qualitative knowledge and understanding of processes and/or system behavior. A prior can also be elicited from expert judgment, or guided by principles of symmetry (scale invariance) or information-theoretical arguments (maximum entropy). Examples of the latter include a Jeffreys prior (Jeffreys, 1946) and the reference prior (Berger et al., 2009; Bernardo, 1979). Finally, for certain choices of the prior distribution, the posterior distribution has the same algebraic form, possibly with different parameter values. For example, if the likelihood function is Gaussian, then a normal prior over the mean will

ensure that the posterior distribution is also Gaussian. Such priors, also called conjugate priors, can thus be deliberately chosen in lieu of analytical tractability. This avoids the need for numerical approximation of the posterior distribution using sampling methods (more of which later).

Prior distributions can be classified as informative or uninformative (or noninformative) depending on their information content for the model parameter or entity of interest. Figure 1.3 portrays graphically the two different types of prior distribution for some hypothetical model with just a handful of unknown parameters. Also shown is their anticipated effect on the marginal posterior distribution of the first parameter for a fixed likelihood function.

The uninformative prior expresses vague and general information about a variable. All its values are deemed equally likely a-priori. Yet, an uninformative prior can communicate objective information on the ranges of the quantity of interest. The classification "uninformative" is therefore somewhat of a misnomer, instead the wording diffuse prior might seem more appropriate. An informative prior expresses specific, or definite, information about a variable of interest. Such distribution voices preference to certain values of the parameter. The effect of this is visible on the posterior distribution as it appears more condensed with a informative prior than with a uninformative prior distribution. This is not always the case, certainly not if the prior distribution and the likelihood function are in disagreement on the statistical distribution of the parameters. The uninformative prior is also referred to as flat or uniform prior.

For the time being, we assume conveniently the use of a multivariate uniform prior distribution, $\mathcal{U}_d(\mathbf{a}, \mathbf{b})$, where $\mathbf{a}$ and $\mathbf{b}$ are d-vectors with lower and upper bound values of the d parameters of $\boldsymbol{\theta}$, respectively

$$P(\boldsymbol{\theta}) \overset{\mathcal{D}}{\sim} \mathcal{U}_d(\mathbf{a}, \mathbf{b}), \tag{1.13}$$

where the vectors $\mathbf{a}$ and $\mathbf{b}$ are defined by the d-dimensional hypercube, $\boldsymbol{\theta}$, or $a_j = \min(\Theta_j)$ and $b_j = \max(\Theta_j)$ where $j = \{1, \ldots, d\}$. The density of the uninformative prior distribution, $P(\boldsymbol{\theta})$, in Equation (1.13) is constant and independent of $\boldsymbol{\theta}$.

The use of a multivariate prior distribution is appropriate when all the different parameters of the model have a similar marginal prior distribution. Examples of such priors include the

multivariate Gaussian, $\mathcal{N}_d(\cdot)$, the multivariate uniform, $\mathcal{U}_d(\cdot)$, the multivariate Student, $\mathcal{T}_d(\cdot)$, the multivariate Gamma, and the Wishart distribution. All these constitute informative priors - with the exception of $\mathcal{U}_d(\cdot)$ (of course one can make each of these distributions almost flat by using infinitely large parameter variances). One of the advantages of a multivariate prior is that they can honor explicitly (linear) correlations among the model parameters (as off-diagonal terms in $d \times d$ covariance matrix). A disadvantage is that they assume the same type of marginal distribution for each of the parameters. This might not be appropriate for multivariate cases with differing individual priors. As alternative, we could use a different univariate distribution, $P(\theta_j) \sim \mathcal{X}(\cdot)$, for each of the parameters, $j = \{1, \ldots, d\}$. This univariate approach enhances considerably our freedom in picking suitable marginal prior distributions, yet assumes parameter independence. The joint density of $P(\theta_j)$'s is therefore equivalent to the product of their individual densities, $P(\boldsymbol{\theta}) = P(\theta_1) \times \cdots \times P(\theta_d)$. This approach allows using at the same time informative and uninformative prior distributions. For the uniform case of Equation (1.13) the joint density can be written as

$$P(\boldsymbol{\theta}) = P(\theta_1) \times \ldots \times P(\theta_d) = c_1 \times \ldots \times c_d \propto 1 \tag{1.14}$$

where each constant, $c_i; i = \{1, \ldots, d\}$, is equivalent to the reciprocal of the range of each parameter. This ensures that the prior distribution, $P(\boldsymbol{\theta})$, integrates to one, $\int_{\boldsymbol{\Theta}} P(\boldsymbol{\theta}) = 1$, and as such is a formal probability distribution. The effect of normalization is clearly visible in Figure 1.3 wherein the larger support of the uninformative (uniform) prior distribution (on right) is counteracted by a density that is much smaller than that of the informative Gaussian distribution (on left). In practice, however we are allowed to work with "improper" priors that do not integrate to unity, as long as we are focused solely on inference of the model parameters. For multiple different working hypotheses, the prior must integrate to one, otherwise the marginal likelihood, $P(\widetilde{\mathbf{Y}})$ is corrupted and we cannot proceed with model selection.

1.4.2 The Likelihood Function

Now the prior distribution has been defined, we are left with the definition of the likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. This function summarizes, in a probabilistic sense, the compatibility of the n observed

data, $\widetilde{\mathbf{Y}}$, and the n model output, $\mathbf{Y} = \mathcal{M}(\boldsymbol{\theta})$ simulated by the parameter values, $\boldsymbol{\theta}$. Likelihood functions play a key role in statistical inference, and the word "likelihood" is used often as synonym for "probability" - yet, in practice the word *probability* is appropriate when describing possible future outcomes for fixed parameter values before data are available. The use of *likelihood*, on the contrary, is appropriate to describe a function of a parameter vector for a given outcome after data are available. If we want to quantify the likelihood of an outcome b in light of some observation, a, then we need to define a probability density function, $f_a(b)$, for the entity a. For example, if $f_a(b)$ is a normal distribution, $f_a(b) \sim \mathcal{N}(a, c)$ then we can calculate the density of this distribution at our outcome b once the (measurement error) variance c of the observation a is known. This computation of the likelihood is easily generalized to a vector of observations, and leaves us with n likelihoods of $\mathbf{Y}(\boldsymbol{\theta})$ evaluated at $\widetilde{\mathbf{Y}}$ using the different distributions of $f_{\widetilde{y}_j}(y_j)$, where $j = \{1, \ldots, n\}$. In practice, it is more insightful to express the likelihood as a function of the residuals, $\mathbf{E}(\boldsymbol{\theta})$, instead rather than the observed and simulated values. This does not affect the actual likelihood values, except centers $f_a(b)$ on $a = 0$ and uses as entry the residual, $b = a - b$, between the observation and outcome. We therefore use instead the notation $f\big(e_j(\boldsymbol{\theta})\big)$ in the remainder of this paper.

We cannot proceed further without making some important assumptions regarding the dependence structure of the n different residuals (and thus likelihoods). For the time being we assume conveniently that the residuals are serially uncorrelated. Then the joint likelihood, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, can be written in multiplicative form as follows

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = f\big(e_1(\boldsymbol{\theta})\big) \times \ldots \times f\big(e_n(\boldsymbol{\theta})\big) = \prod_{t=1}^{n} f\big(e_t(\boldsymbol{\theta})\big), \tag{1.15}$$

where $f(b)$ signifies the zero-mean probability density function evaluated at b. If serial correlation is expected among the residuals, then this can be explicitly accounted for with an autoregressive model of the residuals in the computation of the joint probability density of the n different "events" (likelihoods). The first and/or higher-order correlation coefficients of this model are commonly unknown and join as nuisance variables the inference of $\boldsymbol{\theta}$. The likelihood then acts on the decorrelated residuals. We will revisit serial dependence of the residuals in a later part of this

section. Note, that this approach reduces automatically to Equation (1.15) when the autocorrelation coefficients take on values of zero.

Now we have an expression for the joint likelihood of the residuals (and thus parameters, $\boldsymbol{\theta}$) we need to make an assumption regarding the statistical distribution of the different fs. In the ideal case with a perfect model, input data, and initial states, we certainly expect the residual distribution to match perfectly the distribution of the measurement errors of the system response, $\widetilde{\mathbf{Y}}$. A typical assumption is that these measurement errors follow a Gaussian distribution, $\mathcal{N}(0, \widehat{\sigma}^2)$ with constant variance $\widehat{\sigma}^2$ - and thus $f\big(e_j(\boldsymbol{\theta})\big) \overset{\mathcal{D}}{\sim} \mathcal{N}(0, \widehat{\sigma}^2)\ \forall j \in \{1, \ldots, d\}$. If we substitute $\mathcal{N}(0, \widehat{\sigma}^2)$ in Equation (1.15) then the joint likelihood of the n residuals is simply equivalent to the product of n normal densities

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp\left[-\frac{1}{2}\Big(\frac{\widetilde{y}_t - y_t(\boldsymbol{\theta})}{\widehat{\sigma}}\Big)^2\right]. \tag{1.16}$$

Mathematics teaches us that

$$\prod_{t=1}^{n} \frac{1}{a} = a^{-n} \qquad \text{and} \qquad \prod_{t=1}^{n} \exp(-a_t) = \exp\left(-\sum_{t=1}^{n} a_t\right) \tag{1.17}$$

so we can simplify Equation (2.5) to read

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2) = (\sqrt{2\pi\widehat{\sigma}^2})^{-n} \exp\left(-\frac{1}{2}\widehat{\sigma}^{-2} \sum_{t=1}^{n} \big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2\right), \tag{1.18}$$

where the summation term in the exponent is equivalent to the SSR used commonly as objective function, $F(\boldsymbol{\theta})$, in model calibration. The better the model fits the data, the larger the value of the joint likelihood, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2)$ in Equation (2.5). The vector of parameter values that maximizes the likelihood function is also referred to as maximum likelihood (ML) solution.

The formulation of the likelihood function of Equation (2.5) can suffer from arithmetic underflow, that is, finite multiplication can result in a number that is so close to zero that the computer cannot store this in memory. This can already happen for relatively small n, say $n = 500$, particularly if the

model describes poorly the observed data, and the residuals are large compared to the measurement error standard deviation, $|e_t(\boldsymbol{\theta})| \gg \widehat{\sigma}$ for many elements $t \in \{1, \ldots, n\}$. For reasons of numerical stability it is therefore convenient to work with the log-likelihood, $\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2)$ instead

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}n\log(\widehat{\sigma}^2) - \frac{1}{2}\widehat{\sigma}^{-2}\sum_{t=1}^{n}\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2. \tag{1.19}$$

This log-likelihood formulation is arguably also easier to interpret algebraically. The value of $\widehat{\sigma}^2$ can be defined a-priori by the user or alternatively its value can be inferred simultaneously with the parameters, $\boldsymbol{\theta}$. As last resort we can "integrate out" the measurement error variance in Equation (1.19) using as proxy for $\widehat{\sigma}^2$ the variance s^2 of the error residuals

$$s^2 = \frac{1}{n-1}\sum_{t=1}^{n}\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2. \tag{1.20}$$

We can substitute for $\widehat{\sigma}^2$ the sufficient statistic s^2. This gives us

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}n\log\left(\frac{1}{n-1}\sum_{t=1}^{n}\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2\right) - \frac{1}{2}\sum_{t=1}^{n}\frac{\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2}{\frac{1}{n-1}\sum_{t=1}^{n}\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2}, \tag{1.21}$$

and with $\log(ab) = \log(a) + \log(b)$ this results in

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}n\log\left(\frac{1}{n-1}\right) - \frac{1}{2}n\log\left(\sum_{t=1}^{n}\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2\right) - \frac{1}{2}(n-1). \tag{1.22}$$

A further simplification of

$$\frac{1}{2}n\log\left(\frac{1}{n-1}\right) = \frac{1}{2}n\left(\log(1) - \log(n-1)\right) = -\frac{1}{2}n\log(n-1) \tag{1.23}$$

leads to

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = -\frac{1}{2}n\log(2\pi) + \frac{1}{2}n\log(n-1) - \frac{1}{2}n\log\left(\sum_{t=1}^{n}\left(\widetilde{y}_t - y_t(\boldsymbol{\theta})\right)^2\right) - \frac{1}{2}(n-1) \tag{1.24}$$

We can safely discard those terms from Equation (1.24) that are independent of $\boldsymbol{\theta}$. These terms are normalization constant of $\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. This includes the first, second and fourth term at the right-hand-side, respectively. Without these constants the log-likelihood function reads

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto -\frac{1}{2}n\log\left(\sum_{t=1}^{n}\left(\widetilde{y}_t - y_t(\boldsymbol{\theta})\right)^2\right) \tag{1.25}$$

and the proportionality sign is used as expression for the unnormalized likelihood. This equation is equivalent to

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto \log\left(\sum_{t=1}^{n}\left(\widetilde{y}_t - y_t(\boldsymbol{\theta})\right)^2\right)^{-\frac{1}{2}n}. \tag{1.26}$$

If we are interested in the actual likelihood, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, we end up with

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto \left(\sum_{t=1}^{n}\left(\widetilde{y}_t - y_t(\boldsymbol{\theta})\right)^2\right)^{-\frac{1}{2}n}, \tag{1.27}$$

which is similar to

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto \sum_{t=1}^{n}|\widetilde{y}_t - y_t(\boldsymbol{\theta})|^{-n} \tag{1.28}$$

If we now apply Bayes theorem, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, and assume a uniform prior distribution of the parameters, then the posterior density is equivalent to

$$P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto \sum_{t=1}^{n}|\widetilde{y}_t - y_t(\boldsymbol{\theta})|^{-n} \tag{1.29}$$

This concludes the derivation.

The derivation of Equation (1.29) assumes that the measurement errors of the system response data, $\widetilde{\mathbf{Y}}$ exhibits a constant variance, $\widehat{\sigma}^2$. This assumption might be appropriate for variables such

as temperature and pressure which are known to have a homoscedastic measurement error. This assumption, however is not justified for entities such as windspeed and discharge as the variance of their measurement error increases with their magnitude. We can modify Equation (2.5) to take into account a non-constant measurement error variance

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\sigma}}^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}_t^2}} \exp\left[-\frac{1}{2}\Big(\frac{\widetilde{y}_t - y_t(\boldsymbol{\theta})}{\widehat{\sigma}_t}\Big)^2\right], \tag{1.30}$$

where $\widehat{\boldsymbol{\sigma}}^2 = \{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2\}$ denotes the n-vector of measurement error variances. The log-likelihood of Equation (1.30) now becomes

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\sigma}}^2) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\{\log(\widehat{\sigma}^2)\} - \frac{1}{2}\sum_{t=1}^{n}\left(\frac{\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big)^2}{\widehat{\sigma}_t^2}\right). \tag{1.31}$$

This allows for a different value of the measurement error variance for each observation of $\widetilde{\mathbf{Y}}$.

Now lets imagine a situation in which the residuals $\mathbf{E}(\boldsymbol{\theta}) = \widetilde{\mathbf{Y}} - \mathbf{Y}(\boldsymbol{\theta}) = \{e_1(\boldsymbol{\theta}), \ldots, e_n(\boldsymbol{\theta})\}$ exhibit temporal correlation. This serial correlation, also known as autocorrelation, can exist between values at different times, as a function of the two times, or of the time lag. Lets assume conveniently that the residual mean, $\mu_{\mathbf{E}(\boldsymbol{\theta})}$, and variance, $\sigma^2_{\mathbf{E}(\boldsymbol{\theta})}$ are time invariant. Then, $\mathbf{E}(\boldsymbol{\theta})$ is a wide-sense stationary process, and the (auto)correlation of two residuals, $e_i(\boldsymbol{\theta})$ and $e_j(\boldsymbol{\theta})$, is a function only of the time lag $k = i - j$ between i and j ($i \geq j$). The (auto)correlation coefficient, $\rho(k) \in [-1, 1]$ is then mathematically defined as follows

$$\begin{aligned} \rho(k) &= \frac{\mathbb{E}[\big(e_t(\boldsymbol{\theta}) - \mu_{\mathbf{E}(\boldsymbol{\theta})}\big)\big(e_{t+k}(\boldsymbol{\theta}) - \mu_{\mathbf{E}(\boldsymbol{\theta})}\big)]}{\sigma^2_{\mathbf{E}(\boldsymbol{\theta})}} \\ &= \frac{\sum_{t=1}^{n-k}\big(e_t(\boldsymbol{\theta}) - \mu_{\mathbf{E}(\boldsymbol{\theta})}\big)\big(e_{t+k}(\boldsymbol{\theta}) - \mu_{\mathbf{E}(\boldsymbol{\theta})}\big)}{\sum_{t=1}^{n-k}\big(e_t(\boldsymbol{\theta}) - \mu_{\mathbf{E}(\boldsymbol{\theta})}\big)^2}, \end{aligned} \tag{1.32}$$

where $\mu_{\mathbf{E}(\boldsymbol{\theta})} = \frac{1}{n}\sum_{t=1}^{n} e_t(\boldsymbol{\theta})$. A value of $\rho(k) = 1$ indicates perfect positive correlation, whereas a value of $\rho(k) = -1$ signifies perfect anticorrelation. Two words of caution. First, the (auto)correlation coefficient characterizes only linear relationships between the residuals. Second, the residual variance, $\sigma^2_{\mathbf{E}(\boldsymbol{\theta})}$ must be stable (homogeneity assumption) and larger than zero.

The value of $\rho(k)$ is also referred to as the kth-order (auto)correlation, and can be depicted graphically in a so called autocorrelation function (see Figure 2.3). The autocorrelation is particularly significant at the first few lags, but then deteriorates rapidly at larger distances between the residuals. Residuals that are more than $k = 11$ lags apart appear, on average, uncorrelated. Thus, the temporal correlation, $\rho(k)$, of the residuals depends only on the "distance" between two residuals and not on their actual position in the time series of $\mathbf{E}(\boldsymbol{\theta})$.

We can take into explicit account serial correlation of the residuals in the derivation of the log-likelihood function. For example, lets suppose that the residuals exhibit correlation at the first lag, that is $k = 1$. We can write this serial correlation as an AR(1)-process

$$e_t(\boldsymbol{\theta}) = c + \phi_1 e_{t-1}(\boldsymbol{\theta}) + \epsilon_t, \tag{1.33}$$

where c signifies the bias, $\phi_1 \in [-1, 1]$ is the first-order correlation coefficient, and ϵ_t denotes the remaining errors, $\boldsymbol{\epsilon}(\boldsymbol{\theta}) = \{\epsilon_1, \ldots, \epsilon_n\}$, hereafter also referred to as the "decorrelated" residuals. If we assume that $\epsilon_t \overset{\mathcal{D}}{\sim} \mathcal{N}(0, \widehat{\sigma}_t^2)$, then the expectation $\mathbb{E}\big[e_t(\boldsymbol{\theta})\big] = c/(1-\phi_1)$, and central dispersion $\mathrm{Var}\big[e_t(\boldsymbol{\theta})\big] = \widehat{\sigma}^2/(1-\phi_1^2)$. To illustrate the effect of autocorrelation, please consider Figure 2.5 which displays two different residual time series with (orange line) and without (green line) first-order serial correlation. The differences between the two residual vectors are evident. The uncorrelated residuals traverse randomly up and down the zeroth line (dotted gray line) and do not display any "collective" memory or consciousness. The correlated residuals, on the contrary, show a much strong memory effect with neighboring residuals that take on very similar values.

Lets assume that the residuals do not exhibit a systematic bias, thus $c = 0$. Per Equation (1.33) with $c = 0$ the values of ϵ_t are equivalent to

$$\begin{aligned} \epsilon_t(\boldsymbol{\theta}, \phi_1) &= e_t(\boldsymbol{\theta}) - \phi_1 e_{t-1}(\boldsymbol{\theta}) \\ &= \big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big) - \phi_1\big(\widetilde{y}_{t-1} - y_{t-1}(\boldsymbol{\theta})\big), \end{aligned} \tag{1.34}$$

with $\mathrm{Var}\big[\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big) - \phi_1\big(\widetilde{y}_{t-1} - y_{t-1}(\boldsymbol{\theta})\big)\big] = \mathrm{Var}\big[\epsilon_t(\boldsymbol{\theta}, \phi_1)\big] = \widehat{\sigma}^2$. As we do not have available $\widetilde{y}_0$ we cannot derive in a single step the log-likelihood function of the decorrelated residuals, $\boldsymbol{\epsilon}$. Instead, lets ignore for now $y_1(\boldsymbol{\theta})$ and focus on the last $n-1$ simulated values of $\mathbf{Y}(\boldsymbol{\theta})$. Per Equation (1.31)

the log-likelihood of $\{\epsilon_2(\boldsymbol{\theta},\phi_1),\ldots,\epsilon_n(\boldsymbol{\theta},\phi_1)\}$ equates to

$$\mathcal{L}(\boldsymbol{\theta}|\{\widetilde{y}_2,\ldots,\widetilde{y}_n\},\phi_1,\{\widehat{\sigma}_2^2,\ldots,\widehat{\sigma}_n^2\}) = -\frac{1}{2}(n-1)\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\{\log(\widehat{\sigma}_t^2)\}$$
$$- \frac{1}{2}\sum_{t=2}^{n}\left(\frac{\Big(\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big) - \phi_1\big(\widetilde{y}_{t-1} - y_{t-1}(\boldsymbol{\theta})\big)\Big)^2}{\widehat{\sigma}_t^2}\right). \quad (1.35)$$

which is identical to

$$\mathcal{L}(\boldsymbol{\theta}|\{\widetilde{y}_2,\ldots,\widetilde{y}_n\},\phi_1,\{\widehat{\sigma}_2^2,\ldots,\widehat{\sigma}_n^2\}) = -\frac{1}{2}(n-1)\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\{\log(\widehat{\sigma}_t^2)\}$$
$$- \frac{1}{2}\sum_{t=2}^{n}\left(\frac{\epsilon_t^2(\boldsymbol{\theta})}{\widehat{\sigma}_t^2}\right). \quad (1.36)$$

We are now left with the log-likelihood of the first simulated value, $y_1(\boldsymbol{\theta})$, of $\mathbf{Y}(\boldsymbol{\theta})$. We know that the $\text{Var}\big[e_1(\boldsymbol{\theta})\big] = \widehat{\sigma}_1^2/(1-\phi_1^2)$ and thus the log-likelihood of $y_1(\boldsymbol{\theta})$ is

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{y}_1,\phi_1,\widehat{\sigma}_1^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\widehat{\sigma}_1^2}{(1-\phi_1^2)}\right) - \frac{1}{2}\left(\frac{\big(\widetilde{y}_1 - y_1(\boldsymbol{\theta})\big)^2}{\widehat{\sigma}_1^2/(1-\phi_1^2)}\right). \quad (1.37)$$

As $\log(a/b) = \log(a) - \log(b)$ this equation can be rearranged and simplified to

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{y}_1,\phi_1,\widehat{\sigma}_1^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\widehat{\sigma}_1^2) + \frac{1}{2}\log(1-\phi_1^2)$$
$$- \frac{1}{2}(1-\phi^2)\widehat{\sigma}_1^{-2}\big(\widetilde{y}_1 - y_1(\boldsymbol{\theta})\big)^2. \quad (1.38)$$

The joint log-likelihood of the first-order correlated residuals, $\mathcal{L}(\boldsymbol{\theta}|\{\widetilde{\mathbf{Y}},\phi_1,\widehat{\boldsymbol{\sigma}}^2)$ is now equivalent to the sum of $\mathcal{L}(\boldsymbol{\theta}|\widetilde{y}_1,\phi_1,\widehat{\sigma}_1^2)$ and $\mathcal{L}(\boldsymbol{\theta}|\{\widetilde{y}_2,\ldots,\widetilde{y}_n\},\phi_1,\{\widehat{\sigma}_2^2,\ldots,\widehat{\sigma}_n^2\})$ which yields

$$\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}},\phi_1,\widehat{\boldsymbol{\sigma}}^2) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\{\log(\widehat{\sigma}_t^2)\} + \frac{1}{2}\log(1-\phi_1^2)$$
$$- \frac{1}{2}(1-\phi^2)\widehat{\sigma}_1^{-2}\big(\widetilde{y}_1 - y_1(\boldsymbol{\theta})\big)^2 - \frac{1}{2}\sum_{t=2}^{n}\left(\frac{\Big(\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big) - \phi_1\big(\widetilde{y}_{t-1} - y_{t-1}(\boldsymbol{\theta})\big)\Big)^2}{\widehat{\sigma}_t^2}\right). \quad (1.39)$$

This Equation reduces to the Gaussian likelihood function of Equation (1.31) if $\phi_1 = 0$, that is, the residuals, $\mathbf{e}(\boldsymbol{\theta})$, do not exhibit serial correlation. We can generalize further Equation (1.39) by explicit treatment of bias in the residuals. We will not present this derivation herein as it follows exactly the previous steps, except that $c \neq 0$. The log-likelihood becomes

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, c, \phi_1, \widehat{\boldsymbol{\sigma}}^2) = &-\frac{1}{2}n\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\{\log(\widehat{\sigma}_t^2)\} + \frac{1}{2}\log(1-\phi_1^2) \\ &- \frac{1}{2}(1-\phi^2)\widehat{\sigma}_1^{-2}\big(\widetilde{y}_1 - y_1(\boldsymbol{\theta}) - \underbrace{[c/(1-\phi)]}_{\mathbb{E}\left[e_1(\boldsymbol{\theta})\right]=c/(1-\phi_1)}\big)^2 \\ &- \frac{1}{2}\sum_{t=2}^{n}\left(\frac{\Big(\big(\widetilde{y}_t - y_t(\boldsymbol{\theta})\big) - c - \phi_1\big(\widetilde{y}_{t-1} - y_{t-1}(\boldsymbol{\theta})\big)\Big)^2}{\widehat{\sigma}_t^2}\right). \quad (1.40)\end{aligned}$$

This concludes the derivation of the different likelihood functions. The nuisance variables[1] ϕ_1, c, and the n-vector of measurement error variances, $\widehat{\boldsymbol{\sigma}}^2$, can be defined a-priori by the user, or alternatively, their values can be inferred jointly with the parameters, $\boldsymbol{\theta}$. If the system response data, $\widetilde{\mathbf{Y}}$, exhibit heteroscedastic measurement errors then this could be a pitfall as the number of unknown variables of $\widehat{\boldsymbol{\sigma}}^2$ grows linearly with n. A pragmatic remedy to this problem is to relate $\widehat{\sigma}_t$ to the measured data, $\widetilde{y}_t$, using some predefined measurement error model

$$\widehat{\sigma}_t = \sigma_0 + \sigma_1\widetilde{y}_t. \quad (1.41)$$

where $\sigma_0 > 0$ and $\sigma_1 \in [0, 1]$ are two unknown coefficients that define the intercept and slope of the measurement error function. This approach reduces the number of nuisance variables to four, that is, $\boldsymbol{\alpha} = \{\sigma_0, \sigma_1, c, \phi_1\}$. Nonlinear measurement error models can be used as well, whatever is deemed appropriate in practice. The inference thus involves the estimation of the posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$ using the model, $\mathcal{M}(\boldsymbol{\theta})$ and observed data, $\widetilde{\mathbf{Y}}$.

[1] A nuisance variable is a random variable that is fundamental to the probabilistic model, but that is not of particular itself.

1.4.3 The Posterior Distribution

Now we have defined the prior distribution, $P(\boldsymbol{\theta})$, and the likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, we are left with inference of the (unnormalized) posterior distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, which summarizes our updated knowledge (belief) on the parameters, $\boldsymbol{\theta}$. Unfortunately, in most applications of Bayes' theorem the posterior distribution does not have a closed-form (compact) analytical solution, and we have to resort to sampling methods to approximate the posterior distribution. The underlying principles of this approach are explained in Figure 1.6 using a scatter plot of $M = 160$ samples drawn randomly from a bivariate normal distribution, $\mathcal{N}_2(\mathbf{a}, \Sigma)$ with mean $\mathbf{a} = \{a_1, a_2\}$ and 2×2 covariance matrix,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

As the off-diagonal entries of $\boldsymbol{\Sigma}$ are set to zero, the two parameters of the distribution, θ_1 and θ_2, are independent, and consequently $P(\theta_1) \sim \mathcal{N}(a_1, \sigma_1^2)$ and $P(\theta_2) \sim \mathcal{N}(a_2, \sigma_2^2)$. The plotted samples exhibit several key features. First, the scatter of points exhibits significant variations in sampling density. The sample density is largest in the center of the cloud, and decreases slowly in all radial directions away from this midpoint. Second, a circular pattern emerges of points with equal sampling density. Third, the sampled points do not express a preferred orientation of θ_1 and θ_2. The lack of linear dependence is exemplary for uncorrelated variables (parameters). Fourth, the peaks of the inferred frequency distributions of θ_1 (green line) and θ_2 (blue line) are within the bins of maximum sample density. This coincides with the nucleus (heart) of the point cloud. Finally, the frequency distributions of the two parameters are symmetric and well described with a Gaussian distribution. The key notion of these findings is that a) the sampling density is a proxy for the probability density of the target distribution, b) the orientation of the samples is a measure of the correlation among the variables of the target distribution, and c) the dispersion of the samples is a measure of the variances of the individual variables of the target distribution.

This simple example with the point cloud shows that we can represent any probability distribution with a large number of samples as long as the sampled points satisfy one crucial requirement

and that is that they are distributed exactly according to the underlying density function of the probability distribution. In other words, the sample density at any point of the target distribution must match exactly the density of the distribution at that point. Then, the inferred marginal distributions of the parameters will match their counterparts of the target distribution of interest and the covariance (correlation) structure of the parameters is honored perfectly. Indeed, the inferred distributions of θ_1 and θ_2 in Figure 1.6 match exactly their "true" counterparts, $P(\theta_1)$ and $P(\theta_2)$. The number of samples that is required to represent properly a multivariate probability distribution depends on the shape and dimensionality of this distribution, more of which later.

These results now beg the question of how we should generate the samples? We cannot simply use a random number generator as the resulting samples will likely not be distributed exactly according to the target distribution of interest. Many different methods have been developed to generate samples of an unknown distribution. All these methods rely in some way on Monte Carlo simulation. In the next sections we discuss the application of these methods to approximate the d-variate posterior distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. This distribution, also referred to as the target or limiting distribution, is often high dimensional. The different methods assume a continuous parameter space $\boldsymbol{\theta} \in \boldsymbol{\Theta} \in \mathbb{R}^d$, yet with simple modifications can be used to approximate discrete target distributions.

1.5 Monte Carlo approximation

Monte Carlo methods are a broad class of computational algorithms that use repeated random sampling to approximate some arbitrary d-variate distribution, $F(\mathbf{x})$ with probability density function, $f(\mathbf{x})$. This unknown multivariate distribution, $F(\mathbf{x})$, is equivalent to the posterior distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. For the time being, we use the symbol $\mathbf{x} = \{x_1, \ldots, x_d\}$ to denote the d-variables of the unknown distribution, $F(\mathbf{x})$, which, in our application, constitute the model parameters, $\boldsymbol{\theta}$, possibly augmented with nuisance variables, thus $\mathbf{x} = \{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$.

The basic idea of Monte Carlo methods is to use an alternative distribution, $Q(\mathbf{x})$, which is easy to sample from in practice, and whose probability density function, $q(\mathbf{x})$, approximates as

closely and consistently as possible the unknown density, $f(\mathbf{x})$. The "known" distribution $Q(\mathbf{x})$ is also referred to as the proposal distribution and serves as catalyst to approximate the posterior distribution.

1.5.1 Rejection Sampling

The earliest Monte Carlo method is the acceptance-rejection algorithm (also referred to as rejection sampling) and produces M different samples of the desired target distribution, $F(\mathbf{x})$. The various steps of this method are summarized in the numerical recipe of Algorithm 1, wherein the label Z signifies a draw from the standard uniform distribution, $Z \sim \mathcal{U}(0,1)$. After a sufficient number

Algorithm 1 Rejection Sampling

1: Define a proposal distribution, $Q(\mathbf{x})$, so that $q(\mathbf{x}) \geq f(\mathbf{x})$ if $f(\mathbf{x}) > 0$.
2: Define M and set counter, $i = 1$.
3: **while** $i < M$ **do**
4: Sample randomly a candidate point, $\mathbf{x}_\mathrm{p}$, from $Q(\mathbf{x})$, $\mathbf{x}_\mathrm{p} \sim Q(\mathbf{x})$ and calculate $f(\mathbf{x}_\mathrm{p})$.
5: Compute the acceptance probability, $P_\mathrm{acc}(\mathbf{x}_\mathrm{p}) = f(\mathbf{x}_\mathrm{p})/q(\mathbf{x}_\mathrm{p})$, of $\mathbf{x}_\mathrm{p}$.
6: If $Z \leq P_\mathrm{acc}(\mathbf{x}_\mathrm{p})$ then set $\mathbf{x}_{(i)} = \mathbf{x}_\mathrm{p}$ and counter, $i = i + 1$, otherwise reject $\mathbf{x}_\mathrm{p}$.
7: **end while**

of iterations, rejection sampling produces M different samples that will be distributed exactly according to the unknown distribution, $F(\mathbf{x})$ with underlying density, $f(\mathbf{x})$. Why the accepted samples converge exactly to $F(\mathbf{x})$ is relatively easy to proof mathematically using the envelope principle. Here, instead we provide a visual explanation (see Figure 1.7) and draw samples from a uniform distribution (black line) to approximate the univariate target distribution, $F(x)$, with arbitrary density, $f(x)$ (gray line). It is not difficult to see that the acceptance probability, $P_\mathrm{acc}(x_\mathrm{p}) = f(x_\mathrm{p})/q(x_\mathrm{p})$, of each candidate point, $x_\mathrm{p} \in [1, 5]$, is directly proportional to the density, $f(x_\mathrm{p})$, of the target distribution. Thus if we accept each candidate point with probability, $P_\mathrm{acc}(x_\mathrm{p})$, we account for "bias" in the samples drawn from the proposal distribution, and distribute the accepted samples exactly according to the target density, $f(x)$ of $F(x)$. This does require that the entire target distribution is sampled, that is the space of $\mathbf{x}$ for which $f(\mathbf{x}) > 0$, and that $q(\mathbf{x}) \geq f(\mathbf{x})$, otherwise the acceptance probability can reach values larger than unity, thereby chopping off the peaks of the target distribution. In practice, a multiplier $c \in [1, \infty)$ is used

to inflate the density of the proposal distribution so that it always exceeds the density of the target distribution, $cq(\mathbf{x}) \geq f(\mathbf{x})$. This multiplier enters into the denominator of the acceptance probability, and should thus be chosen wisely, otherwise (e.g. too large) the rejection algorithm can become highly inefficient as most candidate points will be dismissed. The value of c can be determined as largest value of the ratios, $f(\mathbf{x}_\mathrm{p})/q(\mathbf{x}_\mathrm{p})$, of the many different $\mathrm{x_p}$'s. Then, the new acceptance probability, $P_\mathrm{acc} = f(\mathbf{x}_\mathrm{p})/\big(cq(\mathbf{x}_\mathrm{p})\big)$ of each sample x_p can be recomputed and the set of accepted samples re-construed. In practice, the "best" proposal density, $q(\mathbf{x})$, minimizes the value of the constant c, that is $c = \sup_\mathbf{x}\big(f(\mathbf{x})/q(\mathbf{x})\big)$. A perfect agreement between the proposal and target distribution equates to a value of $c = 1$.

The efficiency of rejection sampling depends in large part on the choice of the proposal distribution that is used to generate trial points. This distribution must satisfy two conditions that is a) it envelops the target distribution, and b) its density is at least equal to the density of the target distribution. These two conditions are very difficult to satisfy in practice, without detailed knowledge of the target distribution (as in our Bayes' application). This is especially true for high-dimensional targets with complicated multivariate relationships among the variables. A poorly construed proposal distribution has profound consequences as a (very) large portion of the candidate points will be rejected and go to waste. For high-dimensional target distributions, say $d = 50$ this can lead to acceptance rates on the order of say 0.1%. Consequently, it will take a very large number of iterations to generate a sufficient sample from the target distribution.

1.5.2 Importance Sampling

Importance sampling is an important improvement over the acceptance-rejection algorithm. This algorithm can be written in a few lines (see 'Algorithm 2) and is widely used to compute raw, central and standardized moments of the target distribution, $F(\mathbf{x})$. Whereas rejection sampling produces samples with equal weight that are distributed exactly according to the target distribution, importance sampling returns a collection of weighted samples. This weighted sample cannot be used to draw marginal and or joint histograms of the target distribution, but rather serves to

compute moments of $F(x)$ such as its mean, variance, skewness, and kurtosis. To generate a sample from $F(\mathbf{x})$ we must draw *with replacement* from the importance samples, $\{\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(M)}\}$, using selection probabilities proportional to the importance weights, $\{w(\mathbf{x}_{(1)}), \ldots, w(\mathbf{x}_{(M)})\}$. This process is also known as sampling importance re-sampling (SIR) and produces a collection of samples with equal weights distributed according to the target distribution, $F(\mathbf{x})$. Thus, points with relatively large importance weights have a much higher chance to be selected (with replacement) in this collection of equal weight points than samples with a negligible importance weight. The frequency of appearance of each importance sample in this re-sampled collection of points is thus proportional to the underlying density function, $f(\mathbf{x})$, of the target distribution, $F(\mathbf{x})$. If the importance sample is sufficiently large then re-sampling should provide a reasonable approximation of $F(\mathbf{x})$.

Importance sampling has two main advantages over rejection sampling. First, it does not produce waste as all samples are used to approximate the moments of the target distribution. Second, the density, $g(x)$, of the (importance) sampling distribution, $G(x)$, does not have to be equal to or larger than the target density. The only requirement is that $g(\mathbf{x}) > 0$ if $f(\mathbf{x}) > 0$. This simplifies considerable practical application. Nevertheless, the construction of a proper importance distribution is difficult without detailed knowledge of the target distribution (as in our Bayes' application), and becomes particularly cumbersome in high-dimensional parameter spaces. When the importance distribution is too wide, a large majority of the sampled points will receive negligible weights. On the other hand, when the importance distribution is too narrow, the sampled points will not characterize adequately the target distribution. Indeed, methods such as rejection sampling and importance sampling are rather frugal and inefficient for all but very low dimensional problems. Next, we therefore resort to an alternative class of methods to explore the target distribution.

1.5.3 Markov Chain Monte Carlo simulation

The basis of MCMC simulation is a Markov chain that generates a random walk through the search space and successively visits solutions with stable frequencies stemming from a stationary distribution, $F(\mathbf{x})$. To explore the target distribution, $F(\mathbf{x})$, a MCMC algorithm generates trial

moves from the current state of the Markov chain $\mathbf{x}_{(i-1)}$ to a candidate point, $\mathbf{x}_{\mathrm{p}}$. The earliest MCMC approach is the random walk Metropolis (RWM) algorithm introduced by Metropolis et al. (1953). This scheme is constructed to maintain detailed balance with respect to $f(\mathbf{x})$ at each step in the chain. If $f(\mathbf{a})$ denotes the probability to find the system in state $\mathbf{a}$ and $q(\mathbf{a} \rightarrow \mathbf{b})$ is the conditional probability, $q(\mathbf{b}|\mathbf{a})$, to perform a trial move from $\mathbf{a}$ to $\mathbf{b}$, then the probability $P_{\mathrm{acc}}(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\mathrm{p}})$ to accept the trial move from $\mathbf{x}_{(i-1)}$ to $\mathbf{x}_{\mathrm{p}}$ is related to $P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}} \rightarrow \mathbf{x}_{(i-1)})$ according to

$$f(\mathbf{x}_{(i-1)})q(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\mathrm{p}})P_{\mathrm{acc}}(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\mathrm{p}}) = f(\mathbf{x}_{\mathrm{p}})q(\mathbf{x}_{\mathrm{p}} \rightarrow \mathbf{x}_{(i-1)})P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}} \rightarrow \mathbf{x}_{(i-1)}) \tag{1.42}$$

This principle of *detailed balance* originates from the work of Boltzmann and Maxwell on collision and gas kinetics, respectively. This condition implies that there is no net inflow or outflow of probability among some closed set of possible states, and consequently, there exists a unique equilibrium distribution of the states. Detailed balance is of particular relevance to a Markov chain as it guarantees that the chain, under some regularity conditions, will converge to the exact equilibrium (= target) distribution. Indeed, a chain that maintains detailed balance will visit each state, $\mathbf{a}$, of the stationary (equilibrium) distribution with frequency proportional to its underlying probability density, $f(\mathbf{a})$. This does require the chain to be *irreducible* (it is possible to transition, in one or more steps, from any state to another configuration) to be *a-periodic* (return to a state occurs at irregular times) and to be *positive recurrent* (positive probability to return to a state). For a properly constructed proposal distribution, these three conditions are usually satisfied in practice, except for trivial exceptions. Note that detailed balance is not a necessary condition for convergence of the Markov chain to the target distribution. A few examples will be given later.

If a symmetric jumping distribution is used, that is $q(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\mathrm{p}}) = q(\mathbf{x}_{\mathrm{p}} \rightarrow \mathbf{x}_{(i-1)})$, then it follows from Equation (1.42) that

$$\frac{P_{\mathrm{acc}}(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\mathrm{p}})}{P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}} \rightarrow \mathbf{x}_{(i-1)})} = \frac{f(\mathbf{x}_{\mathrm{p}})}{f(\mathbf{x}_{(i-1)})} \tag{1.43}$$

This Equation does not yet fix the acceptance probability. Metropolis et al. (1953) made the

following choice

$$P_{\text{acc}}(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\text{p}}) = \min\left[1, \frac{f(\mathbf{x}_{\text{p}})}{f(\mathbf{x}_{(i-1)})}\right], \tag{1.44}$$

to determine whether to accept a trial move or not. This selection rule of candidate points has become the basic building block of MCMC algorithms. Hastings (1970) has generalized Equation (1.44) to cases with non-symmetrical proposal distributions

$$P_{\text{acc}}(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\text{p}}) = \min\left[1, \frac{f(\mathbf{x}_{\text{p}})q(\mathbf{x}_{\text{p}} \rightarrow \mathbf{x}_{(i-1)})}{f(\mathbf{x}_{(i-1)})q(\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\text{p}})}\right], \tag{1.45}$$

wherein the forward jump, $\mathbf{x}_{(i-1)} \rightarrow \mathbf{x}_{\text{p}}$, and backward jump, $\mathbf{x}_{\text{p}} \rightarrow \mathbf{x}_{(i-1)}$, do not have equal probability, thus $q(\mathbf{x}_{\text{p}}|\mathbf{x}_{(i-1)}) \neq q(\mathbf{x}_{(i-1)}|\mathbf{x}_{\text{p}})$. This generalization is known as the Metropolis-Hastings (MH) algorithm and broadens significantly the type of proposal distribution that can be used for posterior inference.

Figure 1.8 depicts the evolution (trajectory) of a single Markov chain starting from an arbitrary initial state (black square) for some hypothetical $d = 2$-dimensional target distribution. The gray arrows denote the different trial moves (jumps) of the chain, most of which are accepted (green dots), and some of which are declined (red dots). If a proposal is accepted then the chain moves to this new position, otherwise the chain remains at its "old" state and this position (e.g. values of $\mathbf{x}$) is replicated in the Markov chain. After about thirteen steps, the chain has reached the stationary distribution (in orange). The subsequent positions of the chain are used to approximate the target distribution, $F(\mathbf{x})$. The samples which are stored in the chain have equal weights, and share in common with SIR that their frequency of appearance is directly proportional to the underlying density, $f(\mathbf{x})$, of the target distribution. One crucial difference with rejection and importance sampling is that the sampling (proposal) distribution, $q(\cdot)$, in MCMC simulation does not have to cover or envelope the target distribution. This proposal distribution simply travels with the state of the chain, and centers on the last position to create candidate points.

The core of the RWM algorithm can be written in just a few lines (see Algorithm 3) and requires a symmetric jumping distribution, $q(\cdot)$, a uniform random number generator, $Z \sim \mathcal{U}(0, 1)$, and the

target density, $f(\mathbf{x})$, to create a chain trajectory. In words, assume that the points $\{\mathbf{x}_0, \ldots, \mathbf{x}_{(i-1)}\}$ have already been sampled, then the RWM algorithm proceeds as follows. First, a candidate point $\mathbf{x}_\mathrm{p}$ is sampled from a proposal (jumping) distribution, $q(\cdot)$, that depends on the present location, $\mathbf{x}_{(i-1)}$, and is symmetric, $q(\mathbf{x}_\mathrm{p}|\mathbf{x}_{(i-1)}) = q(\mathbf{x}_{(i-1)}|\mathbf{x}_\mathrm{p})$. Next, the candidate point is either accepted or rejected using the Metropolis acceptance probability (Equation 1.44). Finally, if the proposal is accepted the chain moves to $\mathbf{x}_\mathrm{p}$, otherwise the chain remains at its current location $\mathbf{x}_{(i-1)}$. Repeated application of these three steps results in a Markov chain which, under certain regularity conditions, has a unique stationary distribution with posterior probability density function, $f(\mathbf{x})$. In practice, this means that if one looks at the archived values of $\mathbf{x}$ in the chain sufficiently far from the arbitrary initial state, thus after a burn-in period, then these successively generated states will be distributed according to $F(\mathbf{x})$, the unknown target distribution of $\mathbf{x}$. Burn-in is required to allow the chain to explore the search space and reach its stationary regime (see Figure 1.8).

The RWM algorithm is relatively simple to implement, yet its efficiency is determined in large part by the choice of the proposal distribution, $q(\cdot)$ used to create trial moves (transitions) in the Markov chain. When the proposal distribution is too wide, too many candidate points are rejected, and therefore the chain will not mix efficiently and converge only slowly to the target distribution. On the other hand, when the proposal distribution is too narrow, most candidate points will be accepted, but the covered distance is so small that it will take a prohibitively large number of iterations before the chain has converged to the target distribution. The choice of proposal distribution is therefore of crucial importance and determines the computational cost and practical feasibility of MCMC simulation.

Note, that we assume a fixed computational budget of M iterations of the RWM algorithm. In practice, the chain will continue to evolve until it reaches a stationary distribution as judged by one or more convergence diagnostics (discussed later). The budget of iterations is thus determined on the fly based on the convergence properties of the sampled chain.

1.5.4 Automatic Tuning of Proposal Distribution

In the past decade, a variety of different approaches have been proposed to increase the efficiency of MCMC simulation and enhance the original RWM and MH algorithms. These approaches can be grouped into single and multiple chain methods.

Single-chain methods The most common adaptive single chain methods are the adaptive proposal (AP) (Haario et al., 1999), adaptive Metropolis (AM) (Haario et al., 2001) and delayed rejection adaptive metropolis (DRAM) algorithm, respectively. These methods simulate a single trajectory by drawing candidate points from a Gaussian proposal distribution, $\mathbf{x}_{\mathrm{p}} \sim \mathcal{N}_d(\mathbf{x}_{(i-1)}, s_{(d)}\boldsymbol{\Sigma})$, with covariance matrix, $\boldsymbol{\Sigma}$, which is updated periodically after every m iterations ($m \geq 1$) using all past samples stored in the chain, $\boldsymbol{\Sigma} = \mathrm{Cov}[\{\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(i-1)}\}] + \varphi\mathbf{I}_d$. The variable $s_{(d)}$ signifies the so-called jump rate, and depends on the dimensionality of the target distribution, $\mathbf{I}_d$ denotes the $d \times d$ identity matrix, and $\varphi = 10^{-6}$ is a small scalar that prevents the collapse of the sample covariance matrix to singularity (jumps become zero). This term also guarantees, at least in theory, that the sampled chain is irreducible, because of the unbounded support of the normal jumping distribution with nonsingular (invertible) covariance matrix, $\boldsymbol{\Sigma}$, that is $|\boldsymbol{\Sigma}| > 0$. As a basic choice, the scaling factor is chosen to be $s_{(d)} = 2.38^2/d$ which has proven optimal for Gaussian target and proposal distributions (Gelman et al., 1996) and should give an acceptance rate close to 0.44 for $d = 1$, 0.28 for $d = 5$ and 0.23 for large d.

A summary of the AM method appears below in algorithm 4, wherein the notion $\mathrm{mod}(a, b)$ signifies the modulo operator. This operator returns zero if the quotient k of $a > 0$ and $b > 0$, or $k = a/b$, equates to an integer, or $\mathrm{mod}(a, b) = 0$ if $k \in \mathbb{N}_+$. Thus, the AM algorithm is a special implementation of the RWM algorithm with a transient (multi)normal transition kernel as proposal distribution of the Markov chain. The covariance matrix of the Gaussian proposal distribution is adapted every m iterations using the archived chain samples. This adaptation enhances, sometimes dramatically, the convergence speed of the chain to the stationary distribution, as the jumps will align slowly with the orientation and scale of the target distribution. An important drawback of

adaptation is, however that the AM algorithm does not satisfy the detailed balance condition of Equation (1.42) at every step in the chain. This is easy to see if we compare the (multi)normal proposal distribution of the AM algorithm immediately before and after an update to $\mathbf{\Sigma}$ at iterations $j = km$, where $k \in \mathbb{N}_+$. Then the (conditional) probability, $q(\mathbf{x}_{(j)}|\mathbf{x}_{(j-1)})$, of the forward jump, $\mathbf{x}_{(j-1)} \rightarrow \mathbf{x}_{(j)}$ (with "old" $\mathbf{\Sigma}$), does not equate to the conditional probability, $q(\mathbf{x}_{(j-1)}|\mathbf{x}_{(j)})$, of the backward jump, $\mathbf{x}_{(j)} \rightarrow \mathbf{x}_{(j-1)}$ (with "new" $\mathbf{\Sigma}$). Detailed balance is rapidly restored during the subsequent $m-1$ iterations of the chain as the candidate points are created with a fixed $\mathbf{\Sigma}$. Nonetheless, the resulting chain simulated by the AM algorithm is not truely Markovian.

But why then does the AM algorithm converge to the appropriate limiting distribution? This is because of diminishing adaptation. In words, the transition kernel (multivariate normal distribution) of the AM algorithm converges to a fixed proposal distribution with increasing length of the chain. Indeed, the distance between successive values of $\mathbf{\Sigma}$ decreases to zero as the number of samples in the chain grows without bound. Note that the chain may converge to another than the target distribution if only the recent past is used to generate trial moves (see Haario et al. (2001) for an example). Another viable adaptation strategy is to fix the covariance matrix (say identity matrix) and to tune instead the scaling factor, $s_{(d)}$, during a burn-in period, until a desired acceptance rate is obtained (23% for large d). If adaptation is limited to the burn-in period then the chain transitions in the equilibrium distribution are fully Markovian.

The use of a multivariate normal transition kernel (with/without adaptation) may work well for Gaussian-like target distributions, but may not be adequate to characterize multimodal distributions with long tails, and possibly infinite first and second moments. Experience further suggests that single chain methods have a hard time to explore efficiently multidimensional parameter spaces, particularly when confronted with different regions of attraction and (numerous) local optima. The use of an overly dispersed proposal distribution will help to traverse difficult search spaces and/or sample disconnected modes, yet the resulting chain will converge only slowly due to an improper scaling of the jumps (excessively large). It is also particularly difficult to judge convergence of a single chain trajectory in absence of an independent benchmark against which we can compare the

statistics of the sampled values. Even the most powerful diagnostics do not protect us against a chain that has converged to a local basin of attraction in the parameter space or a chain that explores only one mode of the target distribution. Indeed, single-chain MCMC methods (e.g. DRAM, RWM, AM and AP) suffer many similar problems as local optimization methods (e.g. steepest descent, Newton method, Levenberg-Marquardt) and cannot guarantee an exhaustive exploration of the parameter space in pursuit of the target distribution.

Multiple chain methods: DE-MC Multiple chain MCMC methods simulate different trajectories in parallel to explore the posterior target distribution. This approach has several important advantages, particularly for search spaces with different regions of attraction and numerous local optima, and skewed, tailed, and multimodal target distributions with complex multivariate dependencies among the variables. The use of multiple different chains offers a robust protection against premature convergence, and opens up an array of powerful statistical tests to assess convergence of the sampled values to an equilibrium distribution. One example of an efficient multi-chain method is the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm (Vrugt et al., 2003). This method has found application in environmental modeling and builds on the shuffling concept of the SCE-UA optimization method to distribute, among the parallel chains, information gained about the search space. The use of this shuffling approach introduces a transient transition kernel which expedites convergence of SCEM-UA to the equilibrium distribution, but at the expense of exact reversibility of the sampled chain trajectories. Another novelty of the SCEM-UA algorithm is its explicit treatment of outlier chains, a necessity to traverse efficiently complex search spaces in lieu of the target distribution. If adaptation of the (multi)normal proposal distribution is restricted to the burn-in period only, then the chain transitions simulated by the SCEM-UA algorithm satisfy reversibility. The method then derives an efficient Gaussian proposal distribution for the standard Metropolis algorithm.

Differential Evolution Markov chain, or DE-MC, uses differential evolution as genetic algorithm for population evolution with a Metropolis selection rule to decide whether candidate points should replace their parents or not. In DE-MC, N different Markov chains are run simultaneously in

parallel. If the state of a single chain is given by the d-vector $\mathbf{x}$, then at each generation $i-1$ the N chains in DE-MC define a $N \times d$ matrix (population), $\mathbf{X}_{(i-1)} = \{\mathbf{x}^1_{(i-1)}, \ldots, \mathbf{x}^N_{(i-1)}\}$, with each chain as a row. Then multivariate proposals, $\mathbf{x}^j_{\mathrm{p}}$ in each chain, $j = \{1, \ldots, N\}$ are generated on the fly from the collection of chains, $\mathbf{X}_{(i-1)}$, using differential evolution

$$\mathbf{x}^j_{\mathrm{p}} = \gamma_{(d)}(\mathbf{x}^a_{(i-1)} - \mathbf{x}^b_{(i-1)}) + \boldsymbol{\zeta}_d, \quad a \neq b \neq j, \tag{1.46}$$

where $\gamma_{(d)}$ signifies the jump rate (dimensionality dependent), a and b are integers drawn without replacement from the natural numbers $\{1, \ldots, j-1, j+1, \ldots, N\}$, and $\boldsymbol{\zeta} \overset{\mathcal{D}}{\sim} \mathcal{N}_d(0, c_*)$ is drawn from a normal distribution with small standard deviation, say $c_* = 10^{-6}$. By accepting each proposal with Metropolis probability

$$P_{\mathrm{acc}}(\mathbf{x}^j_{\mathrm{p}}) = \min[1, f(\mathbf{x}^j_{\mathrm{p}})/f(\mathbf{x}^j_{(i-1)})], \tag{1.47}$$

a Markov chain is obtained, the stationary or limiting distribution of which is the posterior distribution.

Because the joint probability density function of the N chains factorizes to $f(\mathbf{x}^1|\cdot) \times \ldots \times f(\mathbf{x}^N|\cdot)$, the states, $\mathbf{x}^1_{(k)} \ldots \mathbf{x}^N_{(k)}$, of the individual chains are independent at any iteration (generation) k after DE-MC has become independent of its initial value. If the initial population is drawn from the prior distribution, then DE-MC translates this sample into a posterior population. From the guidelines of $s_{(d)}$ in RWM the optimal choice of $\gamma_{(d)} = 2.38/\sqrt{2d}$. With a 10% probability the value of γ is set to unity to enable the DE-MC chains to traverse rapidly large search spaces and jump directly between different (disconnected) modes of the equilibrium distribution. Mode-jumping is a desirable property of the DE-MC algorithm as evidenced by the performance of this method on multimodal target distributions.

Algorithm 5 summarizes the DE-MC algorithm in different algorithmic steps, wherein the auxiliary label Z is drawn for each chain by sampling from the standard uniform distribution, $Z \sim \mathcal{U}(0, 1)$. The DE-MC method remedies an important practical problem of the RWM algorithm, namely that of choosing an appropriate scale and orientation for the proposal distribution. Earlier

approaches such as (parallel) adaptive direction sampling solved the orientation problem but not the scale problem. The DE-MC method suffers one critical deficiency, however and that its performance is impaired if one or more of the sampled chains are trapped in an unproductive area of the parameter space in pursuit of the target distribution. This problem is well understood and explained in Figure 1.9 using some arbitrary univariate probability density function. Thus, chains that populate local optima can continue to persist forever if the jumps are insufficient to move the chain outside the space spanned by this optima. The state of this outlier chain not only contaminates the jumping distribution of Equation (1.46) thereby slowing down unnecessarily the evolution and mixing of the "good" chains, but also impair convergence to a limiting distribution. By sampling a disjoint part of the parameter space, the N chains will not reach consensus on the limiting distribution as the statistics of the outlier chain (certainly the mean sampled values) will differ substantially from their counterparts simulated by the other $N-1$ chains.

The chance of a dissident chain increases rapidly with dimensionality of the target distribution (larger number of chains, $N \geq 2d$) and complexity of the underlying density function. A patch is therefore of crucial importance to remedy the searching behavior of the DE-MC algorithm and to expedite convergence on non-smooth density functions with different regions of attraction and numerous local optima. Such a patch has important implications however, as any efforts to remedy outlier chains will violate detailed balance. The treatment of dissident chains should therefore be restricted to a burn-in period (see the SCEM-UA algorithm).

The next section concludes the section on MCMC simulation and presents the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm. This multi-chain algorithm does not suffer from outlier chains and uses subspace sampling with more than one chain pair to enhance, sometimes dramatically, the convergence rate to the target distribution. Many published papers have confirmed the efficiency and robustness of this algorithm for a large array of target distributions involving (among others) complex and/or high-dimensional search spaces, with one or multiple modes and intricate multivariate parameter dependencies. This explains why the DREAM algorithm has found widespread application and use in many different fields of study to reconcile, using Bayesian inference, system models with data. In fact, several publications have shown that DREAM even

provides better solutions than commonly used optimization algorithms. Whereas optimization algorithms are subject to population degeneration, the Metropolis selection rule promulgates chain diversity necessary to explore difficult search spaces in pursuit of the stationary distribution.

Multi-chain methods: DREAM The DREAM algorithm has it roots within DE-MC but uses subspace sampling and outlier chain correction to speed up convergence to the target distribution. Subspace sampling is implemented in DREAM by only updating randomly selected variables (coordinates) of $\mathbf{x}$ each time a proposal is generated. If A is a subset of d^*-dimensions of the original search space, $\mathbb{R}^{d^*} \subseteq \mathbb{R}^d$, then a jump, $\Delta\mathbf{x}^j_{(i-1)}$ in the jth chain, $j = \{1, \ldots, N\}$ at some iteration $i-1$ is calculated from the collection of chains, $\mathbf{X}_{(i-1)} = \{\mathbf{x}^1_{(i-1)}, \ldots, \mathbf{x}^N_{(i-1)}\}$, using differential evolution

$$\begin{aligned} \Delta\mathbf{x}^j_{(i-1),A} &= \boldsymbol{\zeta}_{d^*} + (\mathbf{1}_{d^*} + \boldsymbol{\lambda}_{d^*})\gamma_{(\delta,d^*)} \sum_{k=1}^{\delta} \left(\mathbf{x}^{\mathbf{a}_k}_{(i-1),A} - \mathbf{x}^{\mathbf{b}_k}_{(i-1),A}\right) \\ \Delta\mathbf{x}^j_{(i-1),\neq A} &= 0, \end{aligned} \tag{1.48}$$

where $\gamma_{(\delta,d^*)} = 2.38/\sqrt{2\delta d^*}$ is the jump rate, δ signifies the number of chain pairs that is used to compute the jump, and $\mathbf{a}$ and $\mathbf{b}$ are δ-vectors with integers drawn (without replacement) from $\{1, \ldots, j-1, j+1, \ldots, N\}$. The default value of $\delta = 3$, and results, in practice, in one-third of the proposals being created with $\delta = 1$, another third with $\delta = 2$, and the remaining third using $\delta = 3$. The values of $\boldsymbol{\lambda}$ and $\boldsymbol{\zeta}$ are sampled independently from $\mathcal{U}_{d^*}(-c, c)$ and $\mathcal{N}_{d^*}(0, c_*)$, respectively, the multivariate uniform and normal distribution with, typically, $c = 0.1$ and c_* small compared to the width of the target distribution, $c_* = 10^{-6}$ say. To expedite sampling of multimodal distributions, the default jump rate is switched to unity, $\gamma_{(\delta,d^*)} = 1$, with probability 0.2. The candidate point of the jth chain at iteration i then becomes

$$\mathbf{x}^j_{\mathrm{p}} = \mathbf{x}^j_{(i-1)} + \Delta\mathbf{x}^j_{(i-1)}, \tag{1.49}$$

and the Metropolis ratio of Equation (1.47) is used to determine whether to accept this proposal or not. If $P_{\mathrm{acc}}(\mathbf{x}^j_{\mathrm{p}}) \geq \mathcal{U}(0, 1)$ the candidate point is accepted and the jth chain moves to the new

position, that is $\mathbf{x}^j_{(i)} = \mathbf{x}^j_{\mathrm{p}}$, otherwise $\mathbf{x}^j_{(i)} = \mathbf{x}^j_{(i-1)}$. The default equation for γ should, for Gaussian and Student target distribution, result in optimal acceptance rates close to 0.44 for $d = 1$, 0.28 for $d = 5$, and 0.23 for large d.

The d^*-members of the subset A are sampled at random from the entries $\{1, \ldots, d\}$ (without replacement) and define the dimensions of the parameter space to be sampled by each proposal. This subspace spanned by A is construed in DREAM with the help of a crossover operator as follows. First, a value $\eta \in (0, 1]$ is sampled randomly from a geometric series of r different crossover values, $\boldsymbol{\eta} = \{\frac{1}{r}, \frac{2}{r}, \ldots, \frac{r}{r}\}$, with selection probabilities, $\mathbf{p}_\eta = \{\frac{1}{r}, \ldots, \frac{1}{r}\}$. Then, a d-vector $\mathbf{z}$ with random labels is drawn from the multivariate uniform distribution, $\mathbf{z} \sim \mathcal{U}_d(0, 1)$. All those coordinates l which satisfy $z_l \leq \eta$ are stored in the subset A and span the subspace that will be sampled using Equation (1.48). If A is an empty set, then one dimension of the target distribution will be sampled at random to avoid the jump vector, $\Delta\mathbf{x}^j_{(i-1)}$, to have zero entries everywhere.

We now provide an algorithmic recipe of the DREAM algorithm (see algorithm 6). The auxiliary label Z is drawn for each chain by sampling from the standard uniform distribution, $Z \sim \mathcal{U}(0, 1)$, and $\mathcal{U}\{a, b\}$ denotes the discrete uniform distribution with support $\{a, a + 1, \ldots, b - 1, b\}$, where $a, b \in \mathbb{N}_+$ and $b \geq a$. The variable $\mathcal{F}(\boldsymbol{\eta}|\mathbf{p}_\eta)$ signifies the discrete multinomial distribution on the crossover values, $\boldsymbol{\eta} = \{\eta_1, \ldots, \eta_r\}$, with selection probabilities, $\mathbf{p}_\eta = \{p_{\eta_1}, \ldots, p_{\eta_r}\}$, and the symbol $|A|$ signifies the cardinality, or number of elements, of the set A. The use of a vector of crossover probabilities enables single-site Metropolis (A contains one element), Metropolis-within-Gibbs (A has one or more elements) and regular Metropolis sampling ($A = \{1, \ldots, d\}$), and constantly introduces new directions in the parameter space that chains can take outside the subspace spanned by their current positions. What is more, the use of subspace sampling allows using $N < d$ in DREAM, an important advantage over DE-MC that requires $N = 2d$ chains to be run in parallel. This randomization of the search space, introduces one additional algorithmic variable to the algorithm, namely the desired number of crossover values, r. The default setting of $r = 3$ has shown to work well in practice, but larger values of this algorithmic variable might enhance the convergence rate on high-dimensional target distributions, say $d > 50$, to preserve the frequency

of low-dimensional jumps. Note, more intelligent subspace selection methods can be devised for target distributions with complex multivariate dependencies among the parameters. Correlated parameters should preferably be sampled jointly in tandem, otherwise too many of the (subspace) proposals will be rejected and the search can stagnate.

To enhance search efficiency, the selection probabilities, $\mathbf{p}_\eta = \{p_{\eta_1}, \ldots, p_{\eta_r}\}$, of the different crossover values, $\{\eta_1, \ldots, \eta_r\}$ are adapted during a burn-in period to maximize the distance traveled (Eulicdean norm) by the N chains.

In the past years, several other MCMC algorithms have appeared in the literature which use DREAM as their basic building block but include special extensions to simplify inference (among others) of discrete and combinatorial search spaces, and high-dimensional and CPU-intensive system models. This includes the DREAM$_{(\mathrm{ZS})}$, DREAM$_{(\mathrm{D})}$, DREAM$_{(\mathrm{DZS})}$, DREAM$_{(\mathrm{ABC})}$ and MT-DREAM$_{(\mathrm{ZS})}$ algorithm. These methods are summarized in Vrugt (2016) and have their own respective MATLAB toolboxes.

Convergence Monitoring Per theory, the chains that are simulated by a MCMC algorithm are expected to eventually converge to a stationary distribution, which should be the desired target distribution. But, how do we actually assess that convergence has been achieved in practice, without knowledge of the actual target distribution?

One way to check for convergence is to see how well the chains are mixing, or moving around the parameter space. For a properly converged MCMC sampler, the chains should sample, for a sufficiently long period, the approximate same part of the parameter space, and mingle readily and in harmony with one another around some fixed mean value. This can be inspected visually for each dimension of $\mathbf{x}$ separately, and used to diagnose convergence informally.

Another diagnostic that can be used to monitor convergence is the acceptance rate. A value between 15 - 30% is usually indicative of good performance of a MCMC simulation method. Much lower values usually convey that the posterior surface is difficult to traverse in pursuit of the target distribution. A low acceptance rate can have different reasons, for instance poor model numerics,

or the presence of multi-modality and local optima. Yet, the acceptance rate can only diagnose whether a MCMC method such as DREAM is achieving an acceptable performance, it cannot be used to determine when exactly convergence has been achieved.

Several non-parametric and parametric statistical tests can be used to determine when convergence of the sampled chains to a limiting distribution has been achieved. The most powerful of these convergence tests is the multi-chain $\widehat{R}$-statistic. This diagnostic compares for each parameter $l = \{1, \ldots, d\}$ the within-chain

$$W_l = \frac{2}{N(T-2)} \sum_{j=1}^{N} \sum_{i=\lfloor T/2 \rfloor}^{T} (\mathbf{x}_{(i),l}^{j} - \overline{\mathbf{x}}_l^j)^2 \qquad \overline{\mathbf{x}}_l^j = \frac{2}{T-2} \sum_{i=\lfloor T/2 \rfloor}^{T} \mathbf{x}_{(i),l}^{j} \tag{1.50}$$

and between-chain variance

$$B_l/T = \frac{1}{2(N-1)} \sum_{j=1}^{N} (\overline{\mathbf{x}}_l^j - \overline{\overline{\mathbf{x}}}_l)^2 \qquad \overline{\overline{\mathbf{x}}}_l = \frac{1}{N} \sum_{j=1}^{N} \overline{\mathbf{x}}_l^j \tag{1.51}$$

using

$$\widehat{R}_l = \sqrt{\frac{N+1}{N} \frac{\widehat{\sigma}_+^{2(l)}}{W_l} - \frac{T-2}{NT}}, \tag{1.52}$$

where T signifies the number of samples in each chain, $\lfloor \cdot \rfloor$ is the integer rounding operator, and $\widehat{\sigma}_+^{2(j)}$ is an estimate of the variance of the lth parameter of the target distribution

$$\widehat{\sigma}_+^{2(j)} = \frac{T-2}{T} W_j + \frac{2}{T} B_j. \tag{1.53}$$

To official declare convergence, the value $\widehat{R}_l \leq 1.2$ for each parameter, $l \in \{1, \ldots, d\}$, otherwise the value of T should be increased and the chains run longer. As the N different chains are launched from different starting points, the $\widehat{R}$-diagnostic is a relatively robust estimator.

A related, but more powerful convergence diagnostic is the multivariate variant of the $\widehat{R}$-statistic. This statistic, hereafter referred to as $\widehat{R}^d$-diagnostic assesses convergence of the d parameters simultaneously by comparing their within and between-sequence covariance matrix. Convergence

is achieved when a rotationally invariant distance measure between the two matrices indicates that they are "sufficiently" close. Then, the multivariate $\widehat{R}^d$-statistic achieves a value close to unity, otherwise its value is much larger. In fact, the $\widehat{R}$ and $\widehat{R}^d$-statistic take on a very similar range of values, hence simplifying analysis of when convergence has been achieved. The $\widehat{R}^d$-statistic is particularly useful for high-dimensional target distributions involving complicated multi-dimensional parameter interactions.

Other statistics include the autocorrelation function. The autocorrelation function for each parameter $l = \{1, \ldots, d\}$ is defined as

$$\rho_{l,k}^j = \frac{\sum_{i=1}^{T-k}(\mathbf{x}_{(i),l}^j - \overline{\mathbf{x}}_l^j)(\mathbf{x}_{(i+k),l}^j - \overline{\mathbf{x}}_l^j)}{\sum_{i=1}^{T}(\mathbf{x}_{(i),l}^j - \overline{\mathbf{x}}_l^j)^2}, \tag{1.54}$$

and returns the correlation between two samples k iterations apart in the jth chain, $j = \{1, \ldots, N\}$. Compared to rejection sampling which, per construction, produces uncorrelated samples, MCMC chain trajectories exhibit autocorrelation as the current state of the chain is derived from its previous state. This correlation is expected to decrease with increasing lag k. The autocorrelation function is a useful proxy to assess sample variability and mixing, but does not convey when convergence has been achieved. A high autocorrelation, say $|\rho| > 0.8$, at lags, say $k \geq 5$, demonstrates a poor mixing of the individual chains.

Altogether, joint interpretation of the different diagnostics should help assess convergence of the sampled chain trajectories. Of all these metrics, the $\widehat{R}^d$-statistic is most conservative and strict and provides the best guidance on exactly when convergence has been achieved. This happens as soon as this statistic drops below the critical threshold of 1.2. Suppose this happens at T^* iterations (generations) then the first $(T^* - 1)$ samples of each chain are simply discarded as burn-in and the remaining $N(T - T^*)$ samples from the joint chains are used for posterior analysis. Note, we always recommend to verify convergence of DREAM by visually inspecting the mixing of the different chain trajectories.

In practice, one has to make sure that a sufficient number of chain samples is available for the inference, otherwise the posterior estimates can be biased. For convenience, we list here the

total number of posterior samples, $N(T - T^*)$ (in brackets) one would need for a reliable inference with DREAM for a given dimensionality of the target distribution: $d = 1$ (500); $d = 2$ (1,000); $d = 5$ (5,000); $d = 10$ (10,000); $d = 25$ (50,000); $d = 50$ (200,000); $d = 100$ (1,000,000); $d = 250$ (5,000,000). These listed numbers are only a rough guideline, and based on several assumptions such as a reasonable acceptance rate ($> 10\%$) and not too complicated shape of the posterior distribution. In general, the number of posterior samples required increases with rejection rate and complexity of the target distribution.

1.6 Case Studies

In this section we illustrate the application of Bayesian inference to four different case studies involving modeling of the instantaneous unit hydrograph, the rainfall-runoff transformation, and water movement in the variably saturated zone. These examples focus on the movement and distribution of water in the surface and subsurface, and satisfy the overarching theme of this book. The last two case studies have appeared in literature publications (not everything), therefore we only present herein a short summary of the experimental data, models, numerical setup, and findings. Interested readers are referred to the original publications for further details.

1.6.1 Instantaneous Unit Hydrograph

The first case study considers the modeling of the instantaneous unit hydrograph defined as

$$Q_t = \frac{1}{r_c \Gamma(h)} \left(\frac{t}{r_c}\right)^{(h-1)} \exp\left(-\frac{t}{r_c}\right), \tag{1.55}$$

where Q_t (mm day^{-1}) signifies the simulated streamflow at time t (days), h (-) denotes the number of reservoirs, r_c (days) signifies the recession constant, and $\Gamma(\cdot)$ is the gamma function.

A $n = 25$ - day period with synthetic daily streamflow data was generated by driving Equation (1.55) with an artificial precipitation record using $h = 2$ reservoirs, and a recession constant of $r_c = 4$ days. This artificial data set is subsequently perturbed with a heteroscedastic measurement error

(non-constant variance) with standard deviation equal to 10% of the original simulated discharge values. We now use the n-record of corrupted discharge values, $\widetilde{\mathbf{Y}} = \{\widetilde{y}_1, \ldots, \widetilde{y}_n\}$, to estimate the posterior distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, of the Nash parameters, $\boldsymbol{\theta} = \{h, r_c\}$, using Bayes' theorem in Equation (1.12). We assume a flat prior distribution for the two Nash parameters, $P(\boldsymbol{\theta}) \sim \mathcal{U}_2[1, 10]$, and use the likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\sigma}}^2)$, of Equation (1.30). The n-vector of measurement error variances, $\widehat{\boldsymbol{\sigma}}^2 = \{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2\}$ is derived from Equation (1.41) using the nuisance variables, $\boldsymbol{\alpha} = \{\sigma_0, \sigma_1\}$. We estimate the joint posterior distribution of the Nash model parameters and the nuisance variables, $P(\boldsymbol{\theta}, \boldsymbol{\alpha}|\widetilde{\mathbf{Y}})$, using MCMC simulation with DREAM (algorithm 6).

Figure 1.10 summarizes the results of our analysis and presents a scatter plot matrix of the posterior samples derived with the DREAM algorithm. The main diagonal displays histograms of the marginal distribution of the two Nash model parameters h and r_c and the nuisance variables σ_0 and σ_1, whereas the off-diagonal graphs display bivariate scatter plots of the posterior samples. The x-axes matches exactly the posterior ranges of the parameters and their "true" values, $h = 4$, $r_c = 2$, $\sigma_0 = 0$ and $\sigma_1 = 0.1$. are separately indicated in each histogram with the red cross.

The posterior histograms of the Nash model parameters follow closely a Gaussian distribution with mean that is in excellent agreement (as should be!) with the true values used to create the synthetic 25-day hydrograph. The marginal distributions exhibit a relatively small dispersion (in comparison to prior ranges), which demonstrates that h and r_c are well defined by calibration against the observed (synthetic) streamflow data. The nuisance variables of Equation (1.41) are also well resolved, yet the marginal posterior distribution of the slope of the linear measurement error model, σ_1 is somewhat biased towards larger values. Indeed, the true value of the slope appears to the left of the maximum a-posteriori (MAP) density solution of σ_1 in the left tail of the distribution. The intercept, σ_1, of Equation (1.41) is truncated at zero by the uniform prior distribution, and consequently follows a log-normal distribution. Such distribution is difficult to approximate accurately with the first-order approximation of Equation (1.6), but poses no problems for sampling methods. The bivariate scatterplots of the posterior samples (off-diagonal plots) demonstrate a strong linear (negative) correlation between the Nash model parameters, h and r_c.

The nuisance variables and Nash parameters appear uncorrelated, a testament to the use of an adequate likelihood function.

We next investigate how the posterior parameter uncertainty translates into simulation uncertainty of the Nash model. Figure 1.11 presents a time series plot of the observed data (red dots) and 95% simulation uncertainty ranges of the hydrograph due to parameter (dark gray region) and total uncertainty (light gray region). The Nash model tracks closely the observed data (as is to be expected with synthetic data) with simulation intervals that appear reasonably small. Further analysis of the residuals confirms our assumptions that the residuals are temporally uncorrelated and follow an approximately Gaussian distribution. We next present an example in which the residuals do not satisfy our assumptions.

1.6.2 The Rainfall-Runoff Transformation

The second case study involves modeling of the discharge dynamics of the French Broad River basin at Asheville, North Carlina, USA. This case study has appeared in Schoups and Vrugt (2010) and readers are referred to this work for further details. We summarize briefly the experimental data, the hmodel and its parameters, the prior distribution and likelihood function(s) before we report the findings of our analysis.

The rainfall-discharge relationship of the French Broad is simulated using the hmodel of Schoups and Vrugt (2010). This relatively simple conceptual watershed model is illustrated in Figure 1.12 and transforms mean areal precipitation into runoff emanating from the catchment outlet using four different reservoirs (state variables) and process descriptions for interception, throughfall, evaporation, runoff generation, percolation, and surface and subsurface routing (see Figure 1.12). Runoff generation is assumed to be dominated by saturated overland flow and is simulated as a function of basin water storage without an explicit dependence on rainfall intensity. This assumption is typically valid for temperate climates but may be violated in semiarid watersheds. Snow accumulation and snowmelt are also not accounted for, yet this is not a problem if we focus on "warm" watersheds. Table 2.1 summarizes the seven different parameters of the hmodel, and

their prior uncertainty ranges.

We use a seven-year record of the French Broad River basin from the MOPEX data set with daily data of discharge (mm/day), mean areal precipitation (mm/day), and mean areal potential evapotranspiration (mm/day) to estimate the hmodel parameters, $\boldsymbol{\theta} = \{I_{\max}, S_{\max}, Q_{\max}, \alpha_{\mathrm{E}}, \alpha_{\mathrm{F}}, K_{\mathrm{F}}, \alpha_{\mathrm{S}}\}$. We use a two-year spin-up period to reduce sensitivity of the model to state-value initialization. In other words, only the last five years are used for our analysis.

We assume a uninformative prior distribution, $P(\boldsymbol{\theta})$, for the parameters of the hmodel using the ranges listed in Table 2.1. As a first attempt, we expect conveniently that the discharge residuals are uncorrelated, normally distributed, and with a constant variance. These assumptions lead to the Gaussian likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, in Equation (1.28) - and with our uniform prior leads to the density function of Equation (1.29).

Figure 1.13 displays the results of the DREAM algorithm. The top panel plots the simulation of the maximum likelihood solution for the 1800-day calibration data record, whereas the bottom panel presents a diagnostic check of the corresponding discharge residuals. The hmodel simulation tracks closely the observed discharge data (dots), with 95% intervals (not shown) that appear rather small. Unfortunately, the residuals do not satisfy the three main underlying assumptions of our likelihood function (Gaussian, uncorrelated, constant variance). Indeed, the variance of the residuals increases with flow level, the residual distribution deviates from normality (much more peaked), and the residuals display considerable serial correlation at the first lag.

We now relax the residual assumptions and use the generalized likelihood function to derive the parameters of the hmodel. This requires inference as well of the nuisance variables, $\boldsymbol{\alpha} = \{\sigma_0, \sigma_1, \beta, \xi, \boldsymbol{\Phi}_p\}$. We use $p = 4$ and derive the the posterior distribution of the hmodel parameters and nuisance variables, $P(\boldsymbol{\theta}, \boldsymbol{\alpha}|\widetilde{\mathbf{Y}})$ using MCMC simulation with the DREAM algorithm. Figure 1.14 presents the main results of this analysis. It is evident that the residuals of the maximum likelihood simulation (top plot) now are consistent with the properties of the likelihood function. The residual variance is constant and independent of flow level, the residual distribution is well-described

with a Laplace distribution, and the residuals are temporally uncorrelated.

Figure 1.15 summarizes the effect of the choice of likelihood function on the posterior marginal distribution of the hmodel parameters $I_{\max}$, $S_{\max}$, $Q_{\max}$, α_{F}, K_{F}, and K_{S}. The blue histograms display the results of the Gaussian likelihood function and their green counterparts pertain to the generalized likelihood function. The results in Figure 1.15 highlight several important findings. First, the marginal distributions of the hmodel parameters differ substantially between the two different likelihood function. Second, the posterior histograms derived from the generalized likelihood function exhibit a larger dispersion. Finally, the marginal distributions derived from the generalized likelihood functions are not truncated by the prior distribution. This is particularly evident for the maximum interception, $I_{\max}$, which hits it upper bound of 10 mm for the Gaussian likelihood and takes on much more realistic values of 1-4 mm with the generalized likelihood function.

This concludes the second case study. Interested readers are referred to Schoups and Vrugt (2010) for a much more detailed interpretation and treatment of the results.

1.6.3 Vadose Zone Hydrology

The third and last case study considers the modeling of the soil moisture regime of an agricultural field near Jülich, Germany. Soil moisture content was measured with Time Domain Reflectometry (TDR) probes at 6 cm deep at 61 locations in a 50 $\times$ 50 m experimental plot. The TDR data were analysed and the measured apparent dielectric permittivities were converted to soil moisture values. Measurements were taken on $n = 29$ days between 19 March and 14 October 2009, comprising a measurement campaign of 210 days. For the purpose of the present study, the soil moisture observations were averaged per day to yield a single plot-mean water content time series. Precipitation and other meteorological variables were recorded at a meteorological station located 100 m west of the measurement site.

The HYDRUS-1D model of Simuunek et al. (2008) was used to simulate variably saturated water flow in the agricultural field. This model uses the finite element method to solve numerically

Richards' equation

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z} + 1\right)\right], \tag{1.56}$$

for given (measured) initial and boundary conditions, where θ (cm^3 cm^{-3})[2] signifies moisture content, t (days) denotes time, z (cm) is the vertical (depth) coordinate, h (cm) is the pressure head, and $K(h)$ (cm day^{-1}) represents the corresponding (unsaturated) soil hydraulic conductivity.

Numerical solution of Equation (4.1) requires knowledge of the soil hydraulic properties. We use the van Genuchten-Mualem (VGM) model (Van Genuchten, 1980)

$$\begin{aligned} \theta(h) &= \theta_r + (\theta_s - \theta_r)[1 + (\alpha|h|)^n]^{-m} \\ K(h) &= K_s S_{\mathrm{e}}(h)^{\lambda}[1 - (1 - S_{\mathrm{e}}(h)^{1/m})^m]^2, \end{aligned} \tag{1.57}$$

where θ_s and θ_r (cm^3 cm^{-3}) signify the saturated and residual soil water content, respectively, α (cm^{-1}), n (-) and $m = 1 - 1/n$ (-) are shape parameters, K_s (cm day^{-1}) denotes the saturated hydraulic conductivity, and λ (-) represents a pore-connectivity parameter. The effective saturation, S_{e} (-) is defined as

$$S_{\mathrm{e}}(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r}. \tag{1.58}$$

Observations of daily precipitation and daily potential evapotranspiration are used to characterize the upper boundary condition of our experimental plot. In the absence of detailed knowledge of the lower boundary condition, we assume a constant head, h_{bot} (cm) at the bottom of our modeled soil domain and estimate its value along with the six soil hydraulic parameters of Equation (1.57).

Table 3.1 lists the parameters of the HYDRUS-1D model, their units, upper and lower bounds for the soil domain under investigation, and respective prior distributions. These parameters, $\boldsymbol{\theta} = \{\theta_r, \theta_s, \alpha, n, K_s, \lambda, h_{\mathrm{bot}}\}$, are subject to inference using the soil moisture measurements. The marginal priors of the six soil hydraulic parameters (first six in table) are construed from surrogate data (soil texture) using the Rosetta toolbox of hierarchical pedotransfer functions. A flat prior with

[2]The symbol θ (cm^3 cm^{-3}) is used in the hydrologic sciences to denote volumetric soil moisture content. In present paper we also use this symbol to denote the parameter vector.

ranges between -250 and -10 cm is used for the bottom head, h_{bot}. This guarantees the parameters to take on values that are deemed physically realistic. Folding is the only efficient boundary treatment procedure that does not destroy detailed balance of the sampled chain trajectories.

The initial state of each chain is sampled from the prior distribution, and the $N = 10$ different chains are ran in parallel using the MATLAB parallel computing toolbox.

Figure 1.16 presents histograms of the marginal posterior distributions sampled by DREAM of the HYDRUS parameters θ_s, α, n, and K_{s} of the 100 cm deep soil of our 50×50 experimental plot in Germany. The bottom panel presents a time series plot of simulated soil moisture contents. The dark gray region constitutes the 95% HYDRUS-1D simulation uncertainty due to parameter uncertainty, whereas the light gray region denotes the total simulation uncertainty (parameter + randomly sampled additive error). The observed soil moisture values are indicated with a red circle. The soil hydraulic parameters appear well defined by calibration against the observed soil moisture measurements. Their marginal distributions follow closely their respective normal prior distributions (Table 3.1), yet exhibit somewhat less dispersion. The HYDRUS-1D model closely tracks the observed soil moisture contents with Root Mean Square Error (RMSE) of the posterior mean simulation of about 0.01 $\text{cm}^3/\text{cm}^{-3}$. About 95% of the observations lies within the gray region, an indication that the simulation uncertainty ranges are statistically adequate. The acceptance rate of DREAM averages about 12.6% - about half of its theoretical optimal value of 22-25% (for Gaussian and Student target distributions). This deficiency is explained in part by the high nonlinearity of retention and hydraulic conductivity functions, and numerical errors of the implicit, time-variable, solver of the Richards' equation. This introduces irregularities (e.g. local optima) in the posterior response surface and makes the journey to and sampling from the target distribution more difficult.

1.7 Limits of Acceptability

In the manifesto for the equifinality thesis, Beven (2006) suggested that a more rigorous approach to model evaluation would involve the use of limits of acceptability for each individual observation

against which model simulated values are compared. Within this framework, behavioural models are defined as those that satisfy the limits of acceptability for each observation. Ideally, the limits of acceptability should reflect the observational error of the variable being compared, together with the effects of input error and commensurability errors resulting from time or space scale differences between observed and predicted values. The limits of acceptability approach has been used by various authors, although earlier publications used similar ideas within GLUE based on fuzzy measures. The limits of acceptability framework might be considered more objective than the standard GLUE approach advocated as the limits are defined before running the model on the basis of best available hydrologic knowledge.

For now lets assume that the prior distribution, $P(\boldsymbol{\theta}) \sim \mathcal{U}_d(\mathbf{a}, \mathbf{b})$ is multivariate uniform between some d-vector of values $\mathbf{a}$ and $\mathbf{b}$. For a proposal, $\boldsymbol{\theta}_j$ to be deemed acceptable, $\mathbf{Y}(\boldsymbol{\theta}_j)$ should be contained exclusively within the interval $[\widetilde{y}_t - \Delta_t, \widetilde{y}_t + \Delta_t]$ at each time $t = \{1, \ldots, n\}$. This so called "behavioral simulation space" belongs to the set $\widehat{\Omega}_{(\mathbf{Y})}$ and can be defined as

$$\widehat{\Omega}_{(\mathbf{Y})} = \left\{ \mathbf{Y} \in \mathbb{R}^n \ : \ \mathbf{Y} = \mathcal{M}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}}) \ ; \ \boldsymbol{\theta} \in \widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})} \right\}, \tag{1.59}$$

where $\widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}$ constitutes the posterior (behavioral) parameter set

$$\widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})} = \Omega_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}. \tag{1.60}$$

The conditional parameter set, $\Omega_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}$ is defined as follows

$$\Omega_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})} = \left\{ \boldsymbol{\theta} \in \boldsymbol{\theta} \in \mathbb{R}^d \ : \ \widetilde{\mathbf{Y}} - \mathcal{F}(\boldsymbol{\theta}, \mathbf{x}_0, \widetilde{\mathbf{B}}) = \mathbf{E}(\boldsymbol{\theta}) \ ; \ e_t(\boldsymbol{\theta}) \in [-\Delta_t, \Delta_t] \ , \ t = 1, \ldots, n \right\}, \tag{1.61}$$

and contains solutions, $\boldsymbol{\theta}_j \in \widehat{\Omega}_{(\boldsymbol{\theta}|\mathbf{Y})}$, that satisfy the limits of acceptability of each observation.

If an informative prior distribution is used then the behavioral (posterior) parameter set is computed as the intersection of the prior parameter set, $\Omega_{(\boldsymbol{\theta})}$ and conditional parameter set

$$\widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})} = \Omega_{(\boldsymbol{\theta})} \cap \Omega_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}. \tag{1.62}$$

1.7.1 The DREAM$_{(\text{ABC})}$ Algorithm

Application of the limits of acceptability approach requires the availability of a sampling method that can efficiently search the parameter space in pursuit of behavioral solution set, $\widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}$. Commonly used (population Monte Carlo) rejection sampling methods are rather inefficient in locating behavioral solutions. The chance that a random sample from the prior distribution satisfies the limits of acceptability of each observation is disturbingly small, particularly if the prior parameter space is large compared to the posterior (behavioral) solution space and the number of observations, n is large. Fortunately, an efficient MCMC sampling method, the DREAM$_{(\text{ABC})}$ algorithm, has been developed by Sadegh and Vrugt (2014) to explore efficiently set-theoretic functions such as Equation (1.61).

This selection rule is defined as

$$P_{\text{acc}}(\boldsymbol{\theta}_{(i-1)} \rightarrow \boldsymbol{\theta}_{\text{p}}) = \begin{cases} I\big(f(\boldsymbol{\theta}_{\text{p}}) \geq f(\boldsymbol{\theta}_{(i-1)})\big) & \text{if } f(\boldsymbol{\theta}_{\text{p}}) < n \\ 1 & \text{if } f(\boldsymbol{\theta}_{\text{p}}) = n \end{cases}, \tag{1.63}$$

where $I(a)$ is an indicator function that returns one if the condition a is satisfied and zero otherwise, and the fitness function, $f(\cdot)$, is calculated as follows

$$f(\boldsymbol{\theta}) = \sum_{t=1}^{n} I(|\widetilde{y}_t - y_t(\boldsymbol{\theta})| \leq \Delta_t). \tag{1.64}$$

If the proposal is accepted, then the Markov chain moves to this new position, $\boldsymbol{\theta}_{(i)} = \boldsymbol{\theta}_{\text{p}}$, otherwise it remains at its current location, that is $\boldsymbol{\theta}_{(i)} = \boldsymbol{\theta}_{(i-1)}$.

The fitness of the proposal, $\boldsymbol{\theta}_{\text{p}}$, is equivalent to the number of observations its simulation satisfies within the limits of acceptability. We accept the proposal, $P_{\text{acc}}(\boldsymbol{\theta}_{(i-1)} \rightarrow \boldsymbol{\theta}_{\text{p}}) = 1$, if the fitness of $\boldsymbol{\theta}_{\text{p}}$ is larger than that of the current state of the chain, $\boldsymbol{\theta}_{(i-1)}$, or if the simulation of the proposal is consistently within $\boldsymbol{\Delta} = \{\Delta_1, \ldots, \Delta_n\}$ of the observed values, and thus $f(\boldsymbol{\theta}_{\text{p}}) = n$, otherwise the candidate point is rejected. After a burn-in period in which $f(\cdot) < n$, the convergence of DREAM$_{(\text{ABC})}$ can be monitored with the $\widehat{R}$ diagnostic. A full description of DREAM$_{(\text{ABC})}$ appears

in Sadegh and Vrugt (2014) and interested readers are referred to this publication for further details.

1.7.2 Vadose Zone Hydrology Revisited

We now revisit the third case study and use instead limits of acceptability. Scharnagl et al. (2011) depict in their figure 8 (p. 3055), the 95% ranges of the observed soil moisture data at each measurement time. The value of $\Delta_t; t = \{1, \ldots, n\}$ is simply set equal to half the width of the 95% interval of the distributed moisture content observations. This equates to an average value of the limits of acceptability of 0.047 (cm^3cm^{-3}).

Figure 1.19 presents histograms of the marginal posterior distribution of the six HYDRUS-1D model parameters considered in this study. The bottom panel presents a time series plot of the behavioral simulation set, $\widehat{\Omega}_{(\mathbf{Y})}$. The observed soil moisture data are indicated separately with red dots. The behavioral HYDRUS-1D model nicely tracks the observed soil moisture measurements with behavioral simulation space, $\widehat{\Omega}_{(\mathbf{Y})}$ that encapsulates consistently the observed data. The root mean square error (RMSE) of the behavioral (posterior) mean simulation equates to about 0.0149 cm^3/cm^{-3}, a value somewhat larger than derived separately using the formal Gaussian likelihood function. The behavioral parameter space of most parameters extend a large part of their respective prior ranges with marginal distributions that deviate markedly from normality. The prior ranges are rather narrow and derived from Monte Carlo simulation with the ROSETTA pedotransfer toolbox using textural data (percentages of sand, silt, and clay) as main input variables.

Finally, Figure 1.18 shows how the posterior parameter set translates into uncertainty of the soil water retention (left) and unsaturated soil hydraulic conductivity (right) functions. The light gray region corresponds to the prior parameter set, $\Omega_{(\mathbf{Y})}$ whereas the dark gray is used to denote the behavioral (posterior) solution set, $\widehat{\Omega}_{(\mathbf{Y})}$. Note, we use here the variable $\mathbf{Y}$ to denote the functional space rather than moisture content values. The posterior mean soil hydraulic functions are indicated with the solid black line. The posterior uncertainty of the soil hydraulic functions appears rather large in response to the observed spatial variability of the soil moisture data. This

uncertainty can now be used to simulate the soil moisture variability in the 50×50 experimental plot, simply by drawing soil hydraulic functions from the posterior ranges. Thus, the limits of acceptability framework provides a way to account explicitly for spatial variability. This concludes our section with case studies. Many other examples can be found in different fields of study (e.g. chemistry, ecology, geomorphology, physics, structural engineering) and tackle much more complex problems involving large dimensionalities of the target distribution. A summary with some of these applications is found on Page 276 of Vrugt (2016).

1.8 Marginal Likelihood and Model Complexity

Thus far in this Chapter, we have focused our inferences on a single model but without recourse to the denominator, $P(\widetilde{\mathbf{Y}})$, in Equation (2.4), the so-called marginal likelihood. This normalizing constant in Bayes' theorem ensures that the posterior distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, integrates to unity. The marginal likelihood, or Bayesian model evidence, is of no particular interest for parameter estimation, yet of imminent importance for hypothesis testing. The hypothesis (model), $\mathcal{M}_k$, where $k = \{1, \ldots, K\}$, with largest evidence, $P(\widetilde{\mathbf{Y}}|\mathcal{M}_k)$, is most supported by the available data, $\widetilde{\mathbf{Y}}$.

Bayesian model selection encodes a natural preference for simpler and more constrained models. This approach provides a rigorous justification to the parsimony principle of William of Ockham (1287-1347), an English Franciscan friar, philosopher and theologian, who stated that *"....Entities must not be multiplied beyond necessity"*. This principle of parsimony, also known as Occam's azor, is traceable to the works of philosophers such as Aristotle (384-322 BC), Ptolemy (c. AD 90 - c. AD 168), and consistent with requirements of falsifiability in the scientific method. Indeed, simpler hypothesis (theories) are preferred as they involve fewer assumptions and are therefore easier testable. Thus, a "good" model selection technique must necessarily balance goodness of fit with complexity (number of "free" parameters). Unfortunately, analytical solutions of $P(\widetilde{\mathbf{Y}}|\mathcal{M}_k)$ are available only for certain special cases, which are too limiting to be of practical value in

environmental modeling. We therefore have to resort to sampling methods which approximate numerically the integral of the posterior distribution.

In the Monte Carlo approach, the marginal likelihood can be approximated by the arithmetic mean of the likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, of a large sample of points drawn randomly from the prior distribution. This approximation, however, is not particularly efficient, as many of the random samples drawn from the prior parameter distribution will exhibit insufficient density to contribute to the evidence. A more efficient approach constitutes importance sampling (see algorithm 2). The importance distribution, $G(\boldsymbol{\theta})$, has a known integral of unity, and should satisfy that $g(\boldsymbol{\theta}) > 0$ whenever $P(\boldsymbol{\theta}) > 0$, otherwise certain parts of the target distribution, $P(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ are possible dismissed. The ratio of the density of the unnormalized posterior, $P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, and the density, $g(\boldsymbol{\theta})$, of the importance distribution, $G(\boldsymbol{\theta})$, now details the contribution of some importance sample, $\boldsymbol{\theta}_j$, to the marginal likelihood. The integral of the unnormalized posterior distribution, $P(\widetilde{\mathbf{Y}}|\mathcal{M}_k)$, is thus equivalent to the expected value of $P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})/g(\boldsymbol{\theta})$, or $\mathbb{E}[P(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})/g(\boldsymbol{\theta})]$ and can be approximated numerically

$$P(\widetilde{\mathbf{Y}}|\mathcal{M}_k) \simeq \frac{1}{M}\sum_{j=1}^{M}\frac{P(\boldsymbol{\theta}_j|\mathcal{M}_k)L(\boldsymbol{\theta}_j|\mathcal{M}_k,\widetilde{\mathbf{Y}})}{g(\boldsymbol{\theta}_j)} \tag{1.65}$$

using M different samples $\{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_M\}$ drawn randomly from the importance distribution, $G(\boldsymbol{\theta})$.

The importance estimator of Equation (1.65) is accurate and robust but not without practical problems. Indeed, the efficiency of importance sampling depends critically on the choice of the importance distribution, $G(\boldsymbol{\theta})$. This becomes particularly relevant for CPU-intensive system models and high-dimensional target distributions. To enhance the computuational efficiency of the estimator of Equation (1.65), a two-step approach in which samples from the target distribution (step 1) are used to construct an adequate importance distribution (step 2). Benchmark experiments show that this approach, called Gaussian mixture importance sampling (GMIS) is unbiased, robust, and, efficient, requiring only $M = 10,000$ samples to estimate the evidence, $P(\widetilde{\mathbf{Y}}|\mathcal{M}_k)$, of complex, high-dimensional, target distributions (up to 100 dimensions). The GMIS approach is implemented

in the MATLAB package of DREAM described by Vrugt (2016) and simplifies considerably scientific inquiry through hypothesis testing and model selection.

As last item in this Chapter we provide in Figure (1.19 a textbook example of the relationship between model complexity, and (left graph) the magnitude of the residuals in the calibration (red line) and (independent) evaluation (blue line) period, and (right graph) the value of the maximum likelihood (blue line) and the marginal likelihood (red line). The most important findings are as follows. First, the residuals of the calibration data period decrease steadily in magnitude with increasing complexity of the model. The benefits of using more parameters are largest for parsimonious models, and become increasingly smaller for more complex models. If the model is too complex, then the quality of fit can hardly be improved with the use of more parameters. Second, the residuals of the evaluation period (prediction error) exhibit a characteristic "U" shape. Indeed, adding complexity to simple models will help reduce the prediction error, but only until an optimal complexity of the model is derived, after which the magnitude of the prediction error increases due to model overparameterization. Third, the maximum likelihood value of the calibration period increases with model complexity. The more parameters that are used to explain the calibration data the more flexible the model, and the larger the likelihood maximum. Fourth, the marginal likelihood shows a pattern equivalent to the prediction error. Its value is maximized when the model has an optimal complexity. Finally, to judge the optimum model complexity, traditional metrics of fit require use of an independent evaluation period (left plot), whereas the same conclusions can be drawn from the marginal likelihood using only the observations in the calibration data set.

1.9 Conclusion

This Chapter has reviewed the main elements of Bayesian inference to reconcile dynamic environmental system models with observations, to facilitate prediction in time (forecasting) and space (interpolation), data assimilation, and inference of the model parameters. The prior distribution, formulation of the likelihood function, and marginal likelihood have been discussed extensively with special emphasis on numerical techniques suited to approximate the posterior distribution. This includes

rejection sampling, importance sampling, and recent developments in Markov chain Monte Carlo simulation to sample complex target distributions. We also have highlighted their application to sampling limits of acceptability. Three different case studies with surface and subsurface models were presented to illustrate the application of Bayesian inference to quantification of parameter and model predictive uncertainty. Numerical recipes were provided for each of the numerical techniques to facilitate implementation and use of Bayes analysis.

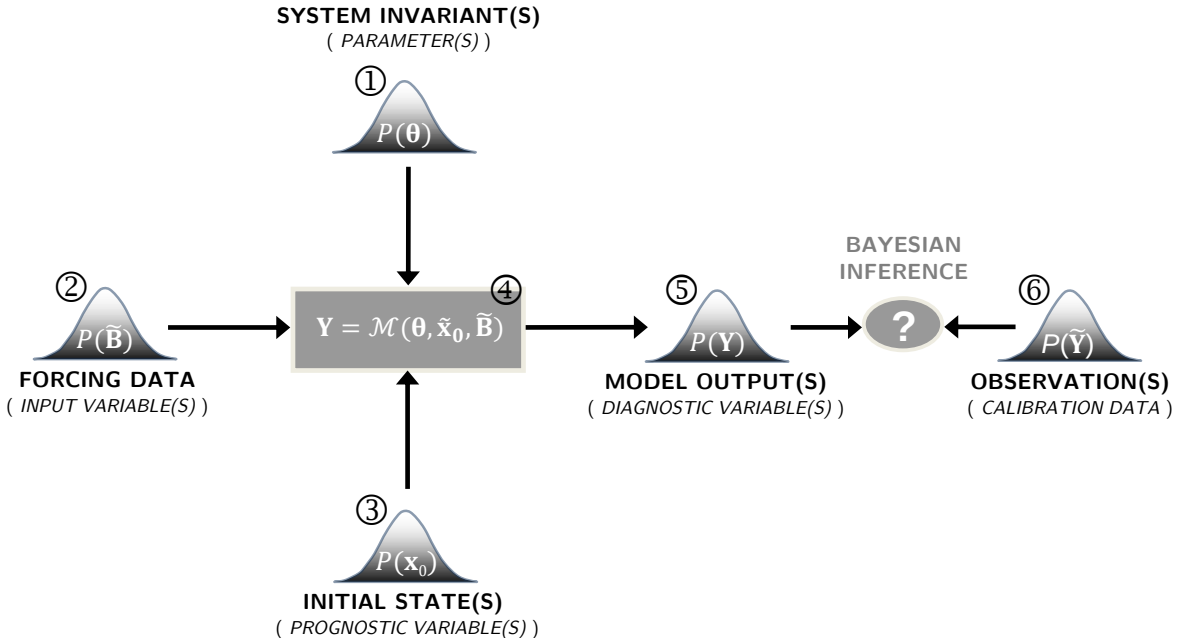

Figure 1.1: Schematic illustration of the most important sources of uncertainty in environmental systems modeling, including (1) parameter, (2) input data (also called forcing or boundary conditions), (3), initial state, (4) model structural, (5) output, and (6) calibration data uncertainty. The measurement error of the calibration data is often prescribed, a rather convenient assumption in most practical situations.

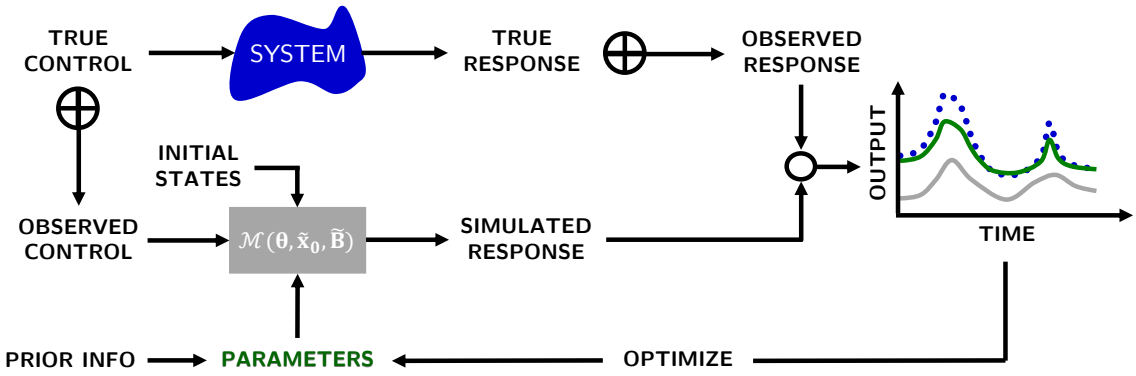

Figure 1.2: Schematic overview of the model calibration problem in the presence of measurement, parameter, and epistemic uncertainty. The model parameters are adjusted iteratively so that the simulated response (solid lines) of the model, $\mathcal{M}(\boldsymbol{\theta}, \widetilde{\mathbf{x}}_0, \widetilde{\mathbf{B}})$, approximates as closely and consistently as possible the observed response (blue dots).

Algorithm 2 Importance Sampling

1: Define an importance distribution, $G(\mathbf{x})$, so that $g(\mathbf{x}) > 0$ if $f(\mathbf{x}) > 0$.
2: Define M.
3: **for** $i = 1, \ldots, M$ **do**
4: Sample randomly a point, $\mathbf{x}_{(i)}$, from $G(\mathbf{x})$, $\mathbf{x}_{(i)} \sim G(\mathbf{x})$ and calculate $f(\mathbf{x}_{(i)})$.
5: Compute the importance weight, $w(\mathbf{x}_{(i)}) = f(\mathbf{x}_{(i)})/q(\mathbf{x}_{(i)})$, of $\mathbf{x}_{(i)}$.
6: **end for**

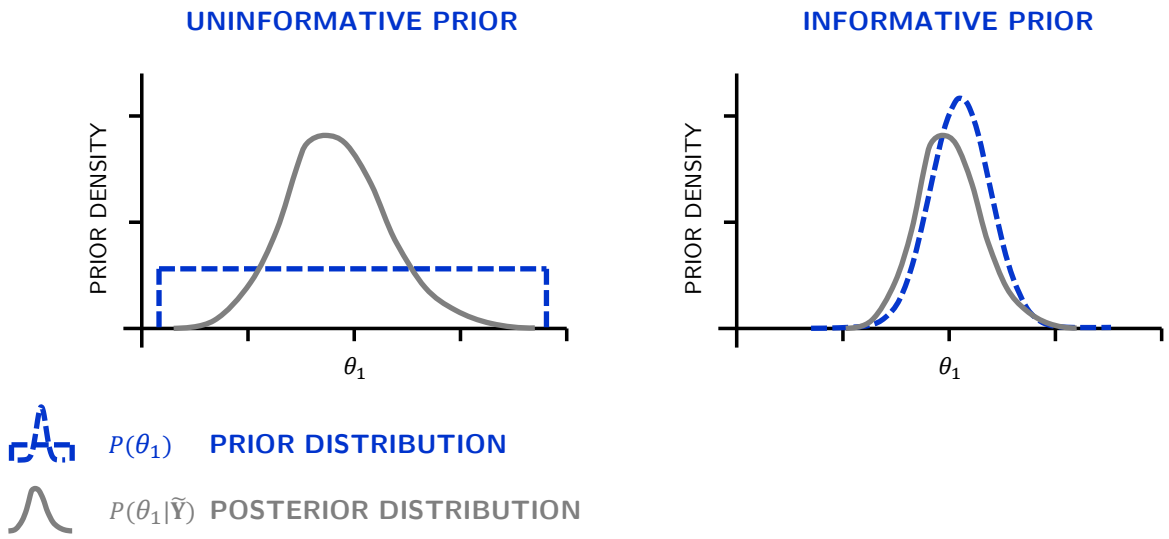

Figure 1.3: The prior distribution, $P(\boldsymbol{\theta})$, reflects all knowledge about the entity $\boldsymbol{\theta}$ before collecting data, $\widetilde{\mathbf{Y}}$, (referred to as evidence) through field and/or laboratory experimentation. An uninformative prior (left) does not favor a-priori any particular value of the parameter, yet can elicit "objective" ranges. An informative prior distribution (right) expresses varying levels of favor to different values of the parameter. Some values of the parameter are deemed more plausible a-priori than other feasible solutions.

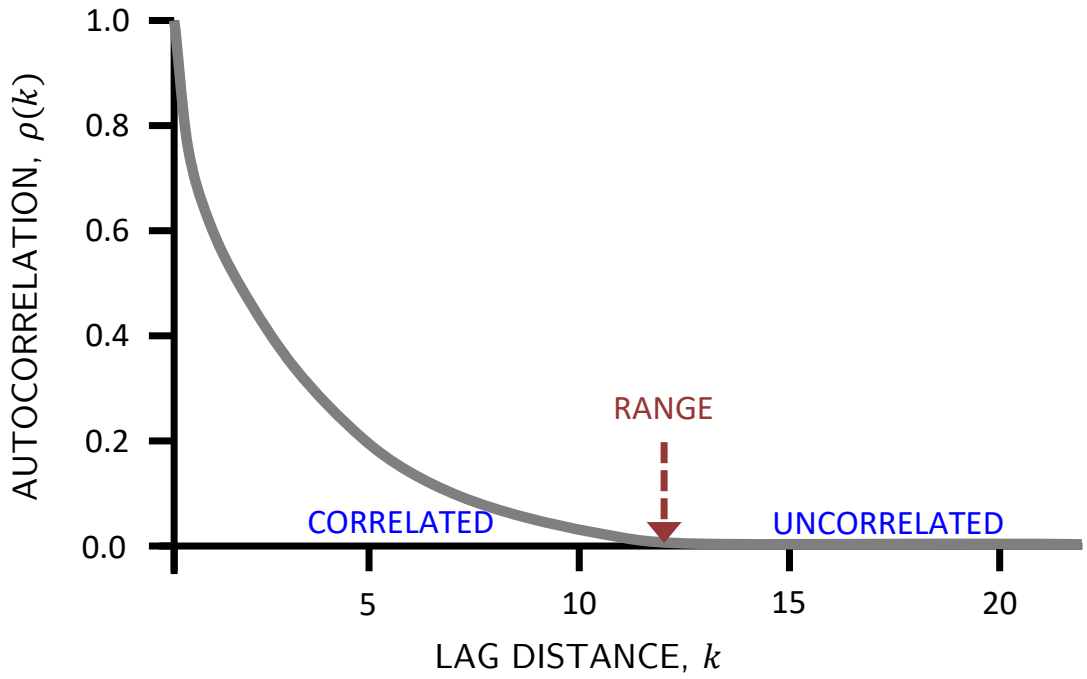

Figure 1.4: Illustrative example of the autocorrelation function of some hypothetical residual time series, $\mathbf{E}(\boldsymbol{\theta})$. The serial residual correlation, $\rho(k)$, is computed at different lags $k = \{1, \ldots, 22\}$ using Equation (1.32). These values are subsequently connected using the gray line. Per definition, the zeroth-order correlation is equal to unity, or $\rho(0) = 1$. This is easily shown as $e_t(\boldsymbol{\theta}) = e_{t+0}(\boldsymbol{\theta})$ and so the numerator in Equation 1.32) simplifies to $\sigma^2_{\mathbf{E}(\boldsymbol{\theta})}$, the denominator.

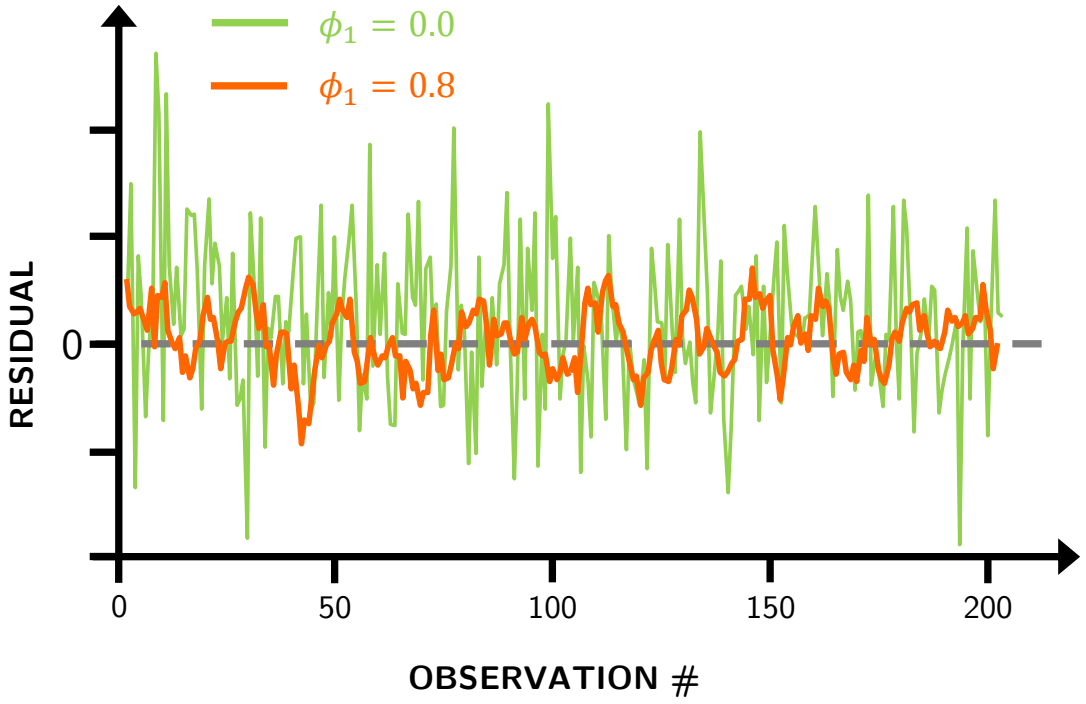

Figure 1.5: The impact of first-order serial correlation on the residuals. The green line lacks serial correlation, hence $\phi_1 = 0$, whereas the orange time series of residuals exhibits strong autocorrelation at the first lag, $\phi_1 = 0.8$. The dashed gray line signifies a perfect match between the model and the data. Note $c = 0$ and $\epsilon_t = 0 \; \forall t \in \{1, \ldots, 200\}$.

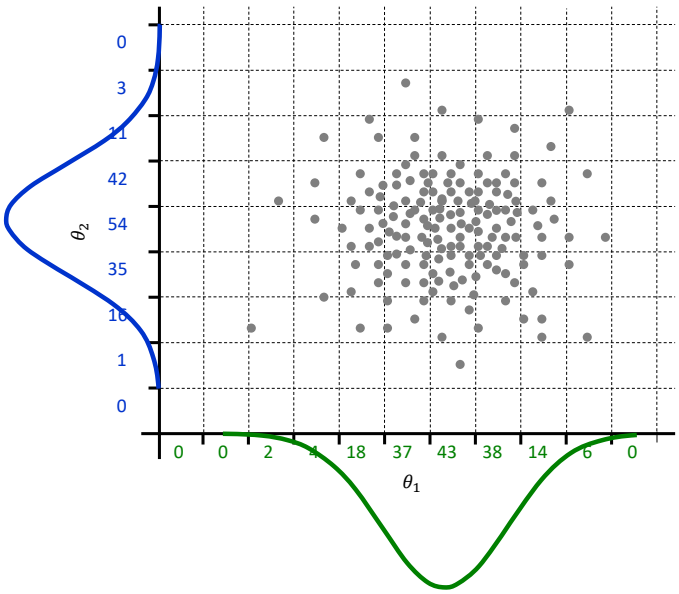

Figure 1.6: Scatter plot of $M = 160$ samples (gray dots) drawn at random from a bivariate normal distribution, $\mathcal{N}_2(\mathbf{a}, \Sigma$, The dashed lines delineate the different bins that are used to construct the frequency distribution (histogram) of θ_1 and θ_2. The green and blue lines depict the marginal distributions of the parameters and are inferred from the sampled points.

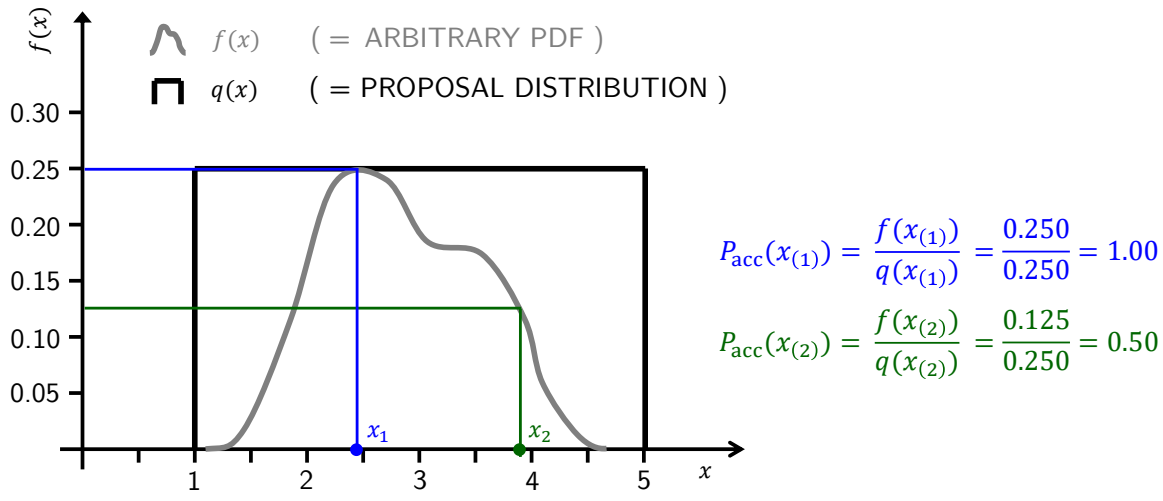

Figure 1.7: Application of rejection sampling to some univariate target distribution, $F(x)$, with arbitrary density function, $f(x)$ (in gray) using as proposal density (in black) the uniform distribution, $Q(x) \sim \mathcal{U}(1, 5)$.

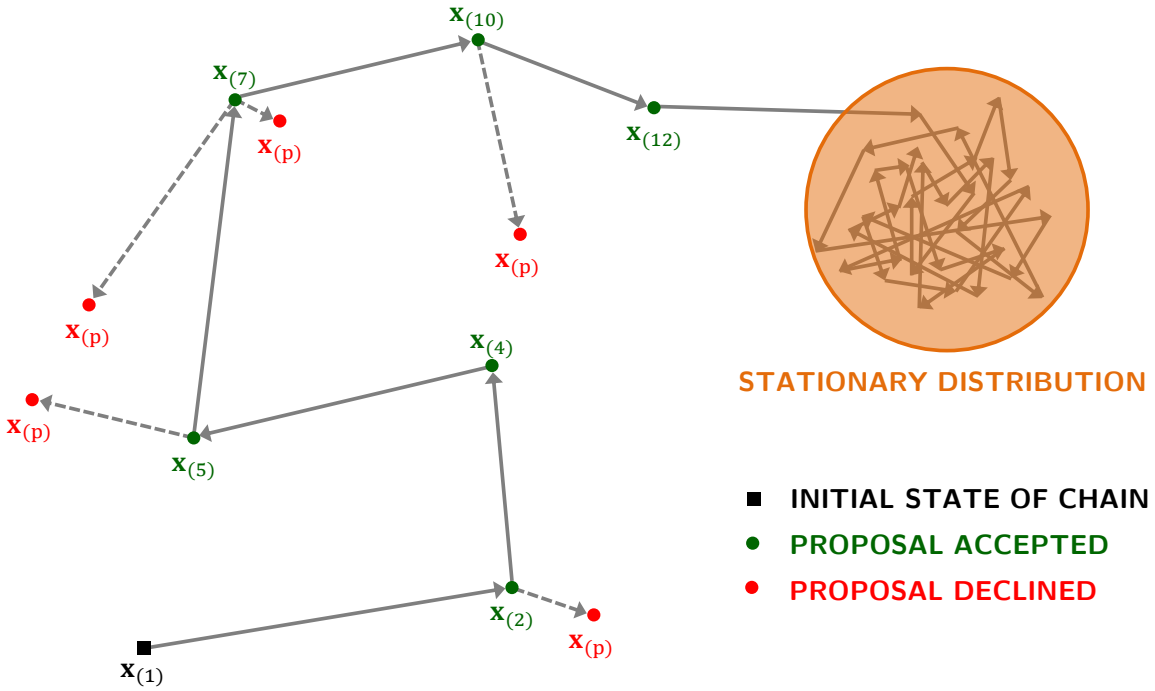

Figure 1.8: Schematic illustration of a chain trajectory derived from MCMC simulation for a two-dimensional target distribution, $f(\mathbf{x})$. The black square signifies the initial state of the chain, and the gray arrows denote the different jumps. Color coding is used to differentiate between proposals (candidate points or trial moves) that have been accepted (green dots) or rejected (red dots), respectively. The orange circle signifies the area of the stationary distribution, which envelops the target distribution. The number of times each chain position appears in the Markov chain (replicates after rejection of proposals) is directly proportional to the density of the target distribution.

Algorithm 3 Random Walk Metropolis

Define the proposal distribution, $q(\cdot)$ (must satisfy $q(a|b) = q(b|a)$) .
Define M.
Draw randomly $\mathbf{x}_{(1)}$, the initial state of the chain, and calculate $f(\mathbf{x}_{(1)})$.
for $i = 2, \ldots, M$ **do**
    Sample randomly a candidate point, $\mathbf{x}_{\mathrm{p}}$, from the symmetric proposal distribution, $q(\mathbf{x}|\mathbf{x}_{(i-1)})$, centered on $\mathbf{x}_{(i-1)}$, thus $\mathbf{x}_{\mathrm{p}} \sim q(\mathbf{x}|\mathbf{x}_{(i-1)})$.
    Compute the target density, $f(\mathbf{x}_{\mathrm{p}})$, at $\mathbf{x}_{\mathrm{p}}$.
    Calculate the Metropolis ratio, $P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}}) = \min\big(1, f(\mathbf{x}_{\mathrm{p}})/f(\mathbf{x}_{(i-1)})\big)$.
    if $Z \leq P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}})$ **then**
        Set $\mathbf{x}_{(i)} = \mathbf{x}_{\mathrm{p}}$ and $f(\mathbf{x}_{(i)}) = f(\mathbf{x}_{\mathrm{p}})$,
    else
        Remain at "old" state, that is $\mathbf{x}_{(i)} = \mathbf{x}_{(i-1)}$ and $f(\mathbf{x}_{(i)}) = f(\mathbf{x}_{(i-1)})$.
    end if
end for

Algorithm 4 Adaptive Metropolis

Calculate $s_{(d)} = 2.38^2/d$, define $d \times d$ covariance matrix, $\mathbf{\Sigma}$, $\varphi = 10^{-6}$, $m \geq 1$, and M.
Draw randomly $\mathbf{x}_{(1)}$, the initial state of the chain, and calculate $f(\mathbf{x}_{(1)})$.
for $i = 2, \ldots, M$ **do**
    Sample randomly a candidate point, $\mathbf{x}_{\mathrm{p}} \sim \mathcal{N}_d(\mathbf{x}_{(i-1)}, s_{(d)}\mathbf{\Sigma})$.
    Compute the target density, $f(\mathbf{x}_{\mathrm{p}})$, at $\mathbf{x}_{\mathrm{p}}$.
    Calculate the Metropolis ratio, $P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}}) = \min\big(1, f(\mathbf{x}_{\mathrm{p}})/f(\mathbf{x}_{(i-1)})\big)$.
    if $Z \leq P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}})$ **then**
        Set $\mathbf{x}_{(i)} = \mathbf{x}_{\mathrm{p}}$ and $f(\mathbf{x}_{(i)}) = f(\mathbf{x}_{\mathrm{p}})$.
    else
        Remain at "old" state, $\mathbf{x}_{(i)} = \mathbf{x}_{(i-1)}$ and $f(\mathbf{x}_{(i)}) = f(\mathbf{x}_{(i-1)})$.
    end if
    if $\mathrm{mod}(i, m) = 0$ **then**
        Adapt covariance matrix, $\mathbf{\Sigma} = \mathrm{Cov}[\{\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(i)}\}] + \varphi\mathbf{I}_d$.
    end if
end for

Table 1.1 Model parameters and their prior uncertainty ranges.

Parameter	Symbol	Minimum	Maximum	Units
Maximum interception	$I_{\max}$	1	10	mm
Soil water storage capacity	$S_{\max}$	10	1000	mm
Maximum percolation rate	$Q_{\max}$	0.1	100	mm/d
Evaporation parameter	α_{E}	0.1	100	-
Runoff parameter	α_{F}	-10	10	-
Time constant, fast reservoir	K_{F}	0.1	10	days
Time constant, slow reservoir	K_{S}	0.1	150	days

Algorithm 5 Differential Evolution Markov chain

Define number of chains $N \geq 2d$, and $c_* = 10^{-6}$.
Set iteration, $i = 2$.
for $j = 1, \ldots, N$ **do**
  Draw randomly $\mathbf{x}^j_{(1)}$, the initial state of the jth chain, and calculate $f(\mathbf{x}^j_{(1)})$.
end for
while chains not converged **do**
  for $j = 1, \ldots, N$ **do**
    Draw without replacement integers a and b from $\{1, \ldots, j-1, j+1, \ldots, N\}$
    Select the jump rate; if $\mathcal{U}(0,1) \leq 0.9$ then $\gamma_{(d)} = 2.38/\sqrt{2d}$ otherwise $\gamma_{(d)} = 1$.
    Create a candidate point, $\mathbf{x}^j_{\mathrm{p}}$, in the jth chain using Equation (1.46)
    Compute the target density, $f(\mathbf{x}^j_{\mathrm{p}})$, at $\mathbf{x}^j_{\mathrm{p}}$.
    Calculate the Metropolis ratio, $P_{\mathrm{acc}}(\mathbf{x}^j_{\mathrm{p}}) = \min\big(1, f(\mathbf{x}^j_{\mathrm{p}})/f(\mathbf{x}^j_{(i-1)})\big)$.
    if $Z \leq P_{\mathrm{acc}}(\mathbf{x}^j_{\mathrm{p}})$ **then**
      Set $\mathbf{x}^j_{(i)} = \mathbf{x}^j_{\mathrm{p}}$ and $f(\mathbf{x}^j_{(i)}) = f(\mathbf{x}^j_{\mathrm{p}})$.
    else
      Remain at "old" state, $\mathbf{x}^j_{(i)} = \mathbf{x}^j_{(i-1)}$ and $f(\mathbf{x}^j_{(i)}) = f(\mathbf{x}^j_{(i-1)})$.
    end if
  end for
  Compute convergence diagnostics
  Update iteration, $i = i + 1$.
end while

Table 1.2 Parameters of the HYDRUS-1D model and their prior uncertainty ranges.

Parameter	Symbol	Lower	Upper	Units	Prior
Residual moisture content	θ_r	0.043	0.091	$\mathrm{cm^3\ cm^{-3}}$	$\mathcal{N}(0.067, 0.006)$
Saturated moisture content	θ_s	0.409	0.481	$\mathrm{cm^3\ cm^{-3}}$	$\mathcal{N}(0.445, 0.009)$
Reciprocal of air-entry value	α	0.003	0.009	$\mathrm{cm^{-1}}$	$\mathcal{N}(0.005, 6.90 \cdot 10^{-4})$
Curve shape parameter	n	1.510	1.849	-	$\mathcal{N}(1.671, 0.042)$
Conductivity at saturation	K_s	0.138	19.962	$\mathrm{cm\ day^{-1}}$	$\mathcal{N}(1.660, 1.386)$
Tortuosity parameter	λ	-5.490	6.270	-	$\mathcal{N}(0.390, 1.470)$
Pressure head at bottom	h_{bot}	-250	-50	cm	$\mathcal{U}(-250, 50)$

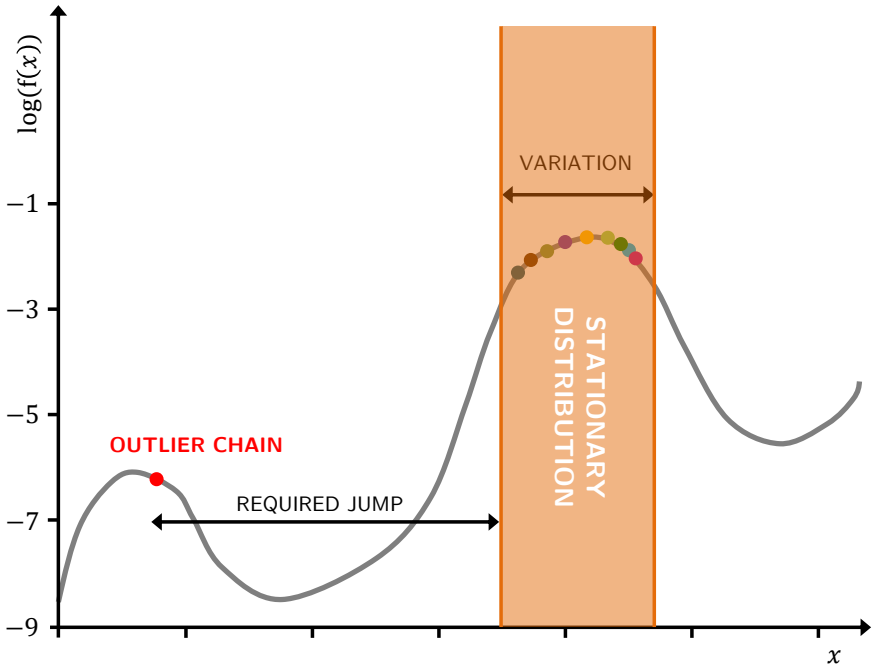

Figure 1.9: Schematic illustration of a dissident chain (red) that is mired in a local basin of attraction of the search space of some univariate probability distribution. This outlier chain cannot transition to the target distribution demarcated in orange. The reasons are as follows. First, the aberrant chain cannot travel in multiple iterations the valley as almost all trial moves will exhibit a negligible acceptance rate. Second, a direct move to a point with equal probability density at the other side of the valley is implausible as the variation among the target chains is insufficient to warrant a large enough jump size with Equation (1.46). Third, a direct move to the target with unit jump rate is impossible as the outlier chain cannot sample its own position ($a \neq b \neq i$) - a requirement for each chain trajectory to satisfy detailed balance. As a consequence, the dissident chain will be trapped forever.

Algorithm 6 DiffeRential Evolution Adaptive Metropolis

Define number of chains $N \geq \lfloor d/2 \rfloor$.
Define algorithmic variables, r, $c = 0.1$ and $c_* = 10^{-6}$.
Compute r crossover values, $\boldsymbol{\eta} = \{\eta_1, \ldots, \eta_r\}$ with $\mathbf{p}_\eta = \{\frac{1}{r}, \ldots, \frac{1}{r}\}$.
Set iteration, $i = 2$.
for $j = 1, \ldots, N$ **do**
    Draw randomly $\mathbf{x}_{(1)}^j$, the initial state of the jth chain, and calculate $f(\mathbf{x}_{(1)}^j)$.
end for
while chains not converged **do**
    for $j = 1, \ldots, N$ **do**
        Draw the crossover value, η, from $\mathcal{F}(\boldsymbol{\eta}|\mathbf{p}_\eta)$.
        Draw a d-vector $\mathbf{z}$ from $\mathcal{U}_d(0, 1)$.
        Store in subset A the indexes l of $\mathbf{z}$ that satisfy $\mathbf{z}_l \leq \eta$, where $l = \{1, \ldots, d\}$.
        If $A = \emptyset$ (empty set), then fill A with random draw from integers $\{1, \ldots, d\}$.
        Compute the cardinality of A, that is $d^* = |A|$.
        Draw the value of δ at random from integers, $\{1, 2, 3\}$, thus $\delta \sim \mathcal{U}\{1, 3\}$.
        Sample δ-vectors $\mathbf{a}$ and $\mathbf{b}$ without replacement from $\{1, \ldots, j-1, j+1, \ldots, N\}$.
        Draw a label $R \sim \mathcal{U}(0, 1)$, if $R \leq 0.8$ set $\gamma_{(\delta,d^*)} = 2.38/\sqrt{2\delta d^*}$ otherwise $\gamma_{(\delta,d^*)} = 1$.
        Create a candidate point, $\mathbf{x}_{\mathrm{p}}^j$, in the jth chain using Equations (1.48) and (1.49).
        Compute the target density, $f(\mathbf{x}_{\mathrm{p}}^j)$, at $\mathbf{x}_{\mathrm{p}}^j$.
        Calculate the Metropolis ratio, $P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}}^j) = \min\big(1, f(\mathbf{x}_{\mathrm{p}}^j)/f(\mathbf{x}_{(i-1)}^j)\big)$.
        if $Z \leq P_{\mathrm{acc}}(\mathbf{x}_{\mathrm{p}}^j)$ **then**
            Set $\mathbf{x}_{(i)}^j = \mathbf{x}_{\mathrm{p}}^j$ and $f(\mathbf{x}_{(i)}^j) = f(\mathbf{x}_{\mathrm{p}}^j)$.
        else
            Remain at "old" state, $\mathbf{x}_{(i)}^j = \mathbf{x}_{(i-1)}^j$ and $f(\mathbf{x}_{(i)}^j) = f(\mathbf{x}_{(i-1)}^j)$.
        end if
    end for
    Compute convergence diagnostics.
    Patch for dissident chains.
    Update iteration, $i = i + 1$.
end while

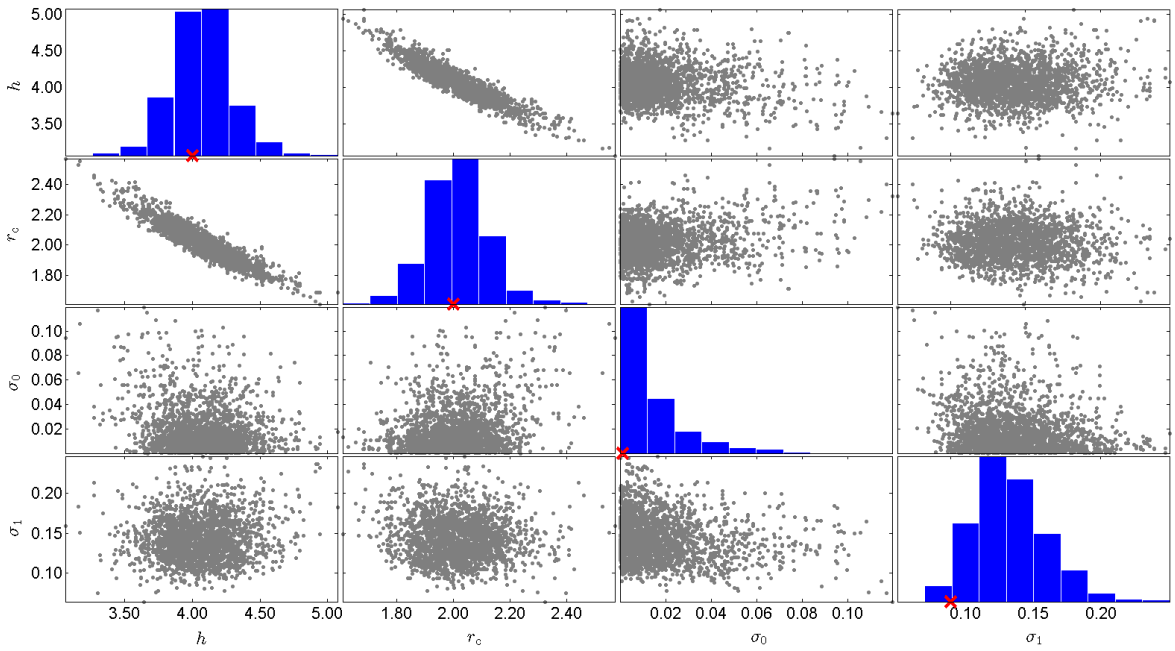

Figure 1.10: Unit hydrograph: Scatter-plot matrix of the posterior samples created the DREAM algorithm. The main diagonal presents histograms of the marginal posterior distributions of the Nash model parameters, h and r_c, and nuisance variables, σ_0 and σ_1. The off-diagonal graphs display bivariate scatter plots of the posterior samples of the different parameter pairs. The true parameter values are separately indicated in each histogram using the red cross.

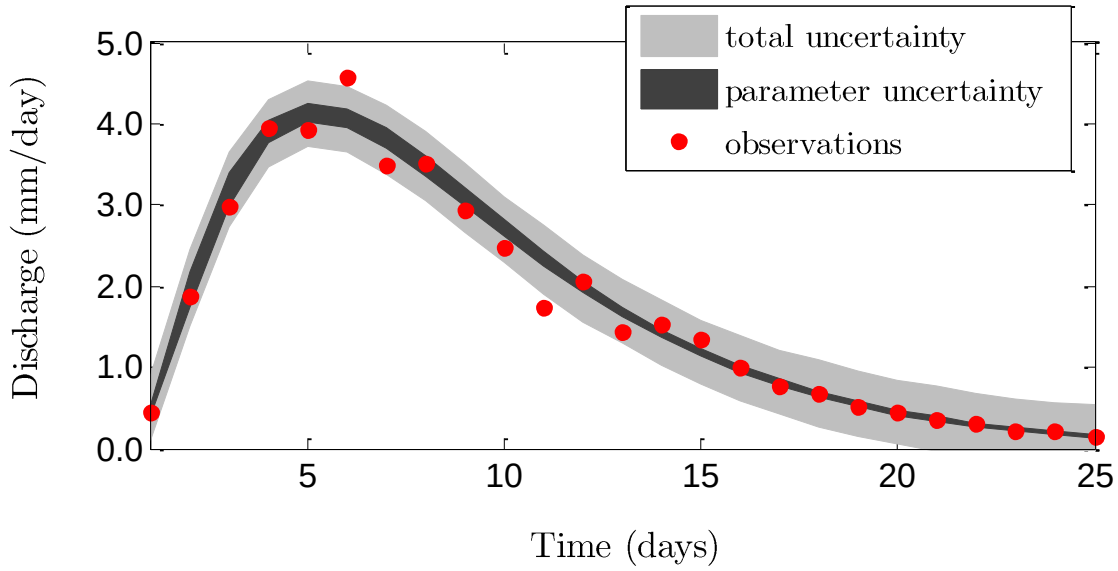

Figure 1.11: Unit hydrograph: Comparison of the observed (red dots) and posterior simulated hydrographs. The light and dark gray regions demarcate the 95% simulation intervals of the Nash model due to parameter and total uncertainty, respectively.

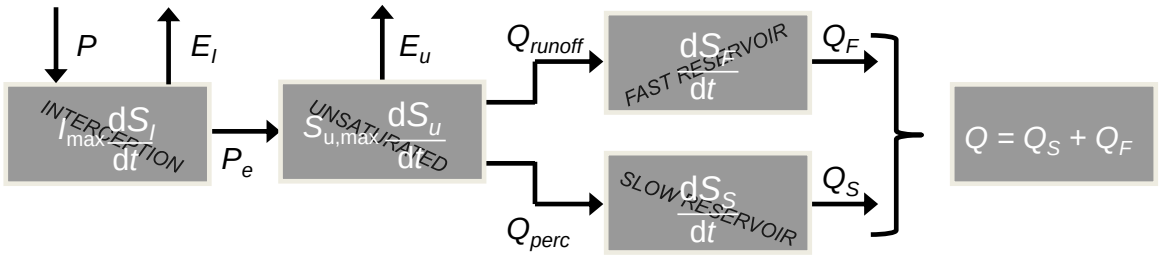

Figure 1.12: The rainfall-runoff transformation: Schematic representation of the hmodel conceptual watershed model.

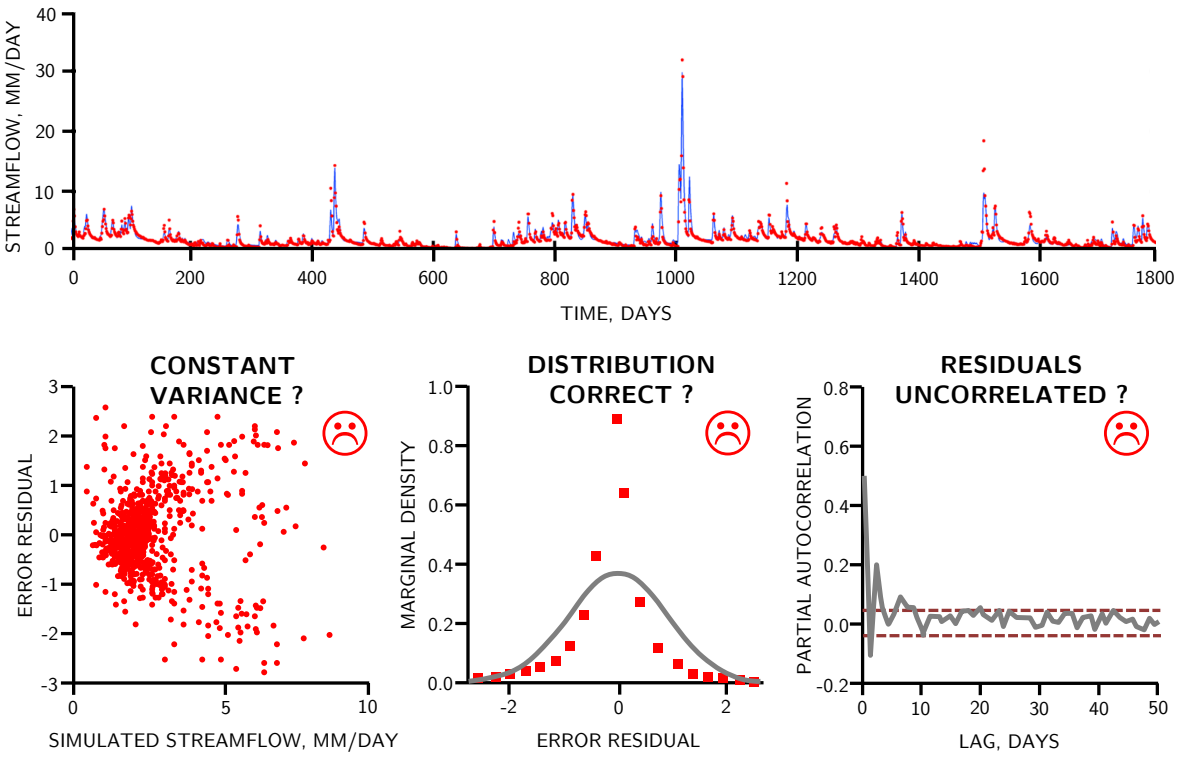

Figure 1.13: The rainfall-runoff transformation: Least-squares calibration with Gaussian likelihood of Equation (1.28) for the French Broad River basin. Time series plot (top panel) of maximum likelihood streamflow simulation (solid blue line) and observations (red dots). The bottom panel analyzes the corresponding residuals. The left plot displays the residuals as a function of the simulated flow level. The middle plot shows the assumed (solid line) and actual (red squares) histogram of the residuals. The right plot summarizes the autocorrelation function of the residuals. The 95% significance levels are separately indicated in this graph with the dashed lines.

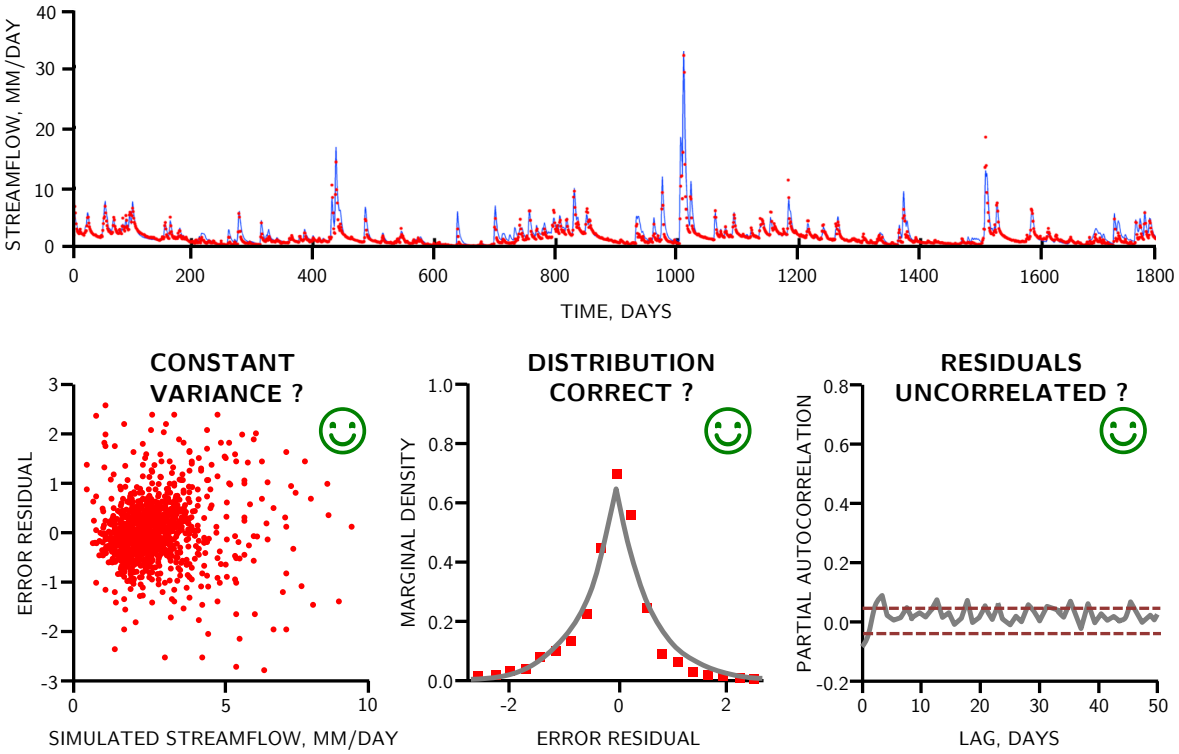

Figure 1.14: The rainfall-runoff transformation: Calibration with the generalized likelihood function for the French Broad River basin. Time series plot (top panel) of maximum likelihood streamflow simulation (solid blue line) and observations (red dots). The bottom panel analyzes the corresponding residuals. The left plot displays the residuals as a function of the simulated flow level. The middle plot shows the assumed (solid line) and actual (red squares) histogram of the residuals. The right plot summarizes the autocorrelation function of the residuals. The 95% significance levels are separately indicated in this graph with the dashed lines

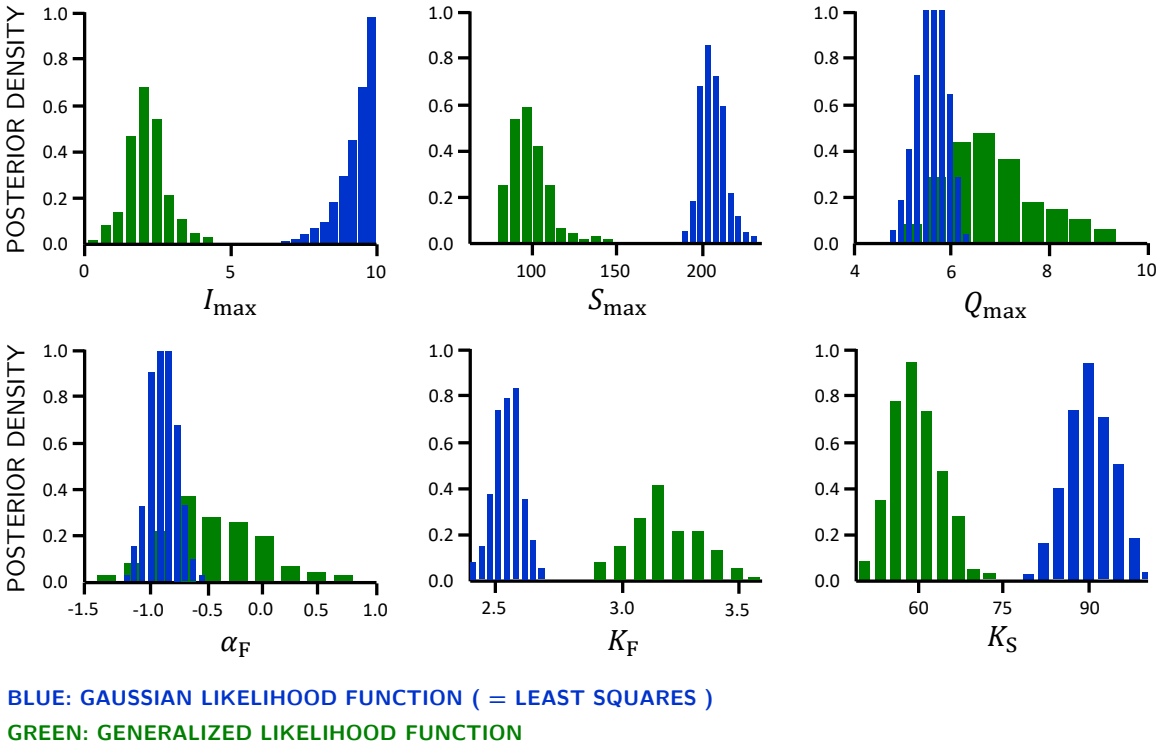

Figure 1.15: The rainfall-runoff transformation: Histograms of the marginal posterior distribution of the hmodel parameters for the French Broad River basin using (in blue) the Gaussian likelihood and (in green) the generalized likelihood.

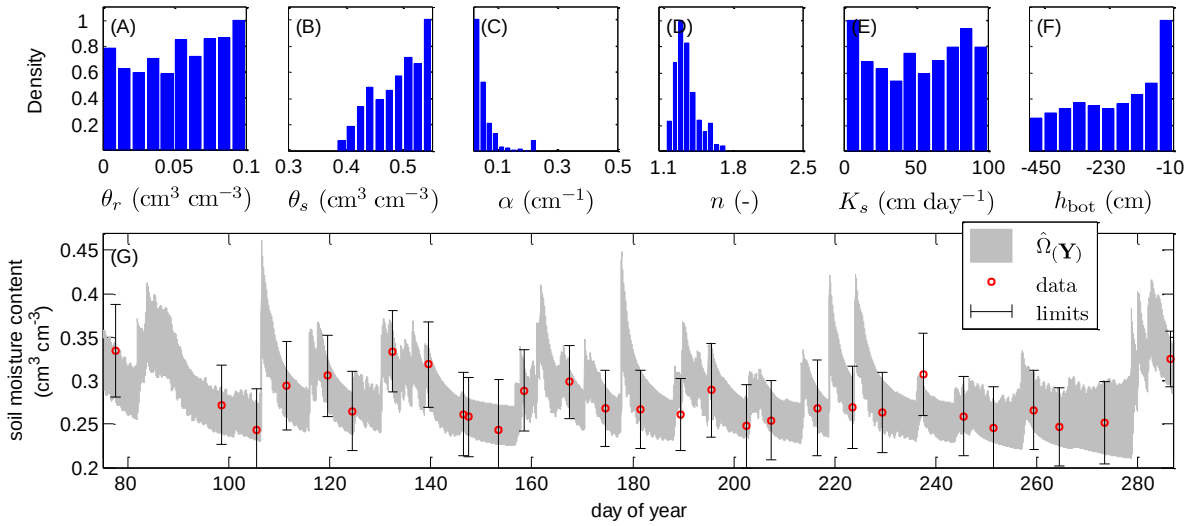

Figure 1.16: Vadose zone hydrology: Histograms of the marginal posterior distribution of the soil hydraulic parameters, (A) θ_s, (B) α, (C) n, and (D) K_s, and (E) HYDRUS-1D 95% simulation uncertainty intervals due to parameter (dark region) and total uncertainty (light gray). The observed soil moisture value are indicated with a red circle.

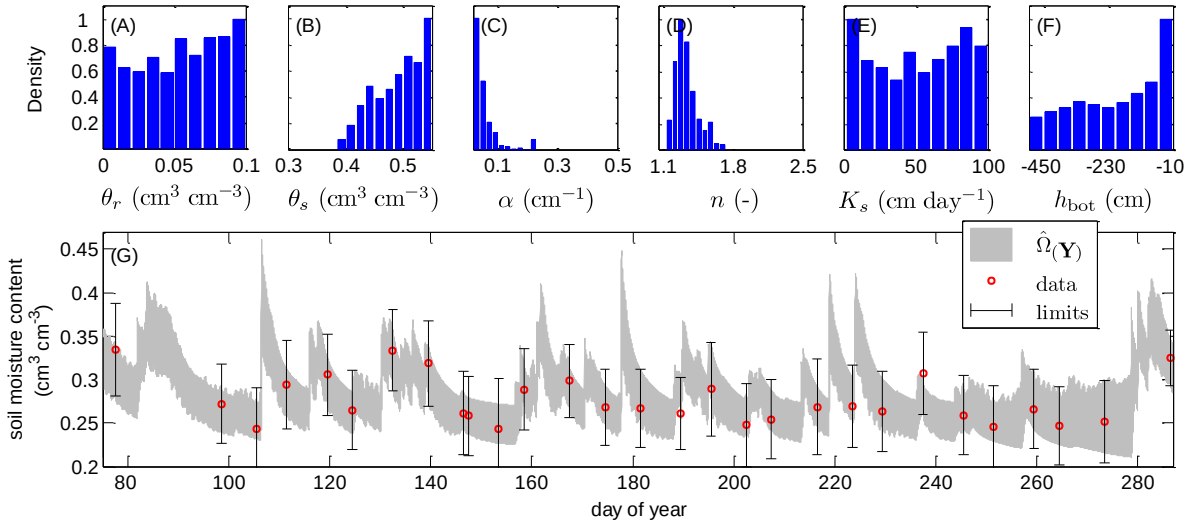

Figure 1.17: Vadose zone hydrology: (top panel) Histograms of the behavioral set, $\widehat{\Omega}_{(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}$, of the soil hydraulic parameters, (A) θ_r, (B) θ_s, (C) α, (D) n, (E) K_s, and (F) $h_{\rm bot}$. The x-axis matches exactly the (uniform) prior distribution. (bottom panel) Comparison of observed (red dots) and posterior simulated, $\widehat{\Omega}_{(\mathbf{Y})}$ (gray region) soil moisture contents.

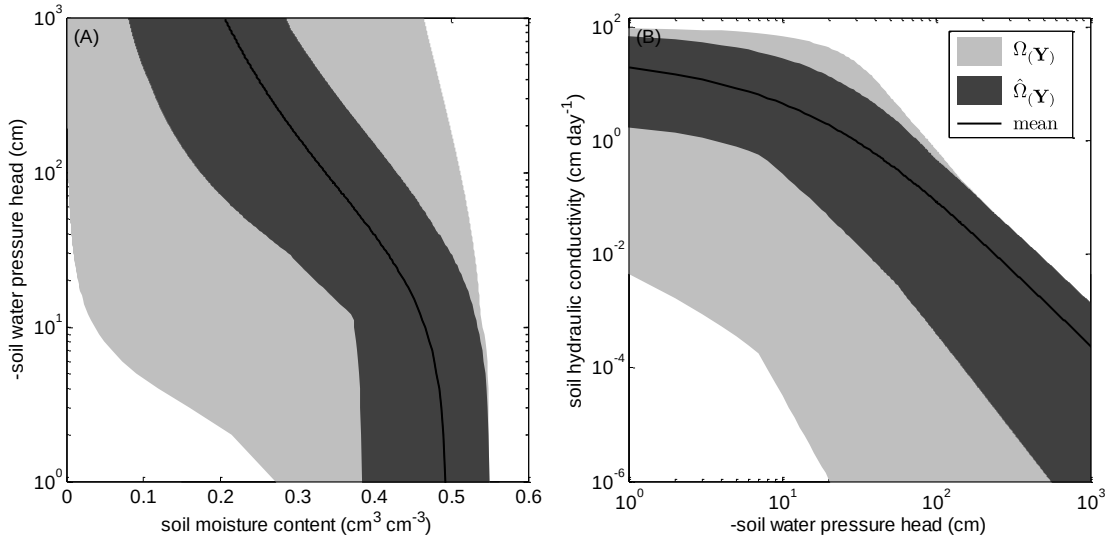

Figure 1.18: Vadose zone hydrology: Comparison of the prior (dark gray) and posterior (light gray) ranges of the (left) soil water retention, and (right) unsaturated soil hydraulic conductivity function. The posterior (or behavioral) mean hydraulic functions are indicated separately with the solid black line.

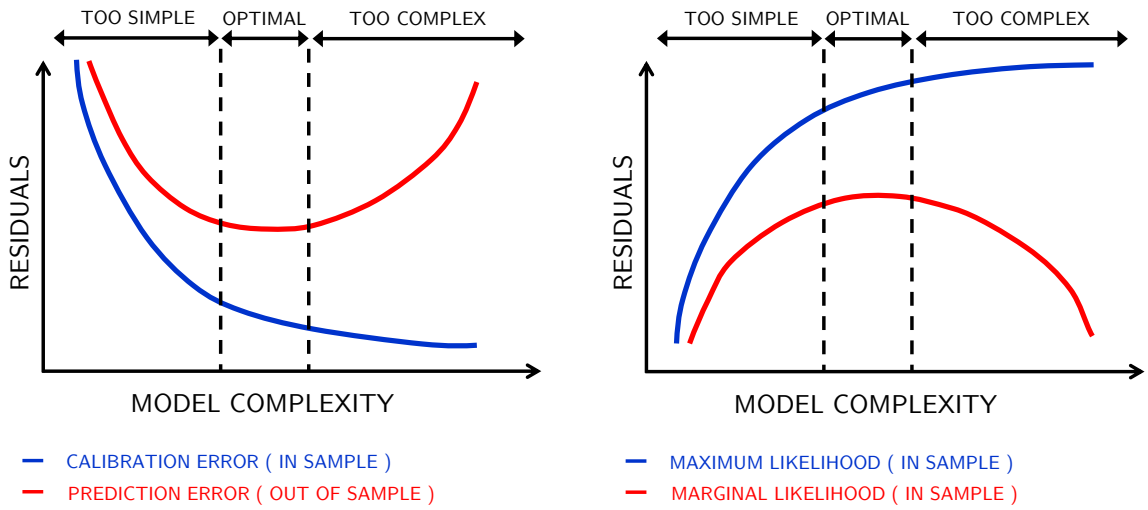

Figure 1.19: Schematic overview of the relationship between model complexity (number of parameters) and (left plot) the magnitude of the error residuals in the calibration and evaluation period, and (right plot) the value of the maximum likelihood, $L(\boldsymbol{\theta}^*|\widetilde{\mathbf{Y}})$, and the marginal likelihood, $P(\widetilde{\mathbf{Y}})$ of the calibration data set.

Chapter 2

Improving Predictions in Ecology with Data Assimilation

by Elias C. Massoud, Elisa Beninca, Jef Huisman, Willem Bouten and Jasper A. Vrugt

Reference:

Massoud, E. C., Huisman, J., Beninca, E., Heerklos, R., Vrugt, J. A.: Probing the limits of predictability: data assimilation improves forecasts of complex dynamics in ecosystems. Ecology

Letters, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017.

Many systems in ecology have shown to exhibit chaotic dynamics. As a consequence, predictability of these systems is limited in the long term. In a previous study Benincà et al. (2008) demonstrate chaos in a long term experiment of a plankton community, and the forecast window for this system was around 15-30 days. Here, we fuse this unique data set with a general coupled predator prey model and successfully model the species dynamics using the data assimilation (DA) technique. To our knowledge, this is the first successful attempt to model predator prey dynamics that exhibit chaotic behavior. DA methods have been introduced in many fields, such as meteorology, but have rarely been used in ecological applications despite the recognition in recent literature of its potential benefits (Peng et al., 2011).

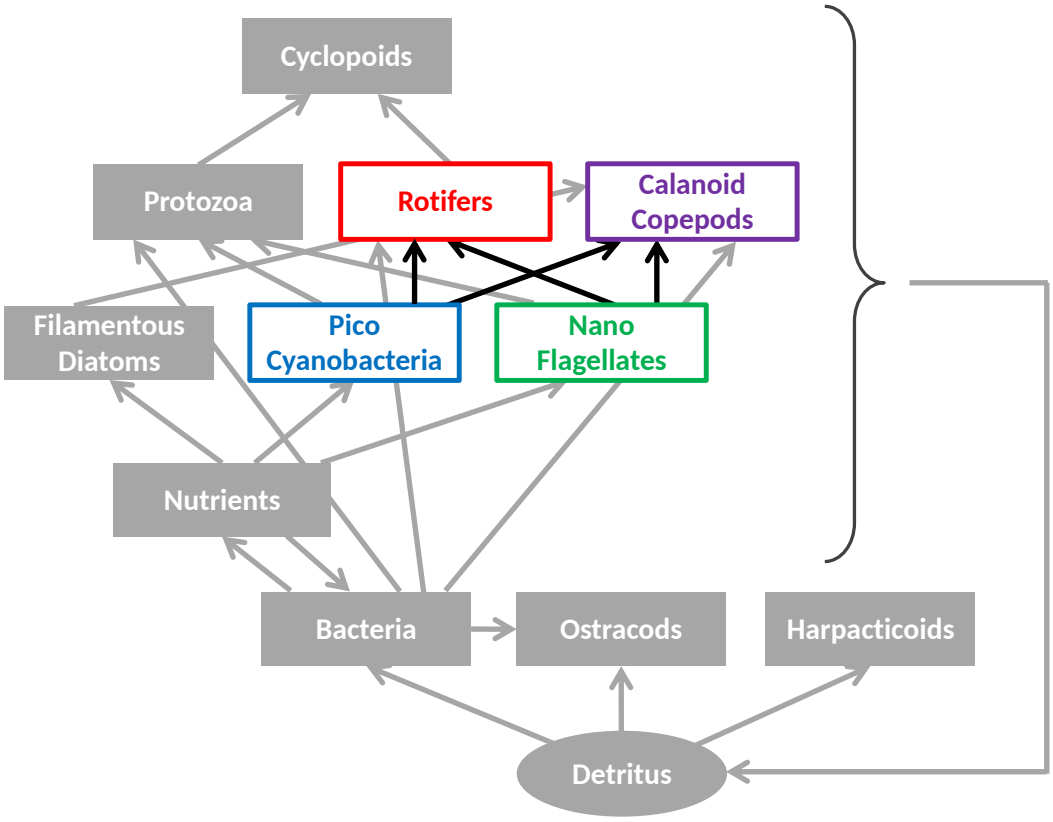

Figure 2.1: Food web structure of the experimental data. Our analysis focuses on two prey species (picocyanobacteria and nanoflagellates) and two predator species (rotifers and calanoid copepods), shown in color. Modified from Benincà et al. (2008).

After successful application of DA, its efficiency is examined when used at reduced frequencies. This is done because the application of DA is costly and requires frequently sampled data to update

the model simulations. This analysis provides insight on measurement design and shows how the accuracy of the model simulations changes with reduced measurement frequencies.

Lastly, we estimate the Ecological Forecast Horizons (EFH) of the model simulations (Petchey et al., 2015), which ultimately show how far into the future can the model projections be trusted without the help from DA. We achieve acceptable projections that span roughly 80 days in some cases, absolutely surpassing the previous benchmark of 15-30 days for this system.

2.1 Manuscript Title - Probing the limits of predictability: data assimilation improves forecasts of complex dynamics in ecosystems

Abstract: The daunting complexity of ecosystems has led ecologists to use mathematical modeling to gain understanding of ecological relationships, to simulate long-term ecological processes, and to predict future system dynamics. In pursuit of mathematical tractability, these models use simplified descriptions of key patterns, processes and relationships observed in nature. In contrast, ecological data are often complex, scale-dependent, space-time correlated, and governed by nonlinear relations between organisms and their environment. This disparity in complexity between ecosystem models and data has created a large gap in ecology between model and data-driven approaches. Here, we explore data assimilation (DA) with the Ensemble Kalman Filter (EnKF) to fuse a two-predator-two-prey model with abundance data from a 2600+ day experiment of a plankton community. We analyze the chaotic population dynamics using different assimilation intervals. Overall, results show that DA improves substantially the predictability and ecological forecast horizon of complex community dynamics.

2.2 Introduction

Ecosystems constitute a complex network of living organisms, which are interconnected and linked with their environment through a myriad of nutrient cycles, mass and energy flows. This may

give rise to emergent and self-organized behavior with structural, spatial, and temporal patterns and dynamics of the organisms that are difficult, costly, and labor intensive to measure directly in the field, particularly at large spatial scales (Hari et al., 2013; Schaeffer et al., 1988). Faced with this daunting complexity, ecologists have developed mathematical models to gain understanding of ecosystem functioning, simulate large-scale experiments, and predict far into the future ecological processes. These models typically use simplified mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated nutrient, energy, mass, and ecological processes in an ecosystem (Running et al., 1999; Vandermeer, 2004). Consequently, most ecological models may explain only a small fraction of the dynamics of the actual ecosystem (Wintle et al., 2003). What is more, due to process and spatial aggregation the parameters of an ecological model often do not represent directly measurable ecosystem quantities and must therefore be estimated indirectly through a calibration process using measurements of ecosystem inputs and outputs. The "calibrated" model can then be used to predict ecological processes over very long periods of time (Fasham et al., 1990; Sitch et al., 2003).

During the past few decades increasingly larger volumes of ecological data have become available in response to continued advances in measurement techniques and the rapid expansion of long-term monitoring networks (Aanensen et al., 2009; Running et al., 1999). This ever increasing wealth of data provides unique opportunities for ecologists to enhance ecosystem understanding and characterization (Reichman et al., 2011; Ter Braak and Van Tongeren, 1995). Yet, ecological data are often complex, high-dimensional, and scale-dependent as governed by local interactions and feedback loops between organisms individually and their environment, and periodic and stochastic changes to the ecosystem (Benincà et al., 2009; Conway et al., 1970).

In this paper, we advocate use of advanced statistical methods to improve treatment of model and measurement errors, and reconcile ecological models with data (Kendall et al., 1999; Peng et al., 2011). This includes state-of-the-art parameter estimation (Ali et al., 2015; Conroy et al., 1995) and data assimilation (DA) methods (Luo et al., 2011; Luo and Schimel, 2011). DA incorporates observations of system behavior into the stace space of a computer model of this system. The prototype of this method, the Kalman filter (KF), was developed by Kalman (1960) for optimal

control of dynamical systems. DA holds great promise in ecology as it will help close the gap between ecosystems models and data, enhance ecological forecasting (McMahon et al., 2009), provide insights into measurement and model errors, and guidance on "optimal" experimental design. For example, via integration with inferential ecosystem models, DA can help further refine wireless sensor networks by weighing the value of an observation against cost of data collection (Clark et al., 2001).

DA has found widespread application and use in many different fields of research, including oceanography (Bertino et al., 2003; Gehlen et al., 2015), marine ecology (Triantafyllou et al., 2013; Xiao and Friedrichs, 2014), hydrology (Vrugt et al., 2005), glaciology (Granzow, 2014), and satellite remote sensing (Dorigo et al., 2007), to name a few. Furthermore, DA has received operational status in real-time weather, traffic, tsunami, and flood prediction systems because of its proven ability to enhance significantly the forecast skill of dynamic system models, and quantify accurately prediction uncertainty. In ecology, the interest in DA has increased rapidly in the past decade (Chen et al., 2008; Mo et al., 2008; Quaife et al., 2008; Williams et al., 2005), to support ecological analysis (Zobitz et al., 2011), to account for model structural, input and output errors (Luo et al., 2011), to constrain model parameters and system states, and improve ecological prediction (Niu et al., 2014), and shed new light on model malfunctioning, and provide guidance on model improvement and data informativeness (Keenan et al., 2011; Sitz et al., 2002; Vrugt et al., 2005). In fact, next generation ecological models are developed conscientiously in anticipation of DA applications (Peng et al., 2011; Williams et al., 2009; Wu et al., 2009; Xu et al., 2012) .

To date, only few authors have used DA to analyze chaotic predator prey dynamics in complex microbial food webs. For example, Lawson et al. (1995) used an adjoint DA method to assimilate synthetic abundance data into a simple predator-prey model, and Lawson et al. (1996) recovered rate parameters, initial conditions, and the amplitude of episodic events in a marine ecosystem model from a 60-day assimilation period using simulated monthly, bi-weekly and weekly distributions of nutrients, phytoplankton and zooplankton. Here, we present the first application of DA to chaotic population dynamics in a long-term laboratory experiment with a marine plankton community

isolated from the Baltic Sea (Benincà et al., 2008). The dataset displayed chaotic population dynamics with species abundances that showed striking fluctuations over several orders of magnitude. Chaos implies that long-term prediction of the population dynamics in our food web is fundamentally impossible. We use the Ensemble Kalman Filter, or EnKF (Evensen, 1994) to fuse the predicted abundances of the two-predator-two prey model of Vandermeer (2004) with measured abundances. The parameters are simultaneously estimated, by minimizing in a loop outside the EnKF, the forecast error of the model. We follow guidelines from Petchey et al. (2015) and compute the ecological forecast horizons of our population model to determine how far into the future useful forecasts can be made (Simmons and Hollingsworth, 2002). Altogether, results demonstrate that joint parameter and state estimation enhances significantly the forecast horizon and predictive ability of a population model.

2.3 Materials and Methods

This section presents our application of DA to provide guidelines for researchers new to model-data fusion methods. We first review our food web and data, then discuss the Vandermeer (VD) population model (Vandermeer, 2004), and then present a brief technical description of the parameter estimation and DA methods used to close the gap between observed and simulated abundances. We conclude this section with a description of the concept of ecological forecast horizons of Petchey et al. (2015) to evaluate DA with different VD model parameterizations.

2.3.1 The data set

We use a microbial plankton community isolated from the Darss-Zingst estuary in the southern Baltic Sea. The structure of this food web is depicted graphically in Benincà et al. (2008) and consists of bacteria, several phytoplankton species, herbivorous and predatory zooplankton species, and detritivores. The plankton community was cultured in a laboratory mesocosm under constant external conditions and sampled twice a week for a period of more than 8 years to count population abundances of the functional groups. A detailed description of the mesocosm experiment appears in

the Supplementary information (SI) of Benincà et al. (2008) and data preprocessing is discussed in Benincà et al. (2009). The final data set comprises 2656 days and consist of $n = 794$ observations of each species' population count (in (mg fwt $L^{-1})^{1/4}$) with constant measurement interval of $\Delta t = 3.35$ days.

The mesocosm data set provides an excellent test bed for DA as the measured species abundances show striking fluctuations over several orders of magnitude, despite the constant external conditions. What is more, Benincà et al. (2008) has demonstrated that the population dynamics exhibit chaos. We focus our attention to the rotifers, calanoid copepods, picocyanobacteria, and nanoflagellates, because these four species have a relatively large presence in our food web, govern a large part of the population dynamics, and their abundances resemble the oscillations typical for coupled predator-prey interactions. The $n = 794$ measured abundances of these $k = 4$ species are stored in a $k \times n$ matrix, $\widetilde{\mathbf{Y}}$, where $\widetilde{}$ signifies observed data.

2.3.2 Coupled predator-prey model

We analyze the dynamics of the two-predator-two-prey communities using the population model of Vandermeer (1982, 2004). This VD model assumes that the prey species interact through Lotka-Volterra competition, and are consumed by predators according to a saturating functional response. If P_1 and P_2 denote the abundances of the nanophytoplankton and picophytoplankton prey species, and Z_1 and Z_2 the abundances of the calanoid copepods and rotifers (competing predators), respectively, then the VD model is given by

$$\begin{aligned}
\frac{\mathrm{d}P_1}{\mathrm{d}t} &= r_1 P_1\Big(1 - \frac{1}{K}(P_1 + \alpha P_2)\Big) - \left(\frac{gP_1Z_1}{H + (P_1 + \beta P_2)} + \frac{g\beta P_1 Z_2}{H + (\beta P_1 + P_2)}\right) \\
\frac{\mathrm{d}P_2}{\mathrm{d}t} &= r_2 P_2\Big(1 - \frac{1}{K}(\alpha P_1 + P_2)\Big) - \left(\frac{g\beta P_2Z_1}{H + (P_1 + \beta P_2)} + \frac{gP_2 Z_2}{H + (\beta P_1 + P_2)}\right) \\
\frac{\mathrm{d}Z_1}{\mathrm{d}t} &= \frac{g(P_1 + \beta P_2)Z_1}{H + (P_1 + \beta P_2)} - mZ_1 \\
\frac{\mathrm{d}Z_2}{\mathrm{d}t} &= \frac{g(P_2 + \beta P_1)Z_2}{H + (\beta P_1 + P_2)} - mZ_2.
\end{aligned} \tag{2.1}$$

This system of four ordinary differential equations describes the coupled interactions of our two competing predators and two prey species and the evolution of their abundances under constant environmental conditions. The VD model has four state variables, namely the abundances, Z_1, Z_2 and P_1, P_2 of the two predators and two preys, respectively, and $d = 8$ parameters, whose values we store in vector $\boldsymbol{\theta}$. This includes the two unitless coefficients $\alpha \geq 0$ and $\beta \in [0, 1]$ which characterize the competition between the zooplankton and their predator selectivity, respectively, the growth rates of the first and second prey, r_1 (day^{-1}) and r_2 (day^{-1}), respectively, the grazing rate, g (day^{-1}), the carrying capacity, K ((mg fwt L^{-1})$^{1/4}$), the mortality rate, m (day^{-1}) and functional response, H ((mg fwt L^{-1})$^{1/4}$).

We write the VD model as a regression function

$$\mathbf{Y} = \mathcal{VD}(\boldsymbol{\theta}), \tag{2.2}$$

which uses as input the d-vector $\boldsymbol{\theta} = \{\alpha, \beta, r_1, r_2, g, K, m, H\}$ of VD model parameters, and returns as output a $k \times n$ matrix, $\mathbf{Y}$ with the simulated abundances of the two predators and two preys, respectively. Bifurcation analyses of the simulated abundances, $\mathbf{Y}$, displays different dynamics, including stable equilibria, limit cycles and chaos depending on the values of the VD parameters (Figures S1-S3). In particular, chaotic dynamics is quite ubiquitous in this model, resulting in complex patterns of synchronous and asynchronous fluctuations of the four species. We refer interested readers to the publications of Vandermeer (1982, 2004) for a detailed explanation of the VD model, and Benincà et al. (2009) for analysis of coupled predator-prey cycles in our food web.

We must choose the VD parameter values so that the simulated abundances, $\mathbf{Y}$, match as closely and consistently as possible the observed data, $\widetilde{\mathbf{Y}}$. This necessitates inference of $\boldsymbol{\theta}$ to minimize the overall "length" of the $k \times n$ matrix of residuals, $\mathbf{E}(\boldsymbol{\theta})$

$$\mathbf{E}(\boldsymbol{\theta}) = \widetilde{\mathbf{Y}} - \mathbf{Y}(\boldsymbol{\theta}) = \{\mathbf{e}_1(\boldsymbol{\theta}), \ldots, \mathbf{e}_k(\boldsymbol{\theta})\}, \tag{2.3}$$

where the $1 \times n$ vector $\mathbf{e}_j$ stores the residuals of the jth species, $j = \{1, \ldots, k\}$.

In this study we examine three different parameterizations. The first originates from wavelet analysis by Benincà et al. (2009) and is listed in Table 1 under "WAVE". The other two parameter

vectors, coined "SODA" and "INTEL", are derived from the measured abundances using inference methods outlined in the next sections.

2.3.3 Bayesian inference of the VD model parameters

In recent decades, Bayesian inference has emerged as a working paradigm for modern probability theory, parameter and state estimation, model selection and hypothesis testing (Vrugt, 2016). Bayesian inference allows for an exact description of parameter uncertainty (and other sources of uncertainty) by treating the parameters as probabilistic variables with joint posterior pdf, $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. This multivariate distribution, the so-called posterior parameter distribution, is the consequence of two antecedents, a prior distribution, $p(\boldsymbol{\theta})$, which captures our initial degree of beliefs in the values of the model parameters, and a likelihood function, $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, which quantifies the level of confidence in the parameter values, $\boldsymbol{\theta}$, in light of the observed data, $\widetilde{\mathbf{Y}}$. Bayes' theorem can be derived from basic axioms of probability

$$p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) = \frac{p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})} \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}), \tag{2.4}$$

where the denominator, $p(\widetilde{\mathbf{Y}}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, acts as a normalizing constant so that $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ integrates to unity.

Without loss of generality, we restrict the model parameters to a closed space, Θ, equivalent to a d-dimensional hypercube, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \in \mathbb{R}^d$, called the feasible parameter space. In the absence of detailed knowledge about the parameter values of our food web, we assume a uniform prior distribution, $p(\boldsymbol{\theta})$, over the ranges listed in Table 2.1. This leaves us with specification of $L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. In the ideal case with a perfect model and initial states, we expect the distribution of the residuals of Equation (2.3) to match exactly the distribution of the measurement errors of the species' abundances, $\widetilde{\mathbf{Y}}$. If we assume the n-measurement errors, $\mathbf{v}_j = \{v_{j1}, \ldots, v_{jn}\}$, of each species, $j = \{1, \ldots, k\}$, to be independent and zero-mean normally distributed, $\mathcal{N}(0, \widehat{\sigma}^2_{\mathbf{v}})$, with constant

variance, $\widehat{\sigma}^2_{\mathbf{v}}$, then the likelihood function becomes

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}, \widehat{\sigma}^2_{\mathbf{v}}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{\mathbf{v}}}} \exp\left[-\frac{1}{2}\widehat{\sigma}^{-2}_{\mathbf{v}} \sum_{j=1}^{k}\sum_{t=1}^{n} (\widetilde{y}_{jt} - y_{jt}(\boldsymbol{\theta}))^2\right], \tag{2.5}$$

where $e_{jt}(\boldsymbol{\theta}) = \widetilde{y}_{jt} - y_{jt}(\boldsymbol{\theta})$ signifies the residual of the jth species' population count at time t. The summation term inside the exponent is equivalent to the sum of squared residuals used widely in model fitting.

The vector of VD parameter values that maximizes the likelihood function is also referred to as the maximum likelihood (ML) solution. With a uniform prior, this ML solution is equivalent to the peak solution of the posterior distribution, $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. We "integrate out" the measurement error variance, $\widehat{\sigma}^2_{\mathbf{v}}$, using as proxy for $\widehat{\sigma}^2_{\mathbf{v}}$ the variance of the ML residuals. This gives (Vrugt, 2016)

$$L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto \left(\sum_{j=1}^{k}\sum_{t=1}^{n} (\widetilde{y}_{jt} - y_{jt}(\boldsymbol{\theta}))^2\right)^{-\frac{1}{2}nk}. \tag{2.6}$$

We are now left with inference of the (unnormalized) posterior distribution, $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}}) \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$, which summarizes our updated knowledge (belief) on the parameters. Unfortunately, in most applications $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$ does not have a closed-form analytic solution. We therefore resort to Markov Chain Monte Carlo (MCMC) simulation and use the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016) to generate samples of the posterior parameter distribution. A detailed description of DREAM appears in Vrugt (2016) and interested readers are referred to this publication for further details.

2.3.4 Joint parameter and state estimation

The assumption of exact knowledge of the initial states of our food web may be appropriate for our mesocosm data set, but the assumption of a perfect model cannot be justified. Indeed, the VD model is a highly simplified description of the population dynamics in our food web, and consequently, it

may not be able to mimic accurately the observed abundances. One should therefore not expect the residuals to satisfy assumptions of normality and independence, but instead exhibit considerable variation in bias, variance, and serial correlation under different population counts. We therefore consider the SODA method (Vrugt et al., 2005) which relaxes the assumption of a perfect model and accounts implicitly for structural errors of the VD model during parameter estimation.

Lets assume, for the time being, that the VD parameter values are known. We write the VD model in a state-space formulation

$$\mathbf{x}_t^{\mathrm{f}} = \mathcal{VD}(\boldsymbol{\theta}, \mathbf{x}_{t-\Delta t}) + \mathbf{q}_t, \tag{2.7}$$

where $\mathbf{x}_t^{\mathrm{f}}$ is a 4×1 vector with forecasted abundances of the two predator and two prey species, respectively, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_d\}$ is the $d = 8$-vector of parameters, and $\mathbf{q}_t$ is a 4×1 process noise vector that accounts for model structural inadequacies. The time step, Δt, is equivalent to the interval of 3.35 days between measured abundances. Thus, based on the state, $\mathbf{x}_{t-\Delta t}$, at time $t - \Delta t$ and the values of the parameters, $\boldsymbol{\theta}$, the VD model predicts the one-observation-ahead abundances using numerical integration of the coupled two-predator-two-prey system in equation (2.1) between $t - \Delta t$ and t days. These VD-predicted abundances are then corrupted with the stochastic noise term, $\mathbf{q}_t$, to create the state forecast, $\mathbf{x}_t^{\mathrm{f}}$, at time t. We can now calculate the likelihood of the forecasted state via recursive implementation of Bayes' theorem in Equation (2.4), details which are found elsewhere.

If the model operator in Equation (5.1) is linear, and the distributions of the model error, $\mathbf{q}_t$, and measurement error, $\mathbf{v}_t$ are multivariate Gaussian, respectively, then the Kalman filter (KF) (Kalman, 1960) provides an optimal estimate of the population counts in our food web at each measurement time t as follows

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t(\widetilde{\mathbf{y}}_t - \mathbf{x}_t^{\mathrm{f}}), \tag{2.8}$$

where $\mathbf{x}_t^{\mathrm{a}}$ is a 4×1 vector with the analysis state, and $\mathbf{K}_t \in [0, 1]$ is a 4×4 matrix of weights called the Kalman gain. The optimal, or analysis, state is our best estimate of the abundances in the food web and lies between the state forecast, $\mathbf{x}_t^{\mathrm{f}}$, and state measurement, $\widetilde{\mathbf{y}}_t$. The second term on the right-hand-side is the innovation vector, $\mathbf{i}_t(\boldsymbol{\theta})$.

The Kalman gain is the relative weight given to the state forecast and state measurement, respectively, and is calculated using

$$\mathbf{K}_t = \frac{\mathbf{Q}_t}{\mathbf{Q}_t + \mathbf{R}} = \mathbf{Q}_t(\mathbf{Q}_t + \mathbf{R})^{-1} \tag{2.9}$$

where $\mathbf{Q}_t$ and $\mathbf{R}$ signify the 4×4 covariance matrices of the model and measurement errors, respectively. In general, if the model errors are large compared to the measurement errors, the gain will be high and the analysis state, $\mathbf{x}_t^{\mathrm{a}}$, of the KF will track closely the measured abundances, $\widetilde{\mathbf{y}}_t$. On the contrary, if the model errors are small relative to the measurement errors, the gain will be small and the analysis state close to $\mathbf{x}_t^{\mathrm{f}}$. The analysis state now enters the VD model to predict the abundances at $t + \Delta t$, and this two-step process of state prediction and state analysis is repeated for every next observation. This recursion negates forecast bias, and helps close the gap between the observed and simulated abundances.

The original Kalman filter requires an analytic formula for the evolution of the model error covariance matrix, $\mathbf{Q}_t$. This formula can only be derived for linear state-space models. For nonlinear models, the extended Kalman filter can be used, which linearizes $\mathbf{Q}_t$ via a tangent linear operator. This approximation, however is notoriously unstable for the VD model as governed by highly nonlinear and/or choatic state transitions. We therefore resort to the Ensemble Kalman filter (EnKF) of Evensen (1994) instead, which approximates the matrix $\mathbf{Q}_t$ by the sample covariance of N different realizations, called an ensemble, of the state forecast. The EnKF is relatively easy to implement. Figure 2.1 presents a brief summary of the method. Interested readers are referred to Evensen (2009) for a detailed explanation of the EnKF. SI-B explicates further our choice for the EnKF rather than the more generic particle filter.

The performance of the KF (or EnKF) depends in large part on the choice of $\mathbf{R}$ and $\mathbf{Q}_t$. We make the common assumption that the state measurement errors are independent, zero-mean normally distributed with variance, $\widehat{\sigma}_{\mathbf{v}}^2$. Consequently, $\mathbf{R} = \widehat{\sigma}_{\mathbf{v}}^2 \mathbf{U}_4$, where $\mathbf{I}_4$ is the 4×4 unit (or identity) matrix. Based on recommendations in Evensen (1994) we use temporally correlated process noise, $\mathbf{q}_t$, in Equation (5.1)

$$\mathbf{q}_t = \rho \mathbf{q}_{t-\Delta t} + \mathbf{w}_t, \tag{2.10}$$

where $\rho \in (-1, 1)$ (-) is the first-order autocorrelation coefficient of the model errors, and $\mathbf{w}_t$ is a 4×1 vector with values drawn randomly from a zero-mean Gaussian distribution with covariance matrix, $\widehat{\sigma}^2_{\mathbf{w}}\mathbf{U}_4$.

Thus far, we have assumed the VD parameter values to be known during state estimation, a supposition that is not particularly realistic for the present application. We therefore use the SODA method and estimate jointly the states and parameters. This method uses an inner EnKF loop for recursive state estimation conditioned on an assumed parameter set, and an outer MCMC simulation loop for batch estimation of the posterior VD parameter distribution, $p(\boldsymbol{\theta}|\widetilde{\mathbf{Y}})$. We assume, again, a uniform prior parameter distribution, $p(\boldsymbol{\theta})$ and compute the likelihood, $L(\boldsymbol{\theta}|\mathbf{Z})$, of the $k \times n$ matrix of state forecast errors, $\mathbf{Z}$, using Equation (2.6) (but with $\mathbf{Z}$ rather than residuals without DA). The ML parameter values derived from SODA will minimize the one-observation-ahead forecast errors of the VD model. Note, that if we set the model error covariance matrix $\mathbf{Q}$ to zero then the Kalman gain, $\mathbf{K}$, will always contain zeros and SODA simplifies to parameter estimation only via Bayes' theorem in Equation (2.4).

We assume an ensemble of $N = 100$ members, $\widehat{\sigma}^2_{\mathbf{v}} = 0.05$, and treat ρ and $\widehat{\sigma}^2_{\mathbf{w}}$ as nuisance variables whose values are estimated along with the VD parameters using the measured abundances, $\widetilde{\mathbf{Y}}$, of the predators and preys. The ML values of "SODA" are listed in Table 2.1.

2.3.5 The "Intelligent" (INTEL) model

The estimated parameter values derived from SODA assume the continued application of DA to the entire 2656 day record. This equates to an average assimilation interval of $\Delta t = 3.35$ days. As data collection may be rather expensive and wasteful, we raise the question here of how frequently must we assimilate the measured abundances to describe accurately population dynamics in our food web? For this, we investigate the relationship between the time interval of two successive population counts, and the forecast skill of the VD model. We also investigate whether we can lower the intensity of data collection, and make intelligent choices regarding measurement timing.

We construct a second and perhaps more "intelligent" model, coined INTEL, which restricts state updating to those 20% of the observations with largest fluctuations in the abundances. In between these "important" observations, the VD model is executed in open loop without state adjustments. We estimate the VD parameters and nuisance variables ρ and $\widehat{\sigma}^2_{\mathbf{w}}$ of INTEL with SODA but assimilating only the 20% key abundance measurements. Table 1 list the ML values of INTEL.

2.3.6 Reduced measurement frequencies

To assess the effect of measurement interval on the predicted population dynamics by the VD model, we evaluate the WAVE, SODA and INTEL parameters with state updates at every $\Delta t = 3.35$, $2\Delta t = 6.70$, $5\Delta t = 5.07$ and $10\Delta t = 33.50$ days, thereby assimilating 100, 50, 20, and 10% of the measured abundances of the predators and preys, respectively. This analysis will provide insights into the relationship between the time interval of successive abundance measurements and the forecast skill of the VD model. These findings provide guidance on experimental design in support of efficient and effective data collection.

2.3.7 Ecological Forecast Horizons

DA should enhance drastically the VD model's ability to track our food web's population dynamics. Yet, state estimation can only be used "in sample" requiring knowledge of the observed abundances. To better understand the "out of sample" performance of the VD model, we use open loop simulation and evaluate the ecological forecast horizon (EFH) of the WAVE, SODA and INTEL parameterizations. Petchey et al. (2015) defined the EFH (P. 597) as *"... the dimensional distance for which useful forecasts can be made"*. Benincà et al. (2008) showed the population dynamics in our food web to have a limited predictability of 15 - 30 days due to chaotic state transitions. We extend the analysis of Benincà et al. (2008) and compute EFHs for the individual predator and prey species using the three different VD parameterizations.

To compute the EFH we need to define a criterion that measures forecast proficiency, and

another, related, criterion that classifies a forecast as being "good" or "bad". This second criterion, coined the forecast proficiency threshold, or FPT, by Petchey et al. (2015), defines the horizon for which model predictions are deemed good enough, and below which the forecasts are considered unacceptable. The forecast horizon is now defined as the first time from the start of simulation at which the forecast proficiency drops below the FPT. In our application, we set the forecast proficiency equivalent to the distance (absolute residual) between the observed and predicted abundances. Without formal guidelines on the choice of the FPT, we use FPT $= \{0.1, 0.2, 0.3, 0.4, 0.5\}$, and report the corresponding forecast horizon for each species and VD parameterization. This provides insights into the relationship between the choice of the FPT and the EFH.

One could argue in favor of an independent evaluation data set to help assess the performance and forecast horizons of the VD model. The use of an evaluation data set should, however not change much our findings. Main reason is the inability of the VD model to explain, without state estimation, the measured abundances. The use of many different starting points of the VD model within the measured record, $\widetilde{\mathbf{Y}}$, should therefore provide robust EFHs of the predators and preys (see SI-C) for a detailed recipe).

We calculate separately the EFH for each predator and prey species and evaluate the forecast errors of the VD model with the INTEL, WAVE, and SODA parameterizations using many different starting points within the calibration data record of observed abundances (see recipe below).

[1] Choose the VD model parameter vector (WAVE, SODA or INTEL)

[2] Define the forecast proficiency threshold (FPT $= \{0.1, 0.2, 0.3, 0.4, 0.5\}$)

[3] For $j = \{1, \ldots, 4\}$ species

  [A] Repeat M times

    [1] Draw randomly an integer, $T \in \{1, \ldots, 601\}$

    [2] Initialize the VD model at $t_{\mathrm{s}} = (T-1)\Delta t$ (days)

    [3] Simulate abundances to $t_{\mathrm{f}} = 2656$ days using a print step of Δt days

    [4] Compute the residual vector of the jth species

    [5] At time, t_{FH}, (days) the absolute residual exceeds, for the first time, FPT

[6] Define the forecast horizon, FH $= t_{\rm FH} - t_{\rm s}$

[B] End (repeat M times)

[C] The EFH of the jth species is equal to the mean of its M forecast horizons [4] End (for loop)

The recipe above computes the forecast horizon of the VD model via open loop simulation within the calibration data set. One might argue that this approach violates the basic principles of the EFH, and that we should have used split sampling to evaluate the performance and forecast horizons of the VD model for an independent data set. Yet, we do not expect out-of-sample VD model evaluation to dramatically change our findings. The main reason for this is the inability of the VD model to track closely, without state estimation, the measured data. The use of M different starting points of the VD model within the calibration data record, $\widetilde{\mathbf{Y}}$, should thus suffice to provide robust forecast horizons of each species.

2.4 Results

To simplify discussion and graphical interpretation, we use color coding in green, blue, and black to differentiate between the results of the WAVE, SODA and INTEL parameter vectors of Table 1, respectively.

2.4.1 Parameter estimates and model outputs

Table 1 compares the WAVE values of the VD model parameters derived from wavelet analysis Benincà et al. (2009) with their ML counterparts of SODA and INTEL. The values in parenthesis report the posterior standard deviations of the VD parameters, and the nuisance variables ρ and $\widehat{\sigma}^2_{\mathbf{w}}$. Figure S4 presents box plots of the marginal posterior distributions of the SODA and INTEL parameters. To simplify plotting each box plot is normalized with the prior range of each parameter. The nuisance variables ρ and $\widehat{\sigma}^2_{\mathbf{w}}$ are required for state estimation by SODA and INTEL. Our DA experiments with the WAVE parameters uses SODA's ML values of ρ and $\widehat{\sigma}^2_{\mathbf{w}}$.

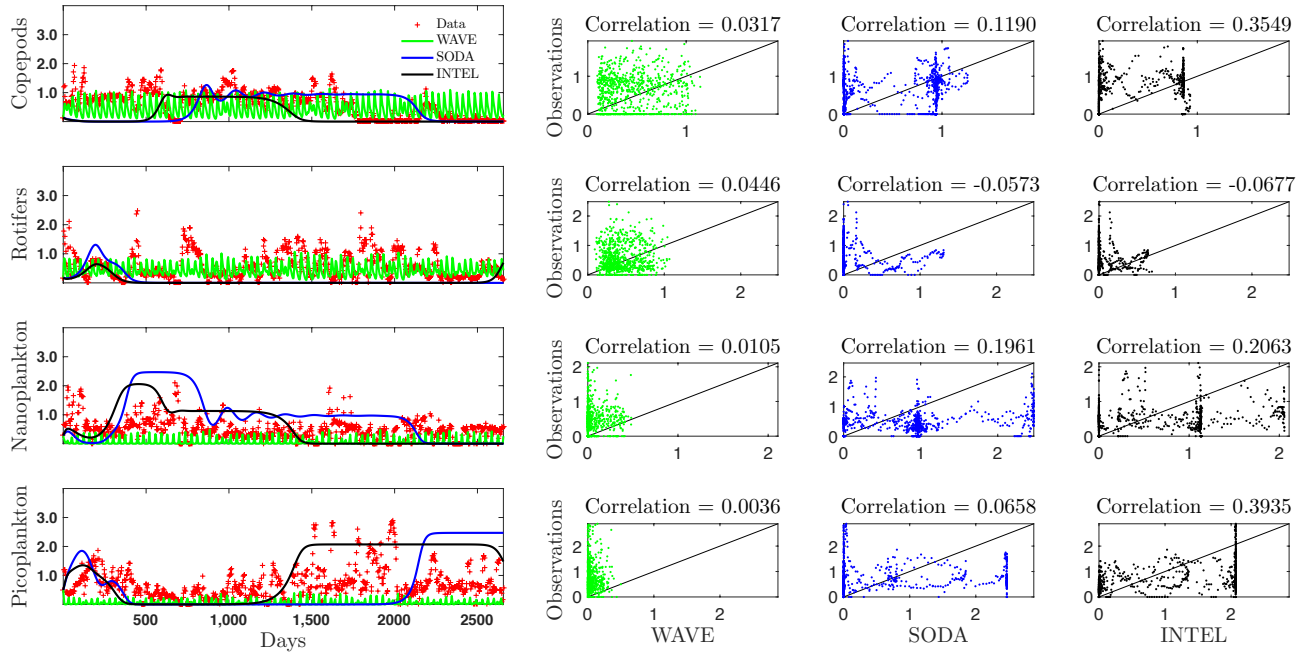

Figure 2.2: Open loop simulation of the VD model for $t \in [0, 2656]$ days using the WAVE, SODA and INTEL parameter values: Time series plots of measured and simulated abundances of the predator and prey species (left panels), and scatter plots of simulated and observed abundances (right panels). The three parameterizations (solid lines) poorly describe the highly nonlinear and chaotic population dynamics of the two predator and two prey species in our microbial food web. Therefore, we investigate the use data assimilation to reduce simulation bias of the VD model, and close the gap between the observed and simulated species abundances.

The SODA and INTEL values of the VD parameters are in good agreement (Figure 9), and correspond reasonably well with their WAVE values. All three parameterizations use values for $\alpha \in [0.1 - 0.3]$, $\beta \in [1.36 - 1.75]$ and $m \in [0.05 - 0.08]$, respectively. This is an encouraging result, as β and α control in large part the population dynamics simulated by the VD model. Note, that WAVE assign much higher values to the growth rates, r_1 and r_2, of the first and second prey, respectively, and the grazing rate, g. It is difficult to pinpoint the exact reasons for this mismatch, yet, these findings suggest that the VD model is structurally deficient, particularly in describing the growth of the preys and the grazing behavior of the zooplankton organisms.

The DA method returns as byproduct estimates of the serial correlation, ρ, and variance, $\sigma^2_{\mathbf{w}}$, of the VD model error. As expected, the model error exhibits some positive (SODA) or negative (INTEL) serial correlation, but this temporal dependency is not very strong. This suggests that

the model bias does not persist far into the future, but rather the sign of the forecast errors changes rapidly from one observation to the next. This is typical for models that are unable to track closely the observed data. The variance of the model error, $\sigma^2_{\mathbf{w}}$, appears to be 4 (INTEL) to 7 (SODA) times larger than the measurement error variance, $\sigma^2_{\mathbf{v}} = 0.05$. This is a very reasonable finding, and leads, on average, to a Kalman gain of about 0.8 or 0.9. In other words, we trust the measured abundances about 5 times more than their simulated values of the VD model.

Figure 7 displays the abundances of the WAVE, SODA and INTEL parameters using open loop simulation with the VD model. It is evident that the VD model cannot track the observed abundances. If we assimilate the measured abundances, then the performance of the VD model enhances dramatically (Figure 8). This is especially true for the SODA and INTEL parameterizations. This finding is not surprising, as the SODA and INTEL parameters were derived from the SODA method using minimization of the one-observation-ahead forecast error. The superiority of the INTEL and SODA parameters is further confirmed in Table 1 which lists the Root Mean Square Error (RMSE) of their forecast errors with and without state estimation. Separately, we also report the mean absolute innovation of each VD model parameterization. As expected, the WAVE parameters exhibit substantially larger forecast errors (on average), and, consequently, require larger innovations to negate prediction bias.

2.4.2 Effect of data assimilation frequency

We evaluate the impact of state estimation frequency on the performance of the VD model, using the WAVE, SODA and INTEL parameters. Panel A of Figure 10 compares in a bar plot the RMSE of the forecast error of each VD parameterization. DA only slightly improves the forecast accuracy of the WAVE parameters, regardless of state estimation frequency. The forecast proficiency of the SODA and INTEL parameters, however steadily improves with increasing number of assimilated abundance measurements. Panel B of Figure 10 displays a bar plot of the mean absolute innovation of each parameterization. In general, the SODA and INTEL parameters have smaller innovations as they closer track the observed abundances. Detailed analysis of the state innovations can help pinpoint model structural errors (Vrugt et al., 2005). Altogether, INTEL is the most efficient

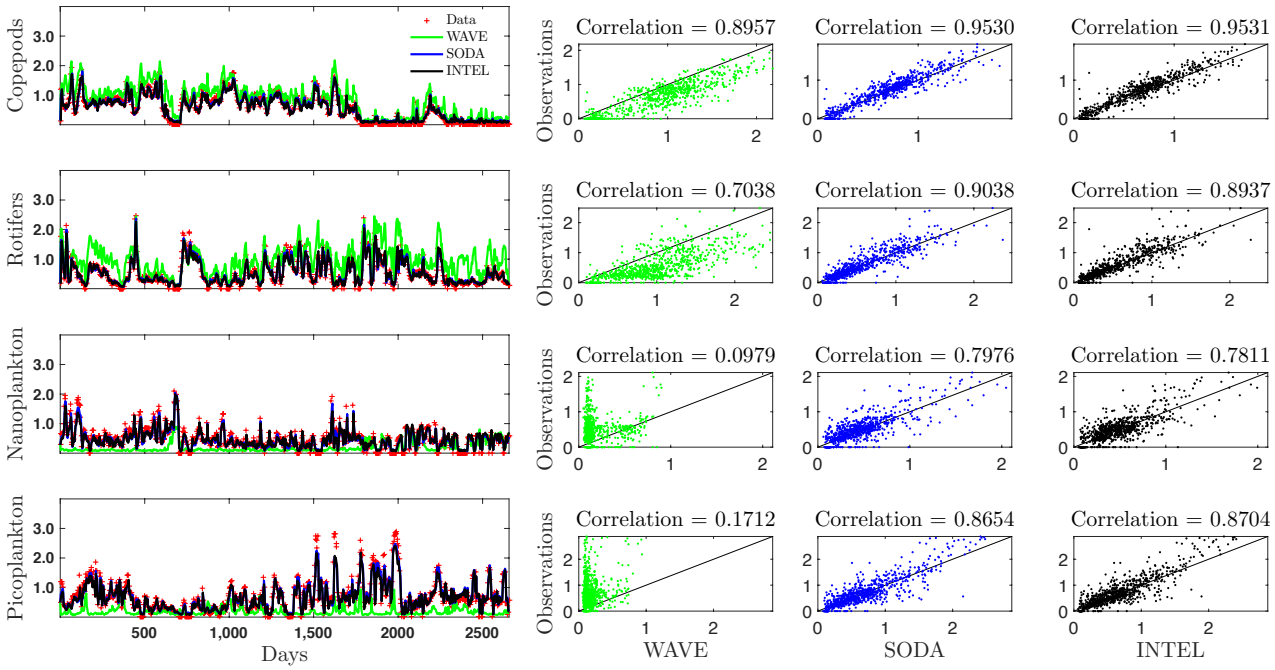

Figure 2.3: Forecasted abundances of the VD model during our 8-year experiment with the WAVE, SODA and INTEL parameters using state estimation with the EnKF. Time series plots of measured and forecasted abundances of the predator and prey species (left panels), and scatter plots of simulated and observed abundances (right panels). Data assimilation enhances significantly the ability of the VD model to track the population dynamics of the predators and preys. This is particularly true for the SODA and INTEL parameterizations, which follow most closely the observed population counts.

use of DA with relatively low forecast errors if state estimation is applied to every fifth forecast. Otherwise, the VD model runs in open loop, and residuals increase rapidly.

2.4.3 Model predictability: forecast horizons

Figure 11 presents the EFH of each species for different values of the FPT. When open loop simulation is used with the VD model, then the SODA and INTEL parameters exhibit, on average, the smallest forecast errors. In other words, for a given proficiency threshold, the SODA and INTEL parameters can predict further into the future than their WAVE estimates. This translates into a forecast horizon of SODA/INTEL that is just a few days larger than WAVE if FPT = 0.1 and grows rapidly to 20 - 40 days for FPT = 0.5. What is more, the zooplankton species exhibit a much larger forecast horizon with SODA/INTEL than their phytoplankton counterparts. This

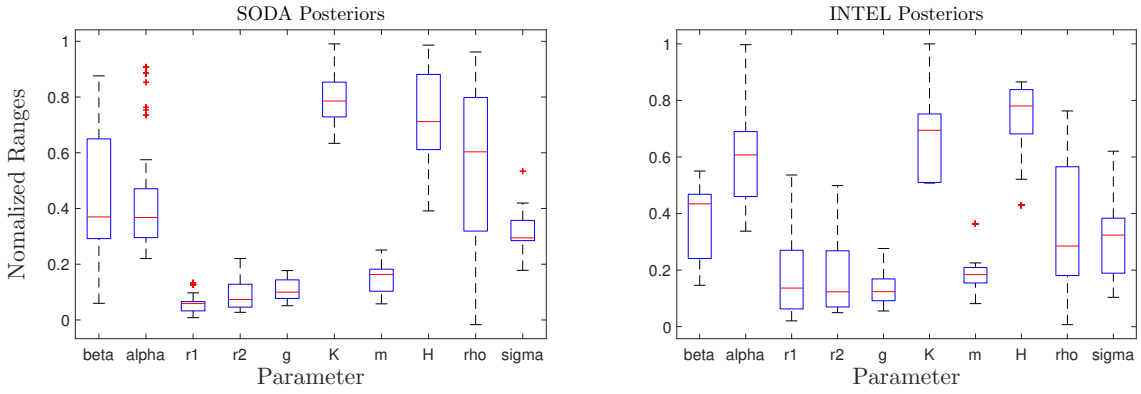

Figure 2.4: Box plots of the marginal posterior distributions of the (A) SODA and (B) INTEL parameters. The box plot of each VD model parameter is normalized by its prior ranges (listed between parentheses in Table 1). The larger the spread of a box plot the larger the posterior parameter uncertainty.

difference amounts to a few days if FPT = 0.1 and increases to about 20 days if FPT = 0.5. Note also that the forecast horizons increase nonlinearly with proficiency threshold.

2.5 Discussion

Species abundances in ecological communities can display complex non-equilibrium dynamics (Becks et al., 2005; Benincà et al., 2015, 2008; Hanski et al., 1993; May, 1973). A characteristic feature of chaotic systems, is that long-term prediction of the system's trajectory is fundamentally impossible (Strogatz et al., 1995). How then should we make predictions for complex multi-species communities?

We have used DA with the EnKF to fuse the simulated abundances of the two predator - two prey VD model with observed population counts. The state adjustments of the EnKF negate, at least in part, structural errors of the VD model, and allow system dynamics to be better described. For optimal predictions, we estimated the VD parameters in an loop outside the EnKF, otherwise the model may not have maximum predictive ability as shown with the WAVE parameters. This led to the SODA and INTEL parameterizations which minimize the forecast errors of the VD model using either all or only the 20% most important abundance observations for state estimation. We then examined the forecast errors of the WAVE, SODA and INTEL parameters using different assimilation frequencies of the abundance measurements.

Symbol	Description	Lower	Upper	Units	WAVE	SODA	INTEL
β	predator coefficient	10^{-4}	1.0	-	0.1	0.12 (0.26)	0.29 (0.14)
α	prey coefficient	0.0	2.0	-	1.5	1.75 (0.21)	1.36 (0.16)
r_1	growth rate of first prey	0.01	2.5	day^{-1}	0.66	0.08 (0.03)	0.13 (0.16)
r_2	growth rate of second prey	0.01	2.5	day^{-1}	0.66	0.09 (0.04)	0.12 (0.14)
g	grazing rate	0.1	2.5	day^{-1}	1.0	0.11 (0.04)	0.24 (0.06)
K	carrying capacity	0.5	2.5	$(\text{mg fwt L}^{-1})^{1/4}$	1.0	2.47 (0.09)	2.07 (0.12)
m	mortality rate	0.01	0.7	day^{-1}	0.066	0.05 (0.06)	0.08 (0.08)
H	parameter of functional response	1.0	3.0	$(\text{mg fwt L}^{-1})^{1/4}$	0.8	1.16 (0.20)	2.39 (0.13)
ρ	first-order autocorrelation of model errors	-1.0	1.0	-	N/A	0.69 (0.26)	-0.35 (0.26)
$\widehat{\sigma}^2_{\mathbf{w}}$	variance of model error	0.0	1.0	-	N/A	0.35 (0.07)	0.20 (0.16)
	RMSE of VD model (no DA: open loop)			$(\text{mg fwt L}^{-1})^{1/4}$	0.63	0.84	0.82
	RMSE of VD model (100% DA)			$(\text{mg fwt L}^{-1})^{1/4}$	0.56	0.21	0.22
	Ensemble mean innovation (100% DA)			$(\text{mg fwt L}^{-1})^{1/4}$	0.38	0.14	0.14

Table 2.1: Description of the VD model parameters, including their lower and upper values, and units. The columns WAVE, SODA and INTEL lists the (optimized) parameter values derived from wavelet analysis, and joint parameter and state estimation with SODA using 100% and 20% of the measured abundances for state updating. The bottom part of the table reports the RMSE of the different VD model parameterizations with and without DA, and the ensemble mean innovation.

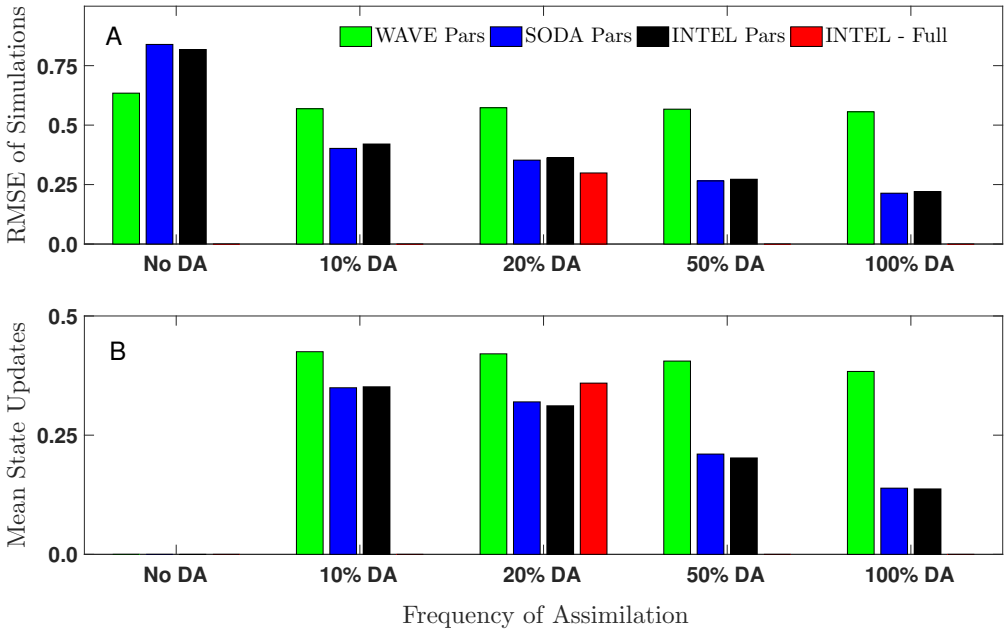

Figure 2.5: A) Root Mean Square Error, or RMSE, of the forecasted abundances of the VD model using the WAVE, SODA and INTEL parameters with different assimilation frequencies, and (B) ensemble mean value of the state innovation. The red bars signify the results of the INTEL model with state estimation restricted to 20% of the time steps with largest fluctuations in the observed data.

DA not only allows ecologists to fuse their models with data, and close the gap between ecosystem theory and observations, but also helps provide guidance on measurement design and frequency and timing of observations. To this end, we introduced an "intelligent" model, called INTEL which restricts state estimation to only the 20% most important measurement times with largest changes in measured abundances. This approach showed to be the most efficient use of DA.

We quantified the forecast proficiency of the VD model using different proficiency thresholds. These findings demonstrated that DA significantly enhances the forecast horizon of each species. The SODA and INTEL parameters had significantly larger forecast horizons than their WAVE counterparts. Nevertheless, we cannot confidently claim why long-term population dynamics in our food web is unpredictable. One could envisage many different reasons. We argue that the answer is twofold. On one hand, sudden, chaotic state transitions, make it fundamentally impossible to predict system dynamics far into the future. On the other hand, structural errors of the VD model limit its forecast horizon even during periods without large, unexpected, changes in population

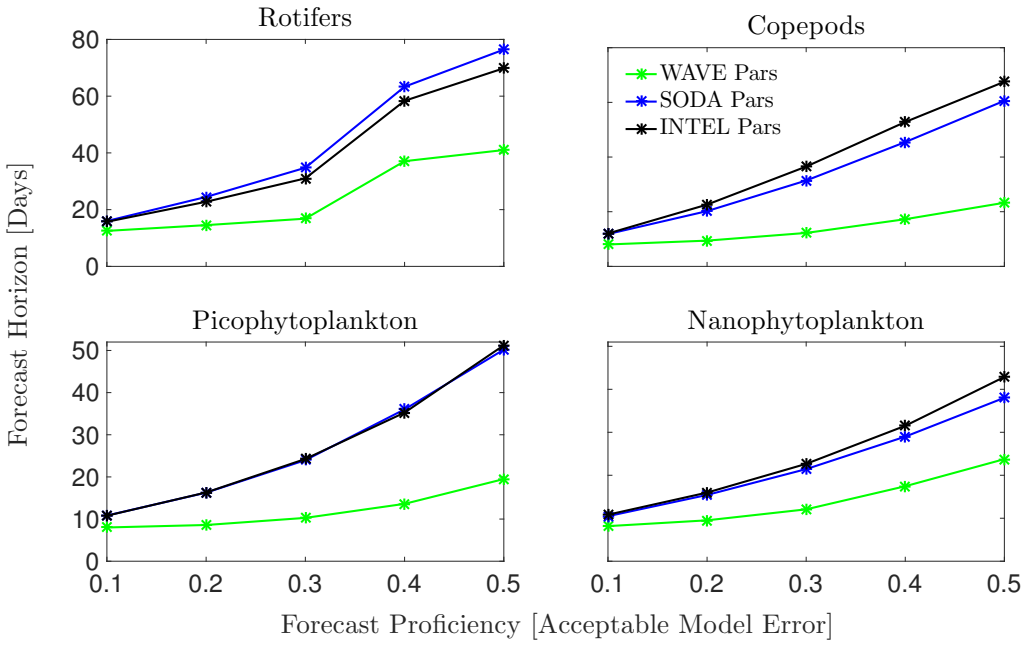

Figure 2.6: Ecological forecast horizon of the VD model derived from open loop simulation with the WAVE, SODA and INTEL parameter values using many different initial states drawn from the mesocosm data set. The forecast horizon is defined as the first time from the start of simulation at which the distance between the state forecast and corresponding observation exceeds a threshold value, called FPT. The forecast horizons quantify the ability of the VD model to produce "good" forecasts, without data assimilation.

counts.

The SODA or INTEL parameters were shown to exhibit smaller prediction errors and larger forecast horizons than their WAVE estimates. Yet, this does not imply that their parameter values more accurately portray system properties. Without state estimation, for example, the SODA and INTEL parameters do not reproduce the species' population oscillations, a feature that the WAVE parameterization is able to replicate almost exactly. The assumption that the two zooplankton species share similar values of g, H and m may be too simplifying. Indeed, rotifers and copepods are two different organisms, with a dissimilar grazing rate, mortality, and prey treatment. A more inclusive parameterization may enhance the ability of the VD model to describe the coupled predator-prey oscillations.

Short-term model predictions may be accurate for chaotic systems, but long-term prediction impossible. This is precisely the reason that DA is an attractive approach, as it will help close the gap between ecological observations and theory, and increase the forecast horizon of ecosystem

models. Our results can serve as a benchmark for DA methods in ecology, and aid measurement design and data collection. We look forward to further advances in ecological forecasting, especially related to systems with chaotic dynamics.

Supplementary information

A: Bifurcation analysis of the VD model

To inspire confidence in the robustness of the (maximum likelihood) values of the WAVE, SODA and INTEL parameters, we performed a bifurcation analysis of the VD model using the prey coefficient, α, predator coefficient, β, and carrying capacity, K, as our bifurcation parameters. All other VD model parameters were kept fixed at their WAVE values listed in Table 1 of the main text. Figures S1 - S3 present the outcome of this analysis. The three bifurcation plots illustrate, how depending on values of α, β and K, the system exhibits different dynamical behaviors, including stable equilibria, limit cycles and chaos. Note, that the SODA and INTEL values of the VD model parameters (see Table 1) occupy the region of chaotic behavior.

The dynamical behavior of the VD model is quite complicated with species that show intriguing mechanisms of (a-)synchronization and competitive exclusion depending on parameter selection. A detailed bifurcation analysis with a larger range of values for the parameters α and β has appeared in Vandermeer (2004), and interested readers are referred to this publication for further information. Briefly, this analysis showed that, effectively, there are two distinct bifurcation patterns: (1) eventual elimination of one of the predators through predator competition, and (2) eventual elimination of one of the predators because of increased resource coupling (potentially representing indirect mutualism between the two consumers). Both bifurcation patterns were observed for many different values of α and β.

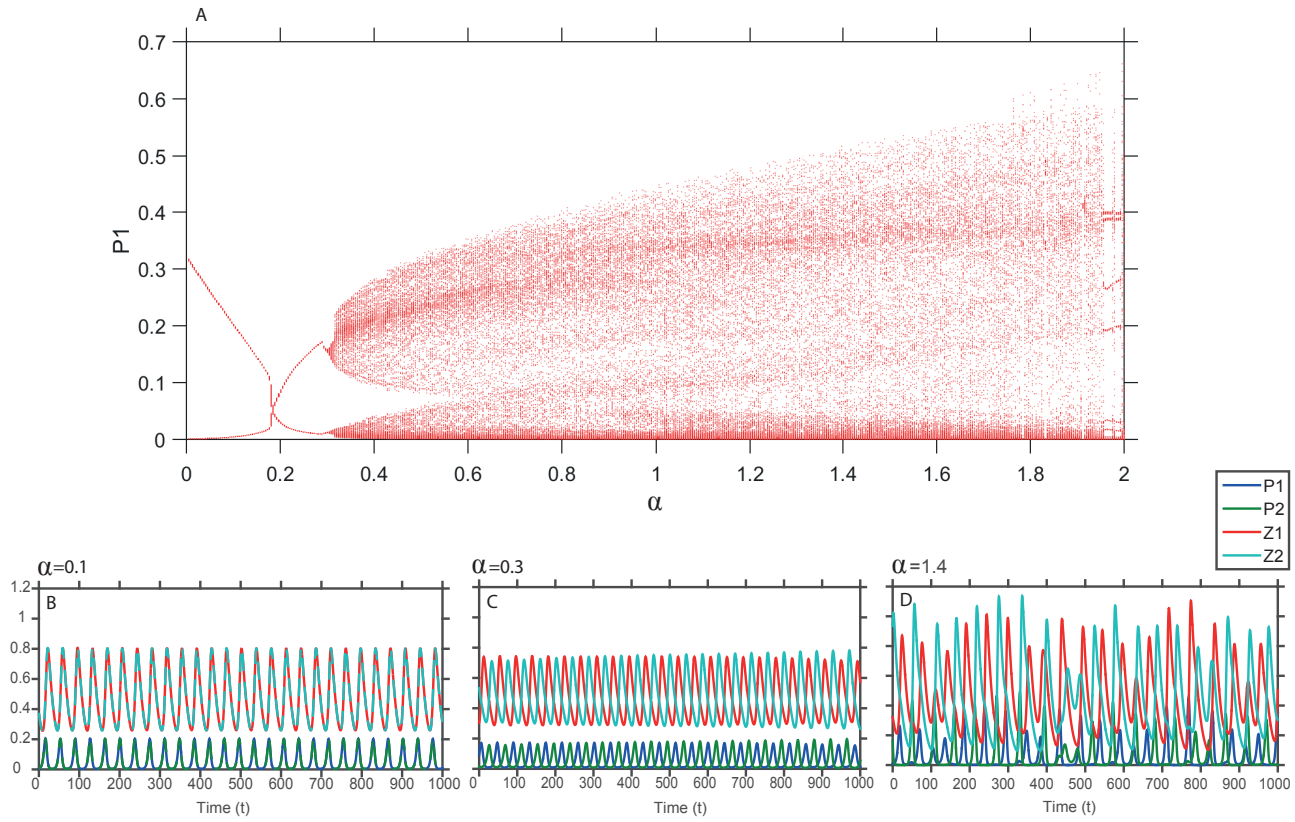

Figure 2.7: Bifurcation analysis of the VD model simulated abundances of the two predator, P_1 and P_2, and two prey, Z_1 and Z_2, species for $t \in [X, X]$ and initial state, $\mathbf{x}_0 = \{P_1, P_2, Z_1, Z_2\} = \{0.28, 0.50, 0.14, 0.18\}$ as a function of the prey coefficient α: (A) Bifurcation diagram that displays the local minima and maxima in the fluctuations of the phytoplankton abundances, P_1, at different values of α; Time series plot of simulated abundances for (B) $\alpha = 0.1$, (C) $\alpha = 0.3$, and (D) $\alpha = 1.4$. All other VD model parameters were kept constant at their WAVE values listed in Table 1. The bifurcation plot was generated with the GRIND Matlab package (`http://www.sparcscenter.org/grind`).

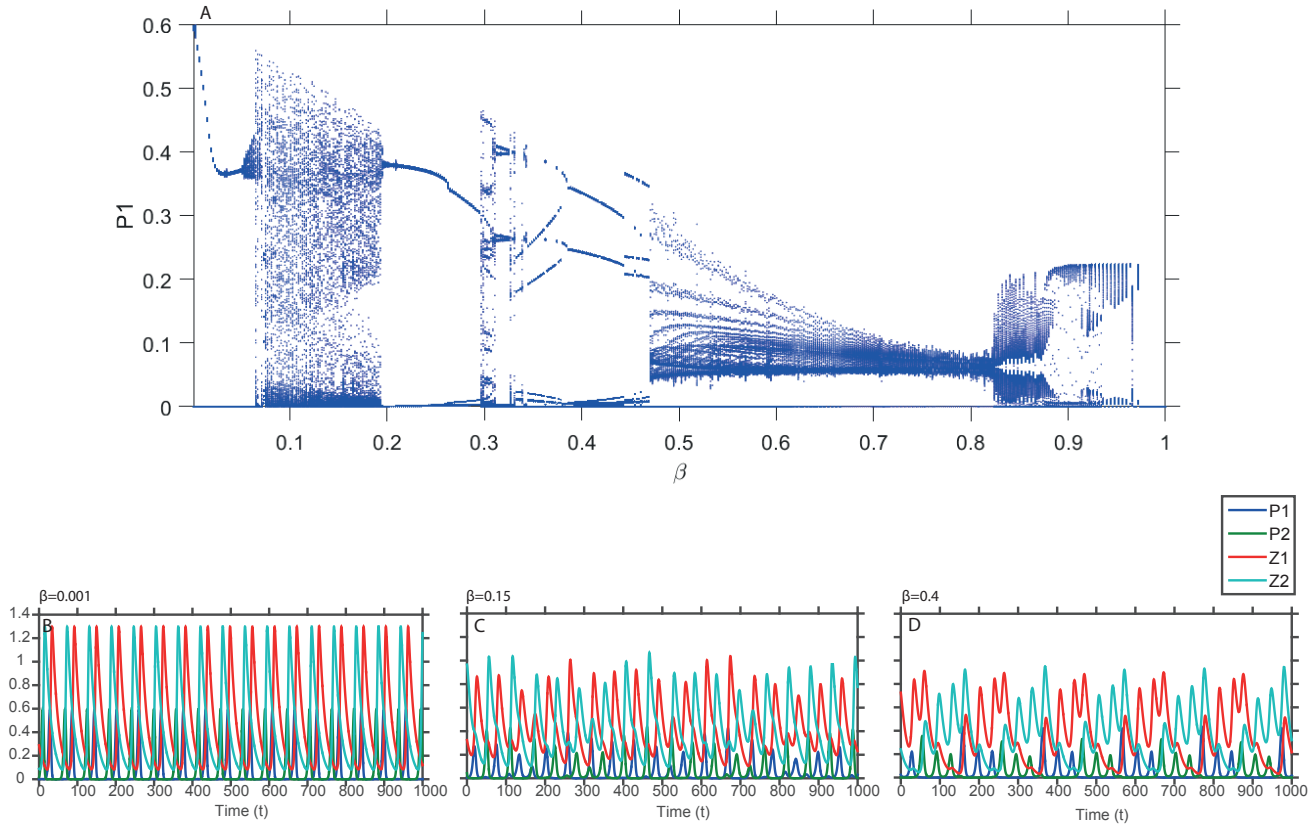

Figure 2.8: Bifurcation analysis of the VD model simulated abundances of the two predator, P_1 and P_2, and two prey, Z_1 and Z_2, species for $t \in [X, X]$ and initial state, $\mathbf{x}_0 = \{P_1, P_2, Z_1, Z_2\} = \{0.28, 0.50, 0.14, 0.18\}$ as a function of the predator coefficient β: (A) Bifurcation diagram that displays the local minima and maxima in the fluctuations of the simulated phytoplankton abundances, P_1, at different values of β; Time series plot of simulated abundances for (B) $\beta = 0.001$, (C) $\beta = 0.15$, and (D) $\beta = 0.4$. All other VD model parameters were kept constant at their WAVE values listed in Table 1. The bifurcation plot was generated with the GRIND Matlab package (`http://www.sparcscenter.org/grind`).

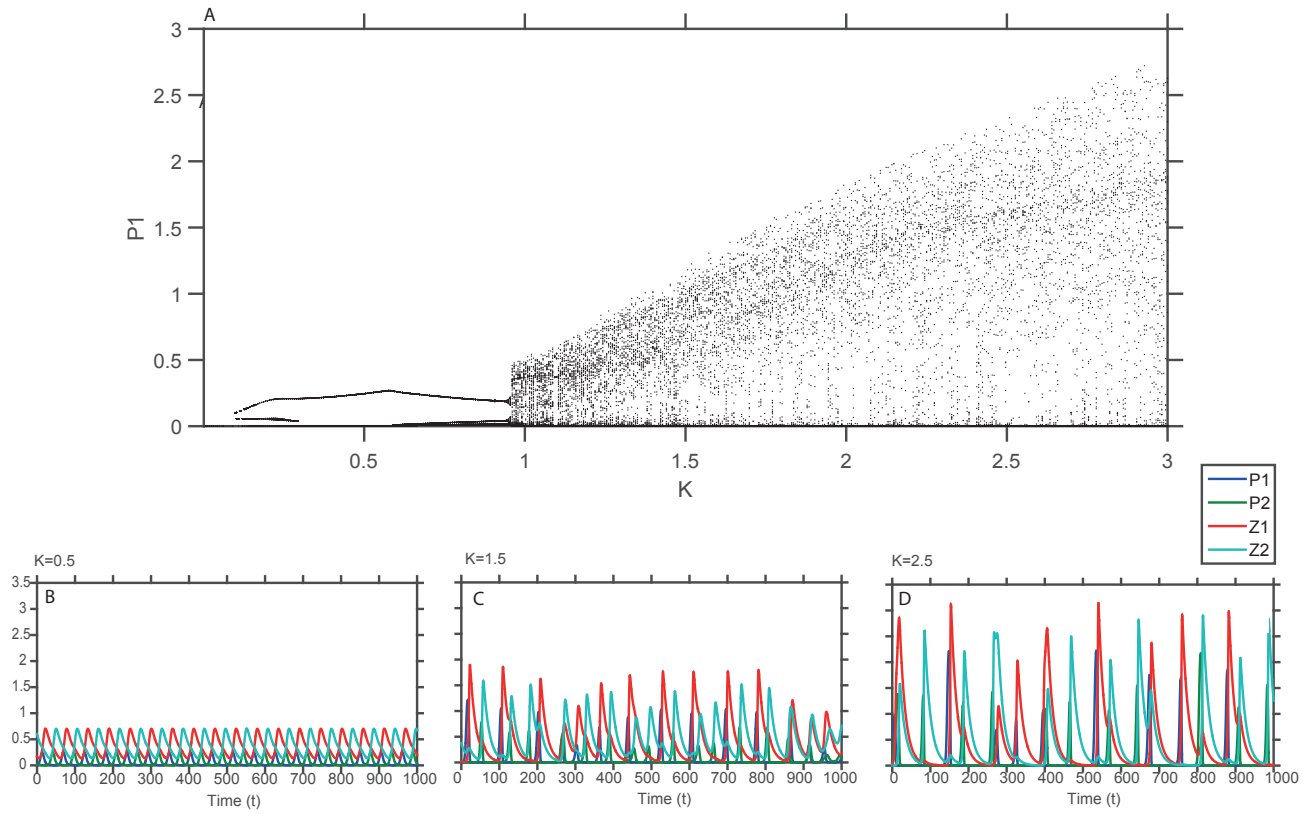

Figure 2.9: Bifurcation analysis of the simulated abundances of the VD model for $t \in [X, X]$ and initial state, $\mathbf{x}_0 = \{P_1, P_2, Z_1, Z_2\} = \{0.28, 0.50, 0.14, 0.18\}$ as a function of the carrying capacity K: (A) Bifurcation diagram that displays the local minima and maxima in the fluctuations of the simulated phytoplankton abundances, P_1, at different values of K; Time series plot of simulated abundances for (B) $K = 0.5$, (C) $K = 1.5$, and (D) $K = 2.5$. All other VD model parameters were kept constant at their WAVE values listed in Table 1. The bifurcation plot was generated using the GRIND Matlab package (`http://www.sparcscenter.org/grind`).

B: The particle filter

We briefly revisit our choice for the EnKF, rather than the more generic particle filter (PF). This DA method has does not use a state analysis step, instead assigns a likelihood to the state forecasts of each of the N particles. The state forecast distribution at each time is then constructed via the likelihoods of the particles. Periodic resampling is used to discard "bad" particles and to rejuvenate the state ensemble. This approach does not make distributional assumptions, and can characterize exactly the evolving posterior state distribution. Yet, we do not recommend the PF in our present application as the VD model is rather deficient in tracking the highly dynamic and chaotic state transitions of the mesocosm data set. This demands frequent resampling of the particles to negate model bias and characterize properly the evolving state distribution. In principle, this would be fine, were it not that resampling via the likelihood of the particles is highly inefficient compared to the state analysis step of the EnKF. Indeed, this method resurrects rapidly a biased ensemble by migrating the ensemble members' state forecasts towards their measured values. Thus, the EnKF exhibits the elasticity we need to track the behavior of highly nonlinear and/or chaotic systems.

2.6 Acknowledgements

The authors of this research greatly appreciate the support from the UC-Lab Fees Research Program Award 237285. All authors would like to thanks R. Heerkloss for sharing with us the mesocosm data set.

Chapter 3

Earth System Models: diagnosing uncertainties in model simulations

by Elias C. Massoud, Chonggang Xu, Rosie Fisher, Ashehad Ali, Nate McDowell, Brad Chrisrofferson, Dan Johnson, Charlie Koven, Jasper A. Vrugt, et al.

References:

Ali, A. A., Xu, C., Rogers, A., Fisher, R. A., Wullschleger, S. D., McDowell, N. G., Massoud, E. C., Vrugt, J. A., Muss, J. D., Fisher, J. B., Reich, P. B., and Wilson, C. J.: A global scale mechanistic model of the photosynthetic capacity, Geosci. Model Dev. Discuss., 8, 6217-6266, doi:10.5194/gmdd-8-6217-2015, 2015.

Massoud, E. C., Xu, C., Fisher, R. McDowell, N.G., et al.: Identification of key parameters controlling vegetation dynamics in a demographically structured land surface model. Journal of Advances in Modeling the Earth System, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017

Johnson, D.J., Needham, J., Xu C., Massoud, E. C., et al. Tree survival in the tropics reveals limited demographic diversity. Nature, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

This chapter of the thesis expands from systems of microscopic scale to a scale that can encompass the entire planet. Here, we investigate Earth System Models, models that are comprised of several components such as land, oceans, atmosphere, and cryosphere. This thesis will look at the land surface component, and extensive results are provided to discuss current uncertainties found in today's Land Surface Models (LSMs).

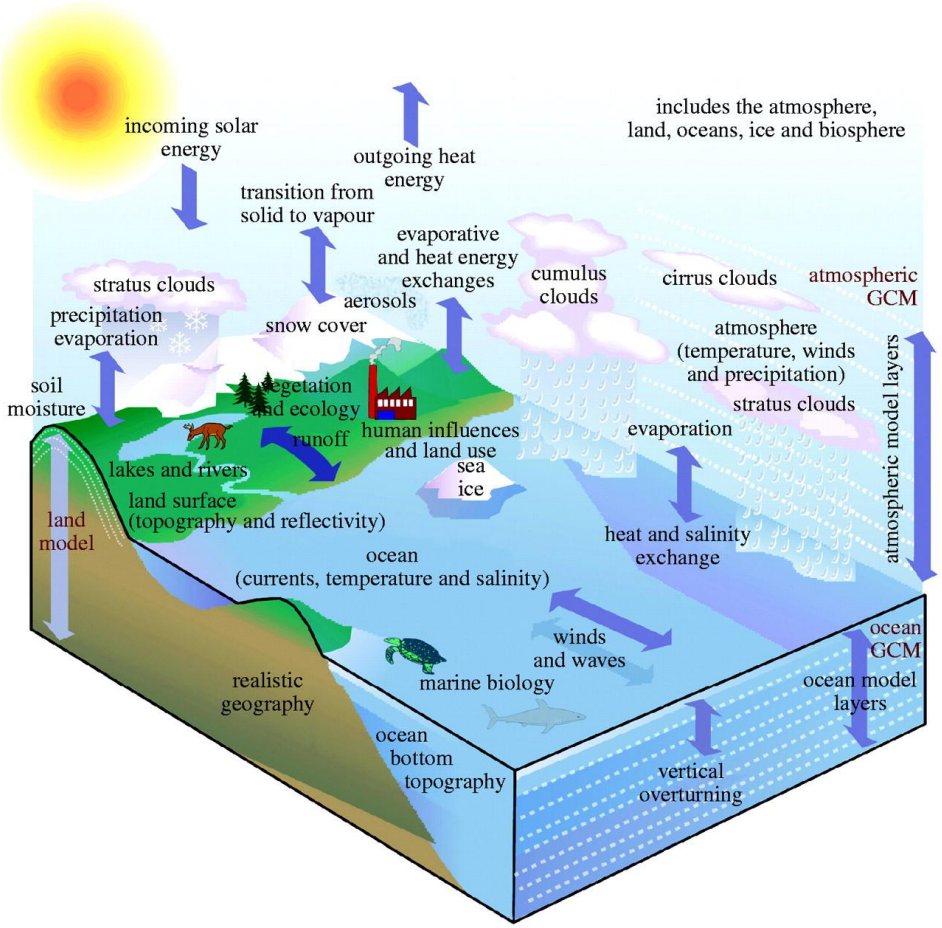

Figure 3.1: Schematic of an Earth System Model. The model formalizes and quantifies concepts of ecological climatology. Ecological climatology is an interdisciplinary framework to understand how natural and human changes in vegetation affect climate. It examines the physical, chemical, and biological processes by which terrestrial ecosystems affect and are affected by climate across a variety of spatial and temporal scales. The central theme is that terrestrial ecosystems, through their cycling of energy, water, chemical elements, and trace gases, are important determinants of climate. The land surface is a critical interface through which climate change impacts humans and ecosystems and through which humans and ecosystems can effect global environmental change. Reference: www.cesm.ucar.edu/models/clm

Three case studies are presented, each looking at different vegetation processes and associated uncertainties found in their respective model representation. The first study investigates a Land Surface model's parameter contribution to overall uncertainties found in the simulations; results provide a diagnosis for reducing uncertainties in LSM parameterizations. The second study related to uncertainties in LSMs introduces a next-generation nitrogen allocation model that represents the photosynthetic capacity at a global scale; this work provides a thorough calibration of the model to a global data set of photosynthetic capacity. Lastly, the third case study related to LSM uncertainties is a project that compares model simulations of tree mortality to a data set comprised a 2+ million observation of traits related to tree mortality in the tropics, and provides direction for future improvement of the representation of mortality in LSMs.

3.1 Manuscript Title - Identification of key parameters controlling vegetation dynamics in a demographically structured land surface model

Abstract: Land Surface Models (LSMs) assess terrestrial feedbacks to climate change, and of critical importance in governing these feedbacks is the representation of vegetation dynamics. In the last decade, vegetation demography has emerged in LSMs to better represent light-competition and thus coexistence of different plant functional types (PFTs). Outputs from these models are typically controlled by many parameters and Sensitivity Analysis (SA) methods are often applied to quantify the impact of each parameter in simulated fluxes and stores. In this study, we use the Fourier Amplitude Sensitivity Testing (FAST) method to conduct a SA on the Community Land Model (CLM) coupled to the Ecosystem Demography (ED), or CLM4.5(ED). While we confirm the importance of the photosynthetic capacity parameter ($V_{c,max25}$) in the model, which is consistently found to be important across terrestrial biosphere models, we also show the importance of carbon storage and allometry parameters that determine vegetation demography through their impacts on survival and growth strategies. Results shown in this study can be utilized in further diagnosis of CLM4.5(ED) as well as other emerging LSMs and for future field measurement efforts that are

directed towards reducing uncertainties in LSM parameterizations.

3.2 Introduction

Earth System Models (ESMs) are abstractions of nature used to simulate physical, chemical, and biological processes across the interacting domains of the atmosphere, ocean, ice and land (Arora et al., 2013; Claussen et al., 2002; Dunne et al., 2012; Hurrell et al., 2013) to estimate past, present, and forecast future conditions and climate. In this paper, we consider Land Surface Models (LSMs) or the land components of ESMs. LSMs consider the biophysical and biogeochemical interactions of the terrestrial surface, and examine how the land's feedbacks with the atmosphere affect the Earth's climate system. Such models typically contain a suite of different parameters to resolve the carbon, water, and energy fluxes and pools at the land-atmosphere interface (Bastidas et al., 1999; Gupta et al., 1999; Masson et al., 2003; Noilhan and Planton, 1989; Sargsyan et al., 2014). Many of these parameters can be estimated directly in the field, but others are difficult or impossible to measure due to various complications such as the lack of a physical meaning, technological limitations, or spatial/temporal aggregation (Entekhabi and Eagleson, 1989; Kumar et al., 2006).

Parameters that are observable in the field are also often subject to large natural variability, including changes through space and time (Fisher et al., 2015; Masson et al., 2003; Wood et al., 1992). For example vegetation parameters can be used to describe different root profiles (Vrugt et al., 2001b) or photosynthetic capacities (Leuning, 2002; Zeng, 2001), however parameter values in models are often pulled from the literature or databases that may not represent local variation or capture seasonal or ontogenetic changes. This can lead to significant divergence in multi-model ensemble projections or uncertainty in model predictions (Dietze et al., 2014b; McDowell et al., 2015; Rogers et al., 2017; Sitch et al., 2008). Since parameters are often defined in simulations with limited prior knowledge of their mean values and variation (Geromel, 1999; Kitanidis, 1986; O'Hagan and Leonard, 1976; Vrugt et al., 2002b, 2003), studies such as model sensitivity analysis are typically required to adequately quantify the impact of the parameters on the model simulations.

Today, many sensitivity analysis techniques are available (Helton, 1993; Saltelli et al., 2000;

Sobol', 1990). Saltelli et al. (2000) classified sensitivity techniques into two groups: local and global sensitivity analysis methods. Local sensitivity analysis techniques examine the response of the output(s) by varying input parameters one at a time and holding other parameters at central values. Global sensitivity techniques examine the response (averaged over the variation of all the parameters) of model output(s) by exploring a finite (or even an infinite) region of the parameter space. The local sensitivity analysis is relatively easy to implement, e.g. requires less model simulations and thus less computing power which may often be a limiting factor. However, the sensitivity index derived by local sensitivity analysis depends on the default values of the other parameters. This assumption that default values are satisfactory is questionable (e.g. Da Rocha et al. (1996); Groenendijk et al. (2011); Schwalm et al. (2010); Sen et al. (2001)) since the discrepancies in LSM predictions are strongly tied (through feedbacks of momentum, energy, mass and biogeochemistry) to the differences in their representation of the land surface (Crossley et al., 2000; Rosolem et al., 2013). Therefore, global sensitivity methods are generally preferred over local sensitivity methods when computing power is not a limiting factor.

LSMs are capable of representing vegetation dynamics through the use of a dynamic global vegetation(Arora et al., 2013; Cox et al., 2000; Friedlingstein et al., 2006; Krinner et al., 2005; Sato et al., 2007). In first generation models, which are still the most common (Fisher et al., 2015), the land surface at a given location is discretized into tiles according to plant functional type (PFT), with each PFT represented by a single big leaf. The abstraction of ecosystems into this simplistic structure makes it difficult to simulate light competition, and thus is often considered less than ideal for simulating processes at the plant to ecosystem scale. In the last decade, second-generation vegetation demographic models have emerged, that can capture coexistence and competition of PFTs driven by light-competition between different sizes of trees (represented as either cohorts or individuals) in a vertical canopy structure (Fisher et al., 2010b; Moorcroft et al., 2001; Scheiter et al., 2013) as well as successional dynamics through the representation of disturbance history (Hickler et al., 2004; Sitch et al., 2003; Thonicke et al., 2001). These next generation models also allow comparison with many more observed vegetation processes than first generation models, but also contain more degrees of freedom causing them to be more complex and subject to high

variability.

Our goal for this study is to apply a global sensitivity analysis, the Fourier Amplitude Sensitivity Test (FAST), on the Community Land Model coupled to the Ecosystem Demography, or the CLM4.5(ED) at a single site, to understand how the model behaves, to provide directions for improved model parameterization, and to assure that simulated output is consistent with our qualitative and quantitative understanding of vegetation dynamics. Specifically, we aim to answer the following question: what are the main parameter controls on vegetation processes such as growth and mortality and on the resultant dynamics of carbon fluxes and stocks? Even though CLM4.5(ED) simulates both carbon and water cycles, in this study, our analysis focused on carbon, because as the source for vegetation growth and dynamics, it exerts a strong control on emergent properties of community dynamics and associated ecosystem processes. The study region for the model simulations is in the Amazon basin and this analysis considers a single PFT that is a typical vegetation type for this region, the Broadleaf Evergreen Tropical Tree. Despite the need for such studies, systematic investigation of the parameter sensitivity of LSMs and/or DGVMs is not standard practice, potentially on account of the high dimensionality involved (although c.f. Fisher et al. (2010b); Pappas et al. (2013); Zaehle et al. (2005)). This is one of few formal global sensitivity analyses for an LSM with cohort-based vegetation demography (see Pappas et al. (2013)) and thus could provide important knowledge for model calibration as well as field measurement efforts.

This study is structured as follows: we first describe the CLM4.5(ED) and the reasoning behind the parameter selection for this study. We then explain the FAST method used to estimate parameter sensitivities in the model. Then, we illustrate the results where we show the model outputs as well as the corresponding parameter sensitivities. Finally, we discuss the results and their implications for diagnosing CLM4.5(ED).

3.3 Materials and methods

3.3.1 CLM4.5(ED)

CLM4.5(ED) is a community-based open-source model that is widely used for understanding climate-vegetation interactions. CLM is the land surface model used within various LSMs, including the Community Earth System Model (CESM) and the Norwegian Earth System Model (NorESM) (Bonan et al., 2011; Lawrence et al., 2011). The Ecosystem Demography (ED) concept is a method for scaling the behavior of forest ecosystems by aggregating individual trees into representative 'cohorts' based on their size, age and plant functional type (PFT), and by aggregating groups of cohorts into representative 'patches' (conceptually similar to a forest plot) which explicitly tracks the time between disturbances (Moorcroft et al., 2001). The main property of the ED concept that differs from most commonly used 'big-leaf' models is the capacity to predict distributions and compositions of plants directly from their given physiological traits described by the model parameterization (Fisher et al., 2015). This is achieved via the means of trait-filtering, whereby plant traits affect plant growth and survival, growth in turn affects the acquisition of light resources, and feeds back onto growth, survival and reproduction. Differences in growth, survival and reproduction rates thus directly control the relative distributions of vegetation types and their traits as well as the overall carbon stocks.

CLM4.5(ED) represents vegetation using size-structured groups of plants (cohorts) which co-exist on various successional trajectory-based land units. CLM4.5(ED) simulates growth by integrating photosynthesis across different leaf layers for each cohort, and mechanistic mortality is simulated based on plant carbon starvation and hydraulic failure (Fisher et al., 2015), in addition to a background mortality rate, tree-fall impact mortality, and fire. The model allocates photosynthetic carbon to different tissues such as leaf, root and stem based on the allometry of different tree species. In this original version of CLM4.5(ED), some processes are relatively under-represented; for example, the soil moisture impact on vegetation growth has a simple representation based on a single modifying factor for the moisture, an indication of the room for improvements in the existing structure of LSMs. CLM4.5(ED) can be simulated at different modes including point mode for

sites, regional mode for watershed or regional scale, and global mode for continental and global scale. See supplementary model description in Fisher et al. (2015) for details on specific components of the model structure.

In our simulations, CLM4.5(ED) aggregates many cohorts into 12 separate size class bins, and outputs the size-resolved information for each bin. There were 3 bins of trees with diameters less than 10 cm and we categorized them as small. There were 4 bins of trees with diameters larger than 10 cm and smaller than 50 cm and we categorized them as medium. Lastly, 5 bins of trees with diameters larger than 50 cm were categorized as large. For sensitivity analysis of each size category (small, medium and large trees), we choose to average the outputs over a 30 year interval in view that the transient and abrupt changes across different size categories at annual resolution could make the FAST analysis only accounts for a small amount of the variance contribution from each parameter.

3.3.2 Sensitivity Analysis: The FAST method

Global sensitivity analysis (SA) aims at quantifying the effects of input variables onto the variance of the response of a physical model such as an LSM. Among the abundant literature on sensitivity analysis methods, the Sobol' indices (Sobol', 1990), have received much attention since they provide accurate information for most models. However, the full description requires the evaluation of 2^n Monte Carlo integrals (Sudret, 2008), which is not practically feasible unless n is low (n here represents the dimensionality of the model, or the number of active parameters). The Fourier amplitude sensitivity test (FAST) is a popular sensitivity analysis technique that is computationally efficient and can be used effectively for nonlinear and nonmonotonic models (Sudret, 2008; Xu and Gertner, 2011). The FAST approach requires a tailored sampling strategy, which is what differentiates this method from other estimators of variance-based indices (Pianosi et al., 2016).

FAST uses a periodic approach to sample the parameter space and a Fourier transformation to decompose the variance of a model output into partial variances contributed by different model parameters. Only the first order sensitivity indices referring to the "main effect" of parameters

were calculated in the original method. In the 1990's, an extended FAST method able to calculate sensitivity indices referring to "total effect" was developed (Archer et al., 1997; Saltelli et al., 1999; Sobol', 1990). This "total effect" of a parameter's sensitivity refers to the sum of a parameter's individual contribution (1^{st} order sensitivity) and the contribution from its interaction with other parameters (higher order sensitivity) on the overall variance of the model output; that is, the total effect includes all the higher order interactions. The FAST method has found widespread use in many different fields of study including sensitivity analysis of the parameters of chemical reaction (Haaker and Verheijen, 2004), nuclear waste disposal (Lu and Mohanty, 2001), erosion [Wang et al., 2001], hydrologic (Francos et al., 2003), atmospheric (Kioutsioukis et al., 2004), matrix population, and forest landscape models (Xu and Gertner, 2009; Xu et al., 2009).

To describe the FAST method used herein, we can conveniently write the CLM4.5(ED) model as a nonlinear regression function, $\mathcal{F}(\cdot)$,

$$\mathbf{Y} = \mathcal{F}(\mathbf{x}, \mathbf{z}_0, \tilde{\mathbf{B}}), \tag{3.1}$$

which returns to the user a matrix, $\mathbf{Y}$, with N simulated values of m different outputs (e.g. GPP or LAI). Inputs to the model include, $\mathbf{x}$, an 87-element vector with parameters, $\mathbf{z}_0$, a vector that defines the initial state of the system (e.g. ground state with no vegetation in this case), and the control matrix $\tilde{\mathbf{B}}$ which summarizes the time-variant forcing variables. Note, that the tilde operator is used for the forcing data to signify explicit use of measured values.

The domain of the independent parameters is a hypercube (high dimensional space)

$$\Omega_n = \left(\mathbf{x} | x_i^{(\text{Min})} < x_i < x_i^{(\text{Max})} \; ; \quad i = 1 \; , ..., \; n\right) \tag{3.2}$$

where $x_i^{(\text{Min})}$ and $x_i^{(\text{Max})}$ are the minimum and maximum values for each parameter. In this study, we focus specifically on the sensitivity of the CLM4.5(ED) model parameters, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and ignore the contribution of the initial states and forcing variables in determining the variance in the model output. Our simulations involve monthly output for a period of $N = 130$ years, and returns m different variables which are used in our parameter sensitivity analysis with FAST. The

select m output variables include monthly basal area, change in diameter at breast height (dDBH), mortality, GPP, NPP, LAI, biomass, etc. The model is capable of producing outputs for various other processes, such as ecosystem-scale water fluxes, but these outputs are not investigated in this paper.

Based on Sobol's definition of variance-based sensitivity metrics (Sobol, 2001), we consider the time series of output of interest m, $\mathbf{Y}_m(t)$, to have corresponding variance at each time step $\mathrm{V}_m(t)$ = VAR($\mathbf{Y}_m(t)$). Then each parameter's contribution to this variance can be expressed as:

$$\mathrm{S}^1_{i,m}(t) = \frac{\mathrm{V}_i(t)}{\mathrm{V}_m(t)} \tag{3.3}$$

where $\mathrm{S}^1_{i,m}$ is the 1st order sensitivity of parameter x_i to output of interest m, V_i is the variance of parameter x_i, and V_m is the total variance of output m at time t.

We show that each parameters' contribution to the variance of the expected value of $\mathbf{Y}_m$ can be calculated as follows (see Xu and Gertner (2011)):

$$\mathrm{V}^i = \mathrm{V}_m\Big(E(\mathbf{Y}_m|\mathbf{Y}_i)\Big), \tag{3.4}$$

$$\mathrm{V}^{i,j} = \mathrm{V}_m\Big(E(\mathbf{Y}_m|\mathbf{Y}_{i,j})\Big), i \neq j, \tag{3.5}$$

$$\mathrm{V}^{i,j,k} = \mathrm{V}_m\Big(E(\mathbf{Y}_m|\mathbf{Y}_{i,j,k})\Big), i \neq j \neq k, \tag{3.6}$$

$$\ldots$$

$$\mathrm{V}^{[1,\ldots,n]ni} = \mathrm{V}_m\Big(E(\mathbf{Y}_m|\mathbf{Y}_{[i,\ldots,n]ni})\Big), \tag{3.7}$$

$$\mathrm{V}^{[1,\ldots,n]} = \mathrm{V}_m\Big(E(\mathbf{Y}_m|\mathbf{Y}_{[i,\ldots,n]})\Big), \tag{3.8}$$

where $\mathrm{V}^i(\cdot)$ is the 1st order parameter contribution of parameter i to the variance of the output, $\mathrm{V}^{i,j}(\cdot)$ is the 2nd order parameter contribution of parameters i and j to the variance of the output, $\mathrm{V}^m(\cdot)$ is the variance of the model output, $E(\cdot)$ are the expected values of the model outputs, and $[1, ..., n]ni$ represents all parameters except i.

For a subset of parameters ($\mathbf{x}_{\text{sub}}$), $V_{\mathbf{x},\text{sub}}$ represents the partial variance in the model output due to the uncertainty in the subset of parameters of $\mathbf{x}_{\text{sub}}$. Therefore, $\mathrm{V}^{[1,\ldots,n]}$ represents the variance in the model output resulting from uncertainties in all the model parameters:

$$\mathrm{V}^{[1,\ldots,n]} = \mathrm{V}_m. \tag{3.9}$$

We define

$$\mathrm{V}_i = \mathrm{V}^i, \tag{3.10}$$

$$\mathrm{V}_{i,j} = \mathrm{V}^{i,j} - \mathrm{V}_i - \mathrm{V}_j, \tag{3.11}$$

$$\ldots$$

$$\mathrm{V}_{[1,\ldots,n]} = \mathrm{V}^{[1,\ldots,n]} - \Big(\sum_{s=1}^{r} \sum_{1\le i_1\le i_2\le\cdots\le i_s\le n} \mathrm{V}_{x_{i_1}\cdots x_{i_s}}\Big), r = 1, 2, \cdots, n-1 \tag{3.12}$$

as partial variances of model output contributed by the first-order (or main) effects, the second order interaction effects, the third-order interaction effects, and so on, until the n^{th} order interaction effects of model parameters (Saltelli and Tarantola, 2002),. Summing all the left and right terms in Equations 10-12, we get the variance decomposition as follows:

$$\mathrm{V}_m = \sum_{i=1}^{n} \mathrm{V}_i + \sum_{i\neq j}^{n} \mathrm{V}_{i,j} + \sum_{i\neq j\neq k}^{n} \mathrm{V}_{i,j,k} + \cdots + \mathrm{V}_{[1,\ldots,n]}, \tag{3.13}$$

which suggests that the total variance resulting from parameter uncertainties can be decomposed into partial variances contributed by the first-order effects, the second-order interaction effects, third-order interaction effects, and until the nth order interaction effects of parameters. Dividing both sides of Equation 13 by $\mathrm{V}^{[1,\ldots,n]}$ (or by V_m), we get

$$1 = \sum_{i=1}^{n} \alpha_i + \sum_{i\neq j}^{n} \alpha_{i,j} + \sum_{i\neq j\neq k}^{n} \alpha_{i,j,k} + \cdots + \alpha_{[1,\ldots,n]}, \tag{3.14}$$

where

$$\alpha_i = \frac{\mathrm{V}_i(t)}{\mathrm{V}_m} = \mathrm{S}^1_{i,m}, \tag{3.15}$$

$$\alpha_{i,j} = \frac{\mathrm{V}_{i,j}(t)}{\mathrm{V}_m} = \mathrm{S}^2_{i,j,m}, \tag{3.16}$$

$$\ldots$$

$$\alpha_{[1,\ldots,n]} = \frac{\mathrm{V}_{[1,\ldots,n]}}{\mathrm{V}_m} = \mathrm{S}^n_{i,m}, \tag{3.17}$$

represent the first-order, second-order, and so on, until n^{th} order sensitivity indices, respectively.

In this study, we identify both the 1^{st} and 2^{nd} order sensitivities of the model parameters. It is possible to identify the specific interactions with FAST (higher orders); however, because of the sample size limitations for a larger tri-variate parameter space, the FAST-based estimation of third-order sensitivity indices would be less reliable. We now can calculate the 1^{st} and 2^{nd} order sensitivities of the model parameters for each output of interest and at each time step. Further details on the FAST method can be found in (Xu and Gertner, 2007, 2011, 2009; Xu et al., 2009).

3.3.3 Parameter Selection

There are more than 200 parameters in CLM4.5(ED). In this study, we focus solely on vegetation components resulting in 87 parameters that are relevant to vegetation processes, including parameters for photosynthetic processes, temperature response, allometry description, radiative transfer, recruitment, turnover and mortality. See Table A1-A4 in the appendix for a complete list of the parameters used in this study, with corresponding description, units, default values, and applied ranges. Refer to Appendix A for the allometry equations, Appendix B for the temperature response curve (photosynthesis) equations, and Appendix C for the carbon storage equations used in CLM4.5(ED).

We use a broad definition of parameter and extract numerous features of the model that were 'hard-wired' in previous CLM versions. The FAST algorithm requires valid ranges to be chosen for

each parameter, which creates the possible space (Ω_n) to sample from. In theory, each parameter has a corresponding observational distribution that produces the ideal space for sampling (LeBauer et al., 2013). However, in this study there are both a large parameter set and a scarcity of appropriate data sources for Amazonian forests for many of the relevant quantities, therefore obtaining a robust data-supported distribution for each parameter was not feasible. Thus, the parameter ranges in this study were generated by applying a uniform distribution over a range that spans +/- 15 % of the default parameter values of CLM4.5(ED) (i.e. default parameter values for tropical evergreen trees). Then using FAST, 5000 parameter combinations were sampled from the parameter space, Ω_n. The sample size was determined using the method of Xu and Gertner (2011) where it is appropriate to use 100 times the number of effective (important) parameters. The 5000 model runs cost about 32 CPU hours for each simulation, and thus we ran our simulations for a total of 160,000 CPU hours on the Los Alamos National Laboratory (LANL) Conejo super computer.

In this analysis, we assume the majority of CLM4.5(ED) parameters to be non-correlated with uniform probability due to the limitation of data for covariate traits for the 80+ parameters in this study. However, we do need to take care of the correlation among parameters in the temperature response functions (Appendix B) in order to generate realistic temperature response curves. These parameters were tested for correlation using a published dataset (Leuning, 2002), which showed that the photosynthetic parameters for activation energy (e.g. $V_{\mathrm{c,max,ha}}$) were not necessarily correlated with the other photosynthetic parameters. However the parameters for deactivation energy (e.g. $V_{\mathrm{c,max,hd}}$) and those related to entropy terms (e.g. $V_{\mathrm{c,max,se}}$) were highly correlated as expected. Thus, each of these parameters' samples were generated from the same location in their relative parameter spaces, which maintains their correlation.

3.3.4 Site Description

In this study, CLM4.5(ED) is tested for a site in a moist-tropical forest in the State of Par, The Amazon, Brazil (7° S, 55° W). We initialized the runs (i.e. $\mathbf{z}_0$ in Equation 1) with a bare ground, or

a state with no vegetation, and simulated the forest dynamics for 130 years, which is enough time for the ecosystem to reach equilibrium dynamics by checking the simulated biomass, basal area and various carbon fluxes. By choosing to start from bare ground and running the model until it reaches a quasi-steady-state size distribution, rather than by examining short runs initialized from observed initial forest size distributions (e.g. Dietze et al. (2014b)), we are deliberately allowing the ecosystem demographic structure itself to be an outcome of the parametric uncertainty rather than a separate, possibly non-self-consistent, initial condition uncertainty. This allows uncertainty to propagate across timescales and therefore gives a more complete estimate of the uncertainty in the system. The climate conditions (i.e. $\tilde{\mathbf{B}}$ in Equation 1) for this site are from Qian et al. (2006) representative of data from 1948-1972 and recycled for the 130 year simulations. CLM4.5(ED) has the potential of simulating dynamics for an ecosystem with many PFT's. In this study, in order to better understand the biophysical, biochemical, and vertical light competition behavior of the model with respect to carbon and water cycling, we focus only on a single broadleaf evergreen tree PFT.

3.4 Results

In this section we will highlight the outputs from CLM4.5(ED) and corresponding uncertainties generated from the 5000 simulations obtained for the FAST analysis. We go on to discuss the sensitive parameters for each output. We first investigate the forest demographic dynamics, diagnosing the growth and mortality processes simulated in CLM4.5(ED). Then, we analyze the forest carbon cycles, specifically the carbon fluxes and stocks in the model simulations.

3.4.1 Forest demographic dynamics: growth and mortality

One of the key properties of CLM4.5(ED) is that vegetation is represented as cohorts of varying sizes for more realistic simulation of light competition in the canopy. To better understand the key drivers of growth and mortality for different size of trees, we group various cohorts of trees into 3 size categories for analysis purpose: small ($diameter < 10$ cm), medium (10 cm $< diameter <$ 50

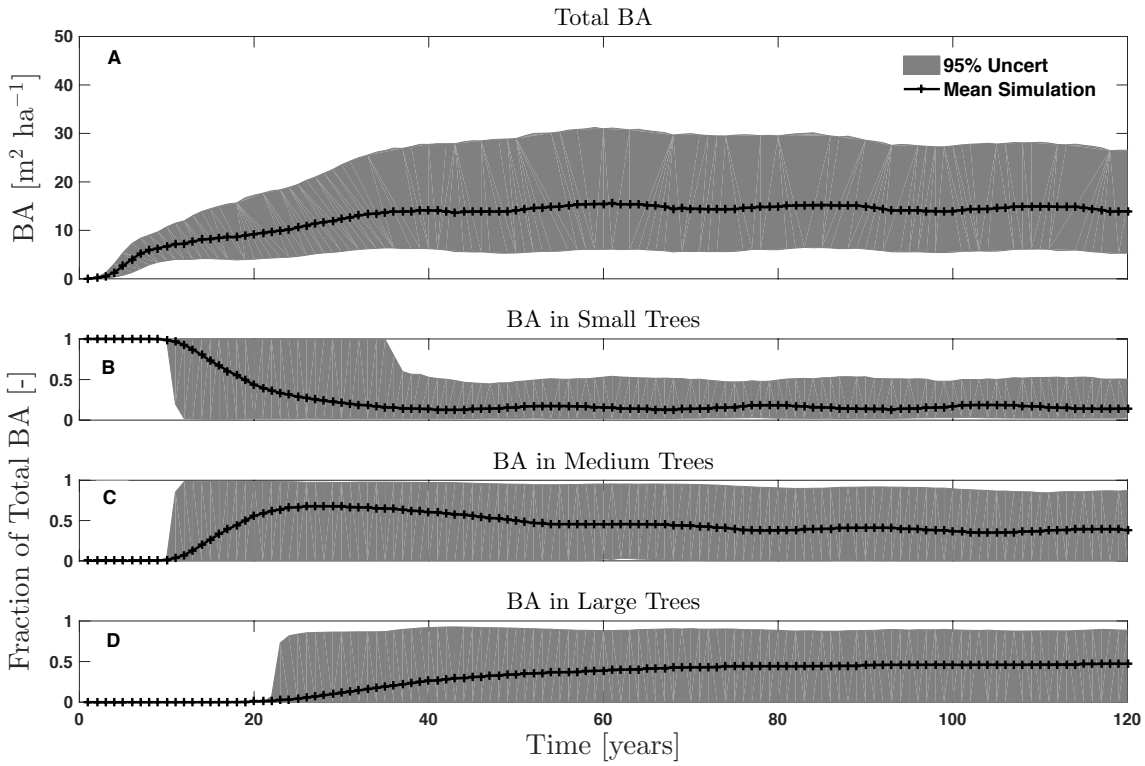

Figure 3.2: Outputs from CLM4.5(ED) for basal area (BA). Figures show the outputs for the various tree sizes considered, including small (*diameter* < 10 cm), medium (10 cm < *diameter* < 50 cm), and large trees (*diameter* > 50 cm). Shown are the mean simulation (black line) with 95% confidence intervals.

cm), and large (*diameter* > 50 cm) trees. Since the model runs were initialized from a near-bare ground state (i.e. with an initial density of half-centimeter diameter saplings), the simulated basal area (BA) of the forest, which is the fraction of stem cross-sectional area per ground surface area showed that (Fig. 1 B-D) the simulated forest grows from small trees (years 0 - ∼20), to medium trees (years ∼20 - 45), and finally to large trees (after 45 - 50 years). Our FAST analysis shows that a key parameter that controls the fraction of BA in different size of trees is the stem allometry coefficient c (Fig. 2). Specifically, a higher value of stem allometry coefficient c, or a higher allocation of carbon to stem (see Eq. A1 in the appendix), will eventually lead to a lower fraction of BA in large trees than in smaller trees in the model simulations (Fig. A1). This is because a higher value of stem allometry coefficient c ultimately leads to a faster reduction in stem growth in large trees (in cm yr^{-1} ha^{-1}), i.e. change in diameter at breast height (dDBH), due to a lower

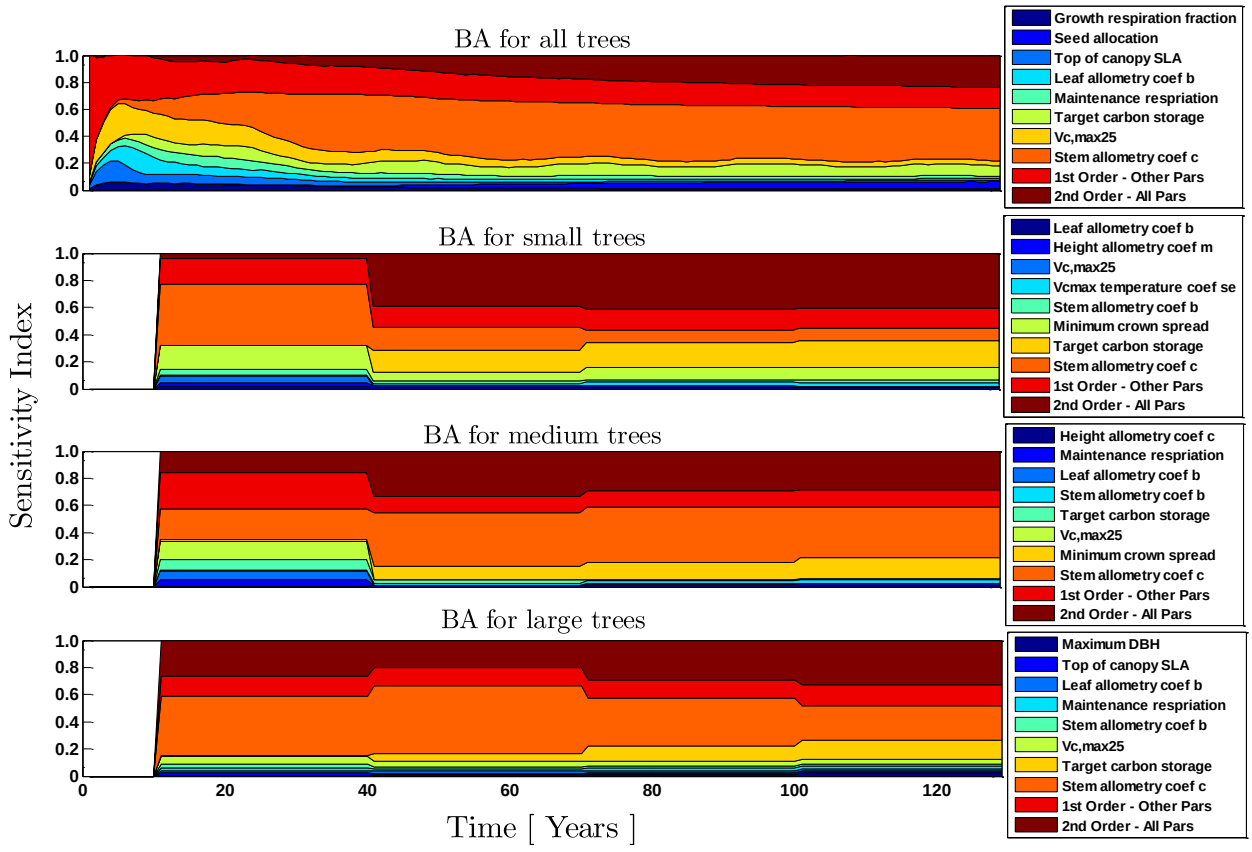

Figure 3.3: Sensitivity index of the model parameters (1^{st} order) for basal area (BA) outputs from CLM4.5(ED) (outputs are in units [m^2 ha^{-1}]). Parameter sensitivities to the CLM4.5(ED) outputs were analyzed for various tree sizes, including small ($diameter < 10$ cm), medium (10 cm $< diameter <$ 50 cm), and large trees ($diameter > 50$ cm). Also shown are sensitivities of the remaining parameters ('1^{st} order - Other Pars') as well as the sensitivity of parameter interactions ('2^{nd} order - All Pars').

proportion of carbon allocated to leaves for productivity (Fig. A2).

The mean of simulated BA and associated uncertainty ranges from the 5000 simulations stabilize after year 100 (Fig. 1A). This could result from the balance of growth and mortality processes represented in the model. For the stem growth in terms of dDBH averaged per tree, small trees observe lower rates of growth compared to medium or large trees (Fig. 3 B-D). However, the fraction of overall stem diameter growth is dominated by the small trees (Fig. A5) due to their high densities (Fig. A3). Our FAST analysis shows that the most sensitive parameters for tree growth were the target storage carbon and stem allometry parameters, however, the importance magnitude varies for different sizes of trees (Fig. 4). We observe that the stem allometry coefficient c showed higher sensitivity for medium and large trees in terms of DBH growth than that for small trees (Fig. 4). The target storage carbon determines the target amount of carbon for the plant to store relative to the leaf biomass (see Appendix C for details). In terms of growth, the target carbon storage parameter is more important for small trees than for medium and large trees, as smaller trees are shaded and thus more vulnerable to changes in the amount of carbon storage that will affect carbon allocations (see Eq. A9 in Appendix C). Yet, for the stem density, or the number of trees in a given area, the target carbon storage parameter is less important for small trees compared to medium and large trees (Fig. A11). Our sensitivity analysis also show specifically important parameters for different sizes of trees. For example, leaf allometry was specifically important for small trees, $V_{\mathrm{c,max25}}$ for medium trees, and seed allocation was solely sensitive for large trees.

Mortality is an important driver for the simulated forest dynamics in CLM4.5(ED). It includes four modes of mortality: 1) fixed background mortality, 2) hydraulic failure based on a threshold of very low soil moisture; 3) carbon starvation resulting from the depletion of carbon storage in plants (see Appendix C for details); and 4) impact mortality resulting from fall of big trees (Fisher et al 2015). In the model, carbon starvation is the main driver for the overall mortality (Fig. A4 in the appendix). Carbon starvation based mortality uses a threshold of carbon storage to trigger mortality (see Appendix C). Under shaded conditions, less carbon storage as determined by target carbon storage, respiration, and NPP could lead to a higher mortality. In terms of the

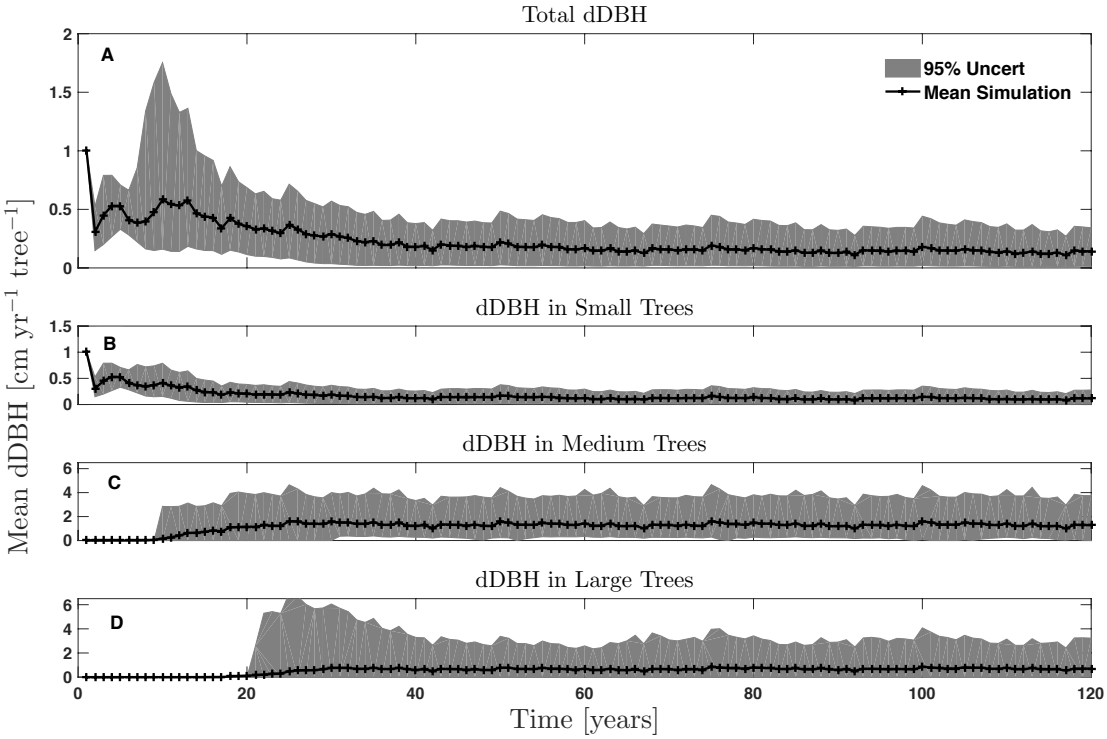

Figure 3.4: Outputs from CLM4.5(ED) for the change in diameter at breast height (dDBH). Figures show the mean dDBH (in cm yr^{-1} tree^{-1}) for the various tree sizes considered, including small (*diameter* < 10 cm), medium (10 cm < *diameter* < 50 cm), and large trees (*diameter* > 50 cm). Shown are the mean simulation (black line) with 95% confidence intervals.

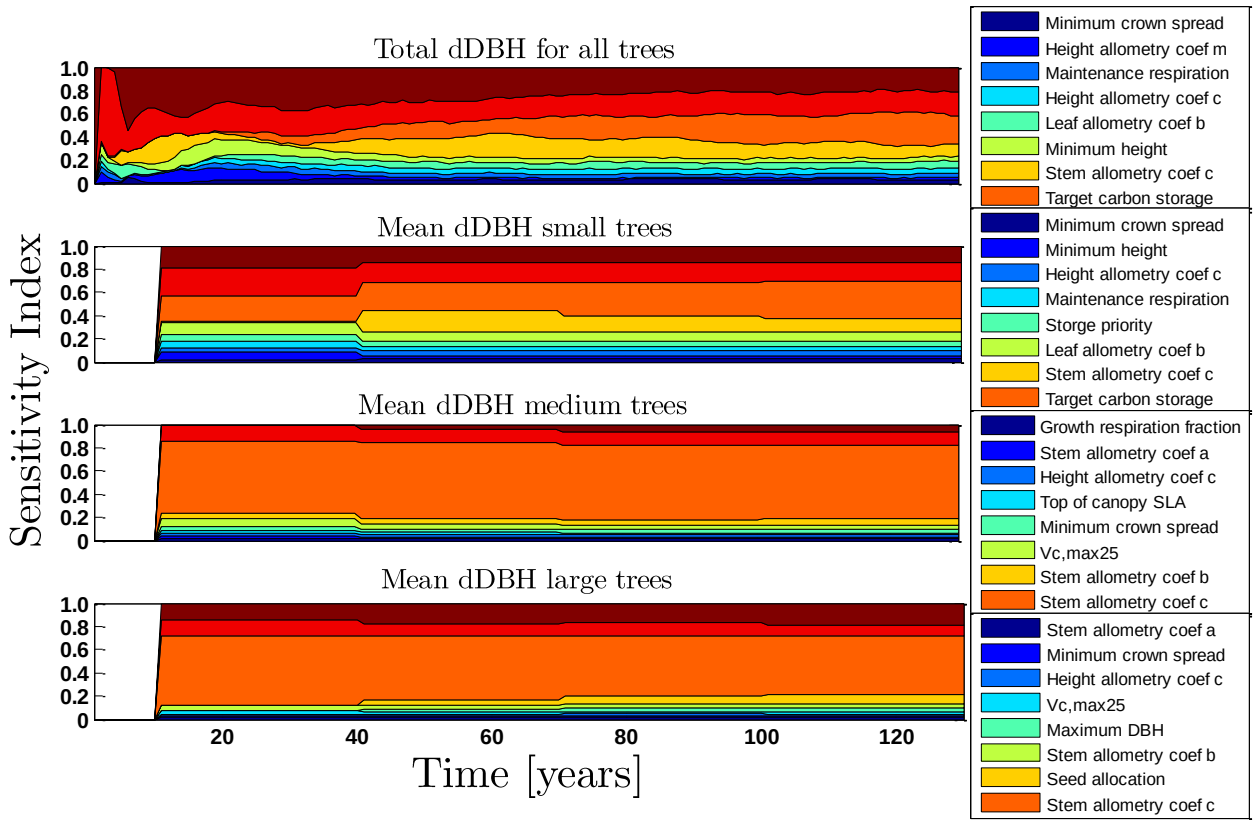

Figure 3.5: Sensitivity index of the model parameters (1st order) for change in diameter at breast height (dDBH) outputs from CLM4.5(ED) (outputs are in units [cm yr^{-1} tree^{-1}]). Parameter sensitivities to the CLM4.5(ED) outputs were analyzed for various tree sizes, including small (*diameter* $<$ 10 cm), medium (10 cm $<$ *diameter* $<$ 50 cm), and large trees (*diameter* $>$ 50 cm). Also shown are sensitivities of the remaining parameters in red ('1st order - Other Pars') as well as the sensitivity of parameter interactions in magenta ('2nd order - All Pars').

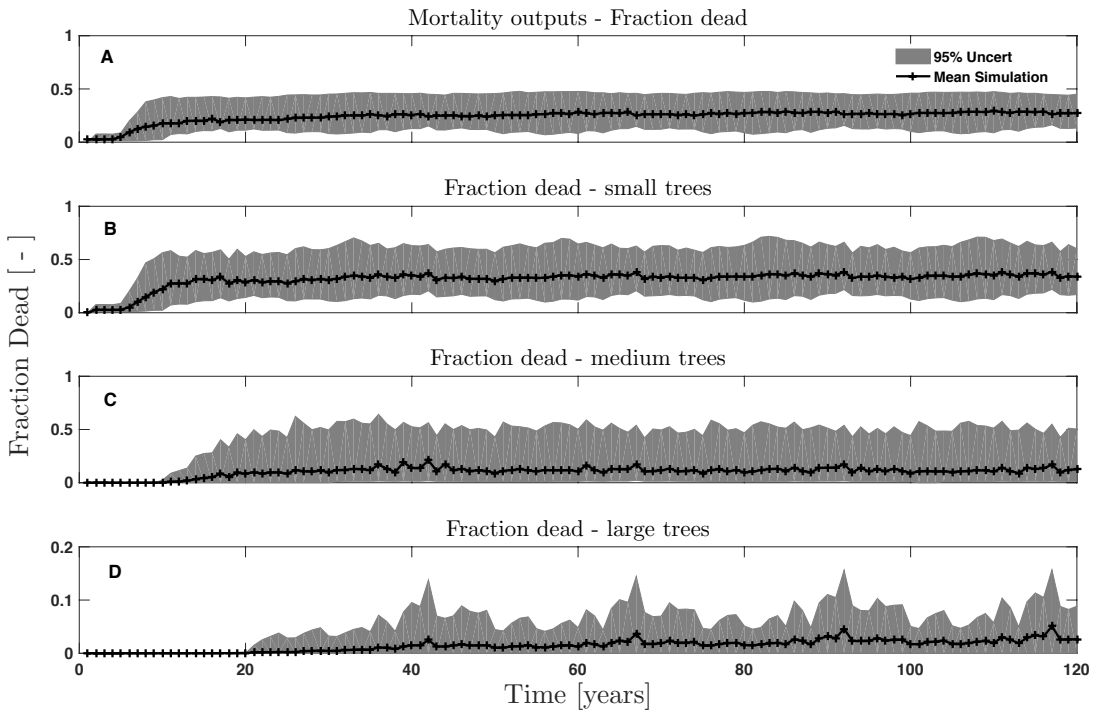

Figure 3.6: Outputs from CLM4.5(ED) for mortality, or the fraction of each trees that die in each year. Figures show the outputs for the various tree sizes considered, including small ($diameter < 10$ cm), medium (10 cm $< diameter <$ 50 cm), and large trees ($diameter > 50$ cm). Shown are the mean simulation (black line) with 95% confidence intervals.

number of plants dead, as expected, it is much higher in the small size of trees (Fig. 5). The sensitivity analysis of predicted mortality rate (percentage of mortality) shows that the dominant parameter for predicting mortality of large trees is the target carbon storage (Fig. 6); however, for small and medium trees, other parameters such as allometric and photosynthetic parameters that could potentially determine their competitive advantages in the canopy are also important (Fig. 6). Specifically, for medium size trees, the mortality rate is affected by both the stem allocation and targeted carbon storage parameters (Fig. 6). For the small trees, important parameters include the photosynthetic parameter ($V_{\mathrm{c,max25}}$), stem allocation coefficient c, leaf maintenance respiration, and stress mortality rate.

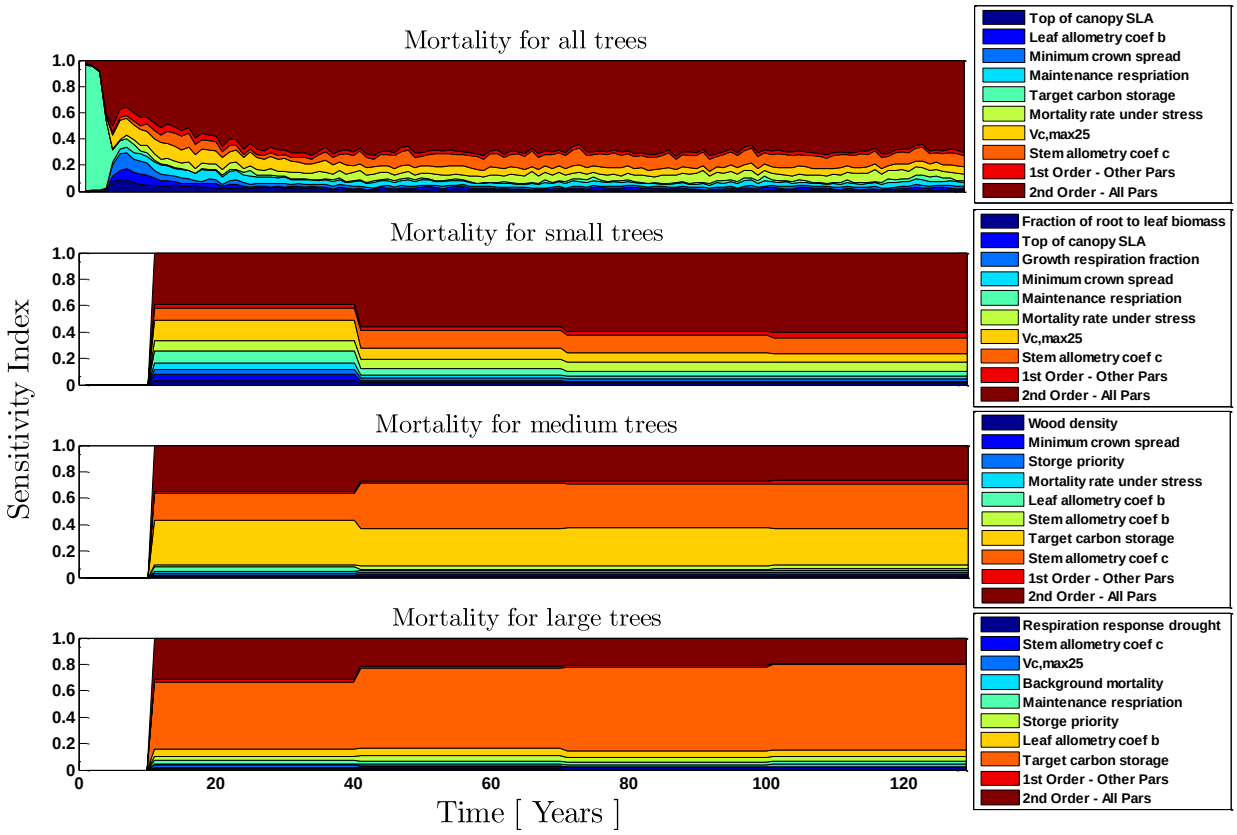

Figure 3.7: Sensitivity index of the model parameters (1^{st} order) for the mortality outputs of CLM4.5(ED) (outputs for mortality are in units [% yr^{-1}] since the fraction of dead trees was considered for this analysis). The sensitivity values shown in this figure reflect parameter influence on the mortality outputs that calculated as the fraction of dead trees in each category, rather than the number of dead trees in each category. Also shown are sensitivities of the remaining parameters ('1^{st} order - Other Pars') as well as the sensitivity of parameter interactions ('2^{nd} order - All Pars').

3.4.2 Forest carbon cycles: carbon fluxes and stocks

The goal of CLM4.5(ED) is to predict carbon fluxes and stocks that can be fed into the ESMs by simulating the vegetation demography. In order to understand the key parametric control on carbon fluxes and stocks, we specifically investigate parameter sensitivities for GPP, NPP, LAI and total forest biomass. Our results show that GPP and NPP levels increased consistently for the first 10 years of the simulations, which was expected for a forest growing from bare ground. However within a fairly short period of 5-10 years, GPP, NPP and LAI and their uncertainty ranges reached a quasi-stable rate (Fig. 7). This amount of time to reach equilibrium is much shorter compared to the basal area (Fig. 1A) and the total biomass accumulations (Fig. 7). Meanwhile, our FAST sensitivity analysis showed that, for carbon fluxes of GPP and NPP, the photosynthetic capacity parameter ($V_{\mathrm{c,max25}}$) was most sensitive, which was not surprising as it controls leaf-level carbon uptake (Fig. 8). Specifically for NPP, the respiration parameters showed high sensitivity, such as the growth and leaf maintenance respiration. For TLAI, the leaf allometry parameter was most sensitive as it determines carbon allocation for leaves and the stem allometry parameter was most sensitive for total biomass as it determines carbon allocation to the stem. A common important parameter is the target carbon storage that is important for GPP, NPP, TLAI and total biomass (Fig. 8). This results from the fact that the target carbon storage is a key driver for mortality especially for medium and large trees in the simulations (Fig. 6), which accounts for a large proportion of biomass (Fig. A6 in the appendix) and GPP (Fig. A7).

Our bi-variate spline analysis (Wahba, 1990) shows that, for $V_{\mathrm{c,max25}}$ and target storage carbon, an increase in either of these parameters will cause an increase in the output of GPP, NPP, TLAI and biomass (Fig. 9). For the parameters related to leaf and stem allometry, however, the relations may differ depending on the output and the year of interest. At year 130, the higher leaf allocation normally leads to higher fluxes (NPP and GPP) but less biomass. Meanwhile, higher stem allocation lead to higher biomass but smaller fluxes (NPP and GPP). This suggests that the trade-offs between carbon allocation to stem vs leaf tissues leads to a corresponding tradeoff between carbon stocks and productivity in the model predictions.

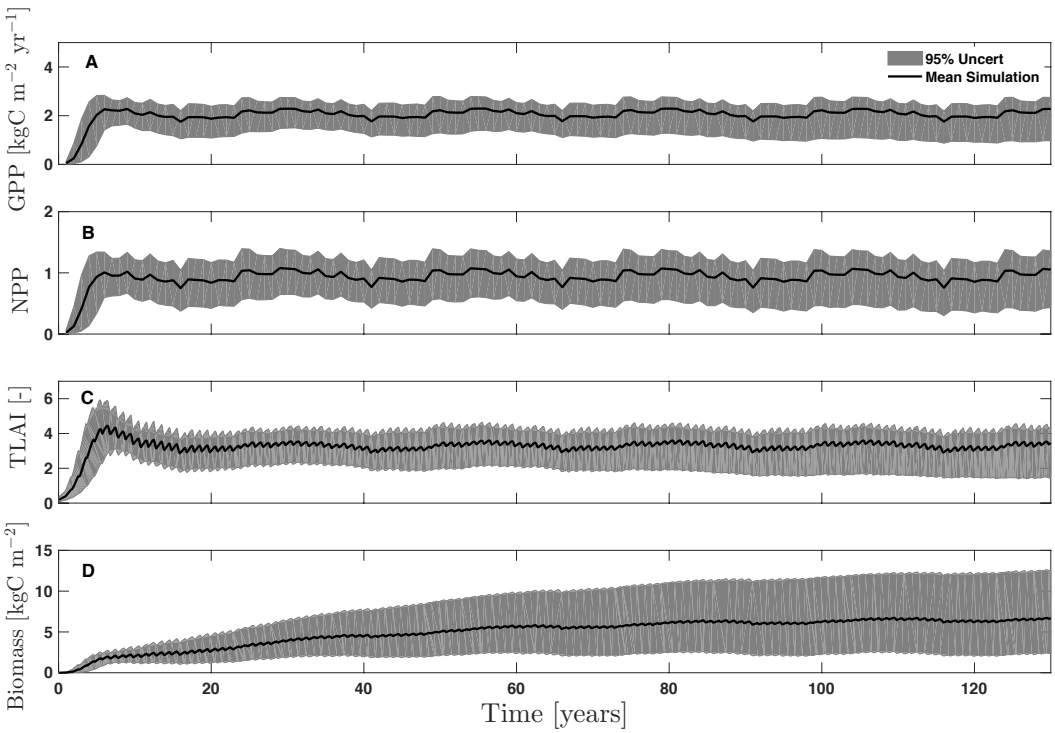

Figure 3.8: Outputs from CLM4.5(ED) model include GPP, NPP, LAI, and biomass. Shown are the mean simulation (black line) with 95% confidence intervals. The system is initialized with a bare ground, and this is shown with initial values of 0 for the different outputs.

3.5 Discussion

While second generation vegetation demographic models such as CLM4.5(ED) provide us great opportunities to predict carbon cycles with LSMs, the larger number of parameters also creates challenges for identifying key processes for further investigation. Our analysis shows similar results to sensitivity analysis on first generation 'big-leaf' vegetation models (e.g., Sargsyan et al. (2014)) in view that photosynthetic capacity, $V_{\mathrm{c,max25}}$, is a key parameter for predicting GPP and NPP fluxes; however, we do show important parameters that are unique to LSMs with second generation vegetation demography. Specifically, results shown here indicate the importance of leaf and stem allometry parameters, which control dynamic carbon allocation strategies based on size, and thus control the general vegetative state and size structure of the forest (reference Waring et al. (1998); Waring and Running (2010)). Our results also show that storage carbon has a key role in vegetation dynamics as it regulates processes controlling survival and mortality in the model (Dietze et al.,

2014a; Fisher et al., 2010a; McDowell et al., 2016). We want to point out that this may not represent the 'true' mechanisms that kill trees in the tropics. For example, Rowland et al. (2015) showed that death from drought in tropical forests is triggered by hydraulics but not carbon starvation. In the CLM4.5(ED), the hydraulic failure is only based on a simple soil moisture threshold (Fisher et al., 2015) but did not consider the plant hydraulics. Ongoing research for CLM4.5(ED) that account for the plant hydrodynamics is expected to better capture the hydraulic failure mortality (e.g. Christoffersen et al. (2016)).

In CLM4.5(ED), storage carbon appears to control the degree to which plants can tolerate conditions in the understory, and might thus be somewhat analogous to a mechanistic control over shade tolerance. In the model, the carbon starvation based mortality use a threshold of carbon storage to trigger mortality (see Appendix C). Under shaded conditions, less carbon storage as determined by target carbon storage, respiration and NPP could leader to a higher mortality. The biological control over shade tolerance in tropical trees remains under significant debate (Kitajima and Poorter, 2010; Sterck et al., 2014; Wright et al., 2010); in this case, the importance of stored carbon indicates that the degree of shade tolerance is a critical input parameter and feature of the model output. Overall, the results indicate that that large trees persist in the forest and small trees are recycled in the model simulations, since the most mortality and growth are seen for small trees and highest levels of BA and biomass are seen for the large trees. This is why target carbon storage is so important; it is the strategy of how much carbon to store in trees for later utilization, which will produce more large trees and thus more persistent trees (less overall recycling of carbon and more storage).

A parameter's individual contribution to the variance in a certain output is the 1^{st} order sensitivity index for that parameter, where the 2^{nd} order sensitivity is that due to bi-parameter interactions. For the CLM4.5(ED) outputs, the parameter interactions that contribute to the overall variance in BA are the stem allometry coefficients b and c, along with respiration parameters such as maintenance respiration or growth respiration fraction (Fig. A8), which is consistent with findings from Dietze et al. (2014b). Similar results are shown for dDBH (Fig. A9). The 2^{nd} order

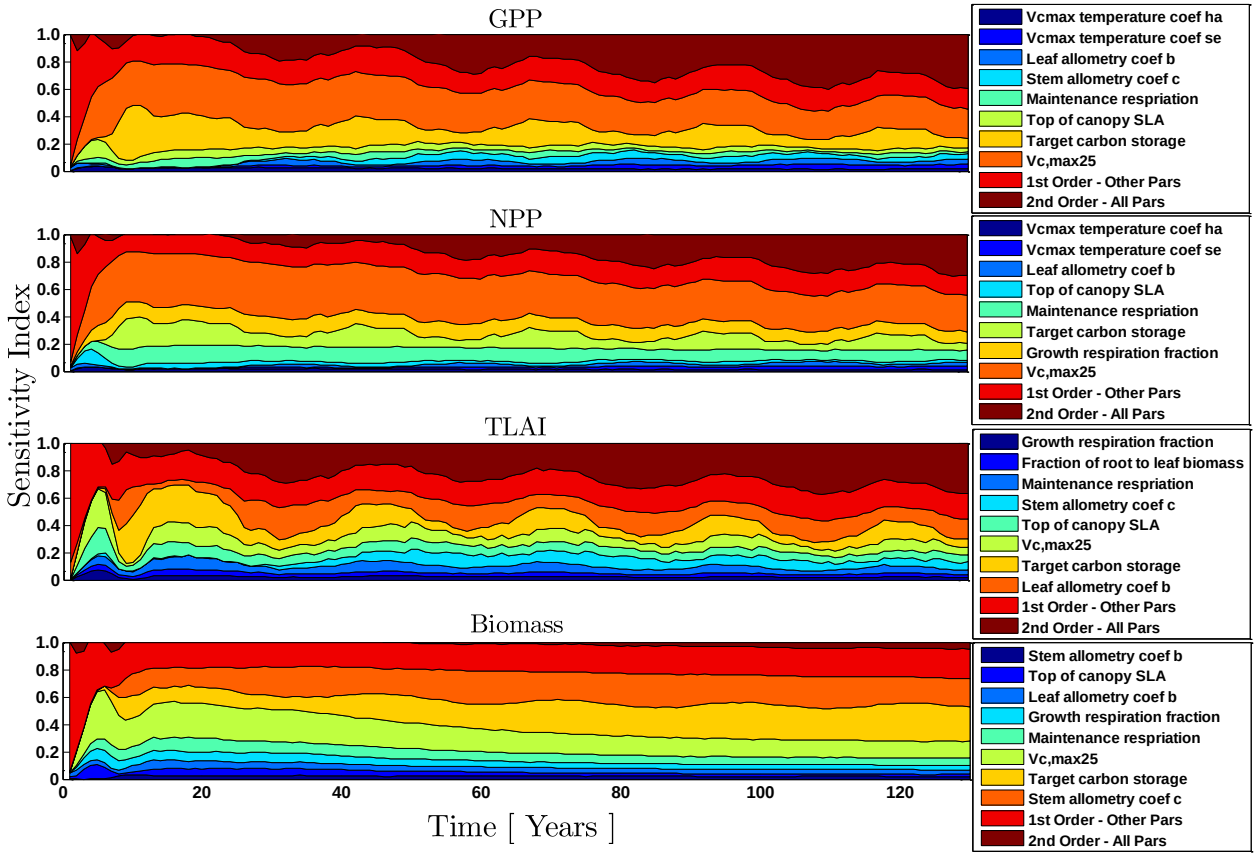

Figure 3.9: Sensitivity index of the model parameters (1st order) for the outputs of interest, including CLM4.5(ED) outputs of GPP, NPP, TLAI, and biomass (units for each output are shown in Fig. 7). Also shown are sensitivities of the remaining parameters ('1st order - Other Pars') as well as the sensitivity of parameter interactions ('2nd order - All Pars').

sensitivities of the parameters to outputs of GPP, NPP, TLAI, and biomass were also estimated (Fig. A10). For GPP and NPP, the leaf allometry coefficient b,$V_{c,max}$ temperature coefficient se, top of canopy specific leaf area, and growth respiration fraction were the most sensitive parameters in regards to 2nd order interactions. For TLAI and biomass, the growth respiration fraction and $V_{c,max25}$ parameters had high 2nd order sensitivity. Therefore, a few parameters seem to contribute the most to the variance in model outputs, and this is seen in both the 1st and 2nd order sensitivity estimations. Thus, parameters that contribute consistently to the variance in the CLM4.5(ED) outputs are namely the photosynthetic parameter ($V_{c,max25}$), the storage carbon parameter, and leaf and stem allometry parameters.

In theory, improvements in older generation models can be achieved when their relatively simple structures are replaced with next-generation process based models. However, this comes at the cost of increased complexity, which requires more detailed testing and assessment of the uncertainty in the more complex models. As shown for many of the outputs, the $V_{c,max25}$ parameter accounted for a large portion of the variance. It is possible, for example, that by replacing the current formulation of the photosynthetic process in the CLM4.5(ED) with a model that more accurately represents the allocation of nitrogen and thus the photosynthetic process (see Ali et al. (2016); Xu et al. (2013)), a reduction in the overall variance of model projections can be achieved. However, it has also been shown that proper parameter constraint can also lead to reduced model sensitivity to $V_{c,max25}$ (Dietze et al., 2014b). Moreover even with additional model process representation the underlying photosynthesis sub-models in DGVMs share similar forms but still result in very different environmental responses of GPP, suggesting additional examination of approaches and process representation is needed (Rogers et al., 2017). Additionally, our analysis shows that the amount of target carbon storage could be an important parameter for predicting carbon stocks and fluxes; however this process has minimal observational constraint globally. Future research efforts focused on better understanding carbon storage and allocation dynamics under different environmental stresses could substantially improve predictions of vegetation demography enabled LSMs, such as CLM4.5(ED).

3.6 Conclusion

Land surface components of ESM's have many parameters that could potentially affect the outcome of their simulations. Quantifying the contribution of model parameters to the overall output variance is critical for model diagnosis and improvement. Here, we used the Fourier amplitude sensitivity test (FAST) to conduct a high-dimensional global sensitivity analysis on the Community Land Model with Ecosystem Demography, CLM4.5(ED). We used an intermediate complexity of simulation: runs were sufficiently long to permit short-term physiological uncertainty to propagate into the long-term forest demographic structure, however we used only a single PFT for each run

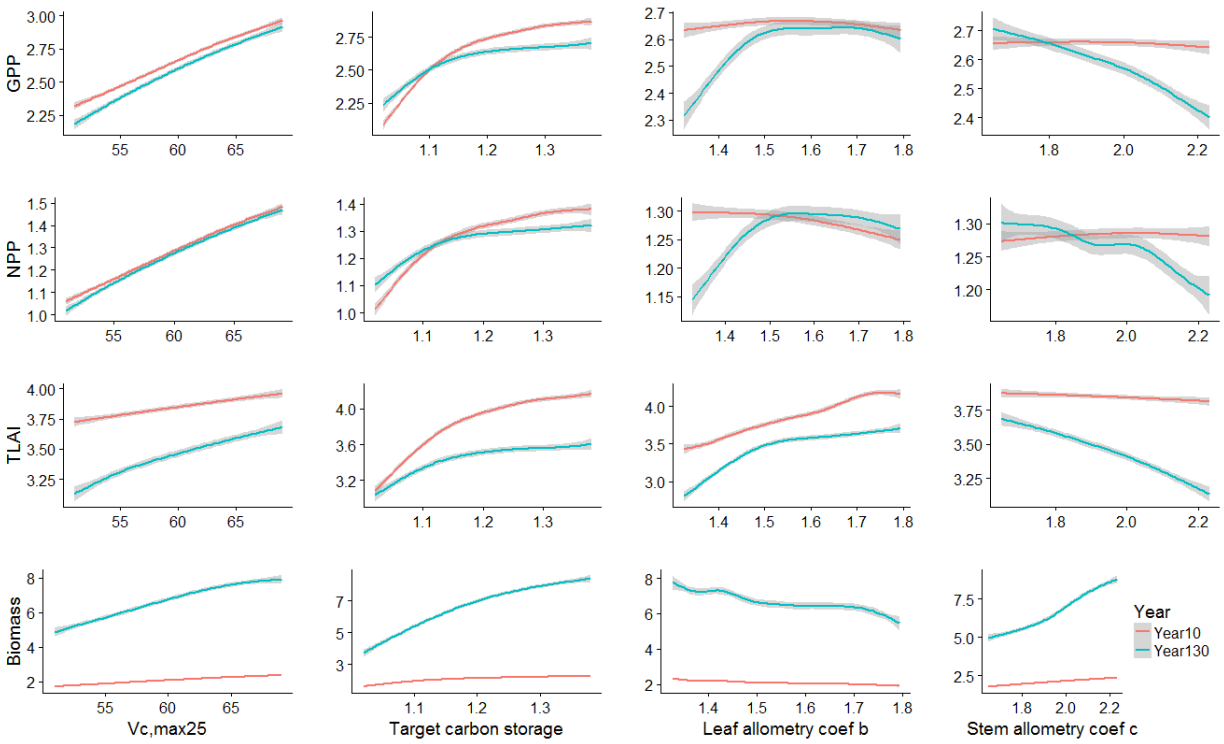

Figure 3.10: Relations between the most sensitive parameters ($V_{c,max25}$, storage carbon, leaf and stem allocation) to outputs of CLM (i.e. GPP, NPP, TLAI, and biomass) at years 10 and 130 of the simulations. Shown are the mean relations, with the 95 % confidence interval in grey envelopes. These figures show how an output will generally increase or decrease when a given parameter is changed. For example, the relation between the photosynthetic parameter ($V_{c,max25}$) and all the outputs were positive, as is the case with storage carbon parameter.

to eliminate from our analysis uncertainty due to competitive dynamics between different PFT strategies.

The most sensitive parameters are those related to photosynthetic processes, followed by survival and allometry parameters (namely $V_{c,max25}$, storage carbon, as a proxy for shade tolerance, and leaf and stem allometry parameters). Specifically, for the CLM4.5(ED) outputs representing carbon exchanges between the land and atmosphere, such as GPP and NPP, the most sensitive parameter was $V_{c,max25}$. For outputs representing the vegetation state of the forest, such as LAI and biomass, the leaf and stem allometry parameters, respectively, were the most sensitive parameters, along with the storage carbon parameter. The outputs of BA, mortality, and dDBH were mostly affected by storage carbon and stem allometry parameters, as well as $V_{c,max25}$. The importance of growth and survival parameters that we find here emphasizes the importance of how long-term ecosystem processes are represented in the governing ecosystem structure of this class of land model.

The FAST analysis provides a promising means of analyzing complex ESM components, and can be a powerful tool in understanding the necessarily high-dimensional representation of living systems within climate projection models. The results of the sensitivity analysis presented here can be utilized in future calibration efforts to reduce the complexity of the parameter-output response surface. We look forward to future diagnosis of CLM4.5(ED), and hope that results presented here can serve to inform the analysis of the structure of the model and its parameters as well as for future measurement campaigns.

Appendix

Appendix A: Allometry equations

The following equations are representative of the cohort based calculations made in CLM4.5(ED). Interested readers are referred to Fisher et al. (2015) for more information. The parameters used for the allometry equations include $dbh2h_{\mathrm{m}}$, $dbh2h_{\mathrm{c}}$, $dbh2bd_{\mathrm{a}}$, $dbh2bd_{\mathrm{b}}$, $dbh2bd_{\mathrm{c}}$, and $dbh2bd_{\mathrm{d}}$ (all are unitless variables). The equations represent processes such as converting changes in structural biomass to changes in height, i.e.:

$$\frac{\mathrm{d}BD}{\mathrm{d}h} = (dbh2bd_{\mathrm{a}})(dbh2bd_{\mathrm{b}})(hite^{dbh2bd_{\mathrm{b}}-1})(DBH^{dbh2bd_{\mathrm{c}}})(density_{\mathrm{wood}}^{dbh2bd_{\mathrm{d}}}) \tag{3.18}$$

where $\frac{\mathrm{d}bd}{\mathrm{d}h}$ is the rate of change of dead biomass (KgC) per unit change of height (m), $hite$ is the height of the tree (m), DBH is the diameter at breast height (cm), and $density_{\mathrm{wood}}$ is the wood density (g cm^{-3}). The equations also represent the processes that convert changes in structural biomass to changes in diameter, i.e.:

$$\frac{\mathrm{d}BD}{\mathrm{d}DBH} = (dbh2bd_{\mathrm{a}})(dbh2bd_{\mathrm{c}})(hite^{dbh2bd_{\mathrm{b}}})(DBH^{dbh2bd_{\mathrm{c}}-1})(density_{\mathrm{wood}}^{dbh2bd_{\mathrm{d}}}) \tag{3.19}$$

where $\frac{\mathrm{d}BD}{\mathrm{d}DBH}$ is the rate of change of dead biomass (KgC) with change in DBH (cm). As long as the DBH has not reached the maximum threshold yet (max_{dbh}). The rate of change of height (m) with change in DBH (cm) is calculated as follows:

$$\frac{\mathrm{d}H}{\mathrm{d}DBH} = 10^{dbh2h_{\mathrm{c}}}(dbh2h_{\mathrm{m}})(DBH^{dbh2bd_{\mathrm{m}}-1}); DBH < max_{\mathrm{DBH,pft}} \tag{3.20}$$

and then if the DBH reached the maximum, the changes in structural biomass to changes in diameter becomes:

$$\frac{\mathrm{d}BD}{\mathrm{d}DBH} = (dbh2bd_{\mathrm{a}})(dbh2bd_{\mathrm{b}})(DBH^{dbh2bd_{\mathrm{b}}-1})(DBH^{dbh2bd_{\mathrm{c}}})(density_{\mathrm{wood}}^{dbh2bd_{\mathrm{d}}}) \tag{3.21}$$

Appendix B: Temperature response curve

The parameters used for the temperature response curve equations include the equation to calculate the maximum carboxylkation rate, $V_{\mathrm{c,max25}}$, the maximum electron transport rate, J_{max}, and the Triose phosphate use (TPU) limited carboxylation rate, TPU (also all parameters here are unitless) (Fisher et al., 2015). The temperature response equations for $V_{\mathrm{c,max,z}}$, $J_{\mathrm{max,z}}$, and $TPU\mathrm{z}$ are:

$$V_{\mathrm{c,max,z}} = V_{\mathrm{c,max,25}}(e^{\frac{vcmaxha}{(0.001rgas)(t_{frz}+25)}})(1-\frac{t_{frz}+25}{t_{veg}})(\frac{vcmaxc}{1+e^{-vcmaxhd+(vcmaxse)(t_{veg})}}) \tag{3.22}$$

$$J_{\mathrm{max,z}} = J_{\mathrm{max,25}}(e^{\frac{jmaxha}{(0.001rgas)(t_{frz}+25)}})(1-\frac{t_{frz}+25}{t_{veg}})(\frac{jmaxc}{1+e^{-jmaxhd+(jmaxse)(t_{veg})}}) \tag{3.23}$$

$$TPU_{\mathrm{z}} = tpu25(e^{\frac{tpuha}{(0.001rgas)(t_{frz}+25)}})(1-\frac{t_{frz}+25}{t_{veg}})(\frac{tpuc}{1+e^{-tpuhd+(tpuse)(t_{veg})}}) \tag{3.24}$$

where t_{frz} is the freezing point of water in Kelvin (273.15 K).

Appendix C: Carbon storage in CLM4.5(ED)

Carbon storage, b_{store} (in kg C/cohort) plays a very important role in both the growth and mortality (Fisher et al., 2015). Specifically, CLM4.5(ED) assumes a target carbon storage determined by the multiplication of leaf biomass (b_{leaf}) and cushion factor (S_{cushion}; variable *cushion* in Table A3). At the specific time, the carbon balance for growth and storage is calculated as follows,

$$C = NPP - T_{md}f_{md,min} \tag{3.25}$$

where T_{md} is the maintenance respiration and $f_{md,min}$ is the minimum fraction of the maintenance demand (storage priority parameter in Table A1) that the plant must meet each time step, which represents a life-history-strategy decision concerning whether leaves should remain on in the case of low carbon uptake (a risky strategy) or not be replaced (a conservative strategy).

The fraction of the carbon balance for each cohort allocated to the carbon storage pool (f_{store}) will be determined by the follow equations:

$$f_{\text{store}} = e^{(-f_{\text{tstore}})^4} \tag{3.26}$$

where

$$f_{\text{tstore}} = \max\big(0, \frac{b_{\text{store}}}{S_{\text{cushion}} b_{\text{leaf}}}\big) \tag{3.27}$$

Thus, the target carbon storage parameter, S_{cushion}, can affect carbon allocations. Specifically, a higher value of S_{cushion} will lead to a higher allocation of carbon to storage and thus lower allocation to growth at the specific time step.

Carbon storage also plays an important role for the mortality. Specifically, carbon starvation mortality (M_{cs}) is calculated as follows:

$$M_{\text{cs}} = S_{\text{m}} \max\big(0, 1 - \frac{b_{\text{store}}}{b_{\text{leaf}}}\big) \tag{3.28}$$

where S_{m} is the stress mortality factor (i.e., *stress_mort* in Table A.3).

Appendix D: Growth relations to stem allometry

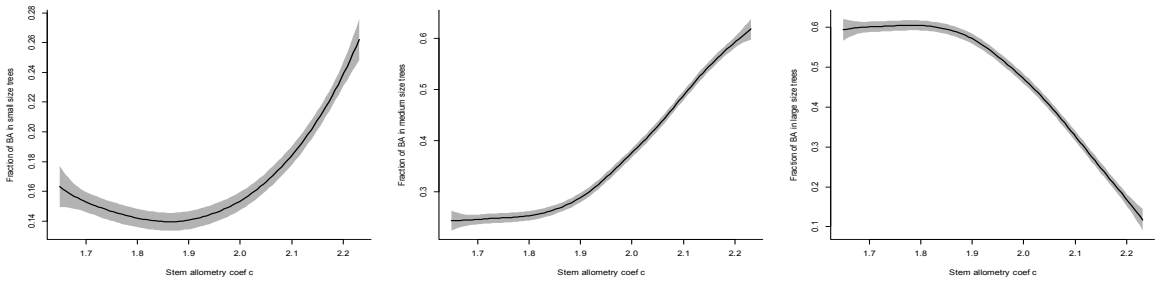

Figure 3.11: Impact of stem allometry on basal area (BA) distribution across trees of different sizes. The figure shows results for the simulations years 100-130, and the 95% uncertainty of these relations.

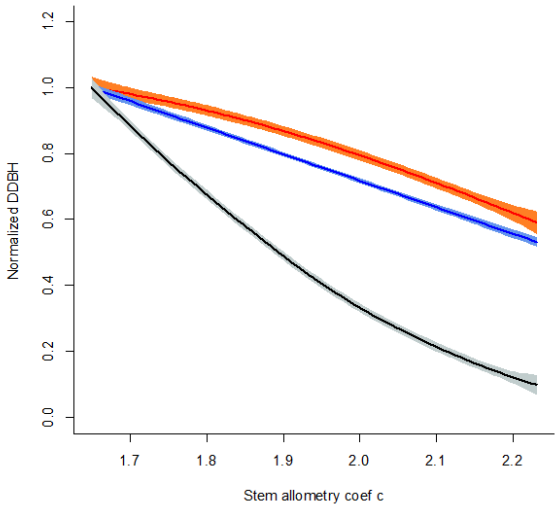

Figure 3.12: Impact of stem allometry on tree diameter growth (dDBH). The values in the figures are normalized by the expected value at lowest stem allometry coefficient c. The red curve are used for small trees, blue for medium trees, and black for large trees. The figure shows results for the simulations years 100-130, and the 95% uncertainty of these relations.

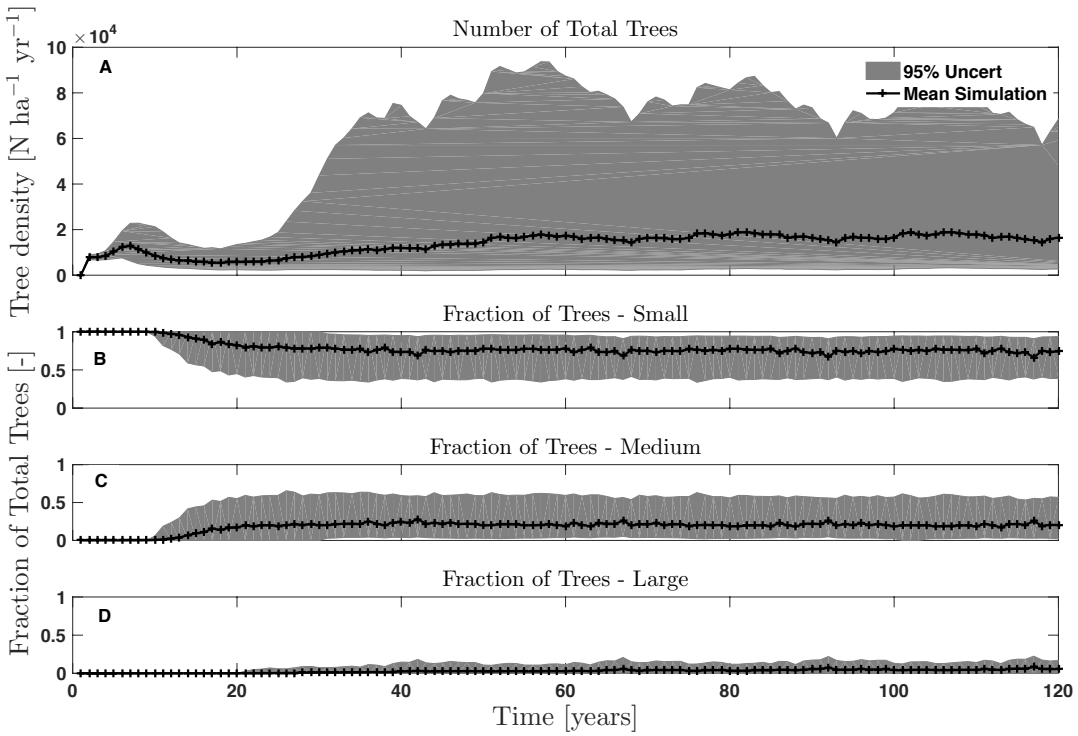

Figure 3.13: Outputs from CLM4.5(ED) for tree density, or the number of trees per area (NPLANT). Figures show the outputs for the various tree sizes considered, including small (*diameter* < 10 cm), medium (10 cm < *diameter* < 50 cm), and large trees (*diameter* > 50 cm). Shown are the mean simulation (black line) with 95 % confidence intervals.

Appendix E: Mortality Outputs

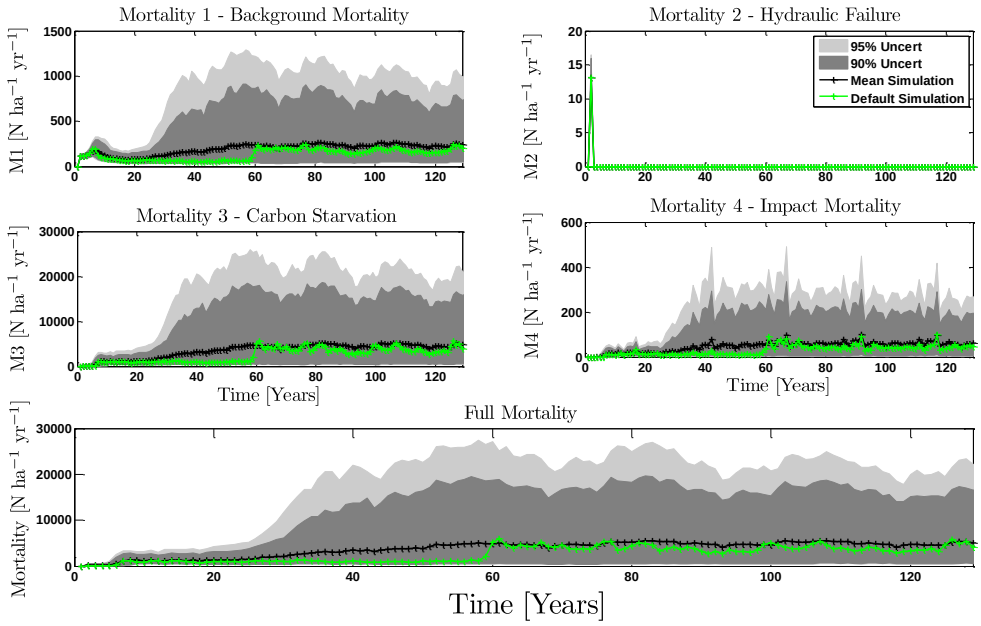

Figure 3.14: Mortality outputs from CLM4.5(ED), including the mechanisms of M1 - Background Mortality, M2 - Hydraulic Failure, M3 - Carbon Starvation, and M4 - Impact Mortality. The bottom panel shows the total mortality, which is the sum of M1-M4. An additional possibility for mortality in CLM4.5(ED) is from fire disturbances, however the fire sub-routine of the model was turned off since the study site was in the Amazon. Shown are 95 % (light grey) and 90 % (dark grey) confidence intervals, along with the mean simulation (black lines) and the simulation using the default parameter set (green lines).

Appendix F: CLM4.5(ED) fractional outputs of small, medium, and large trees

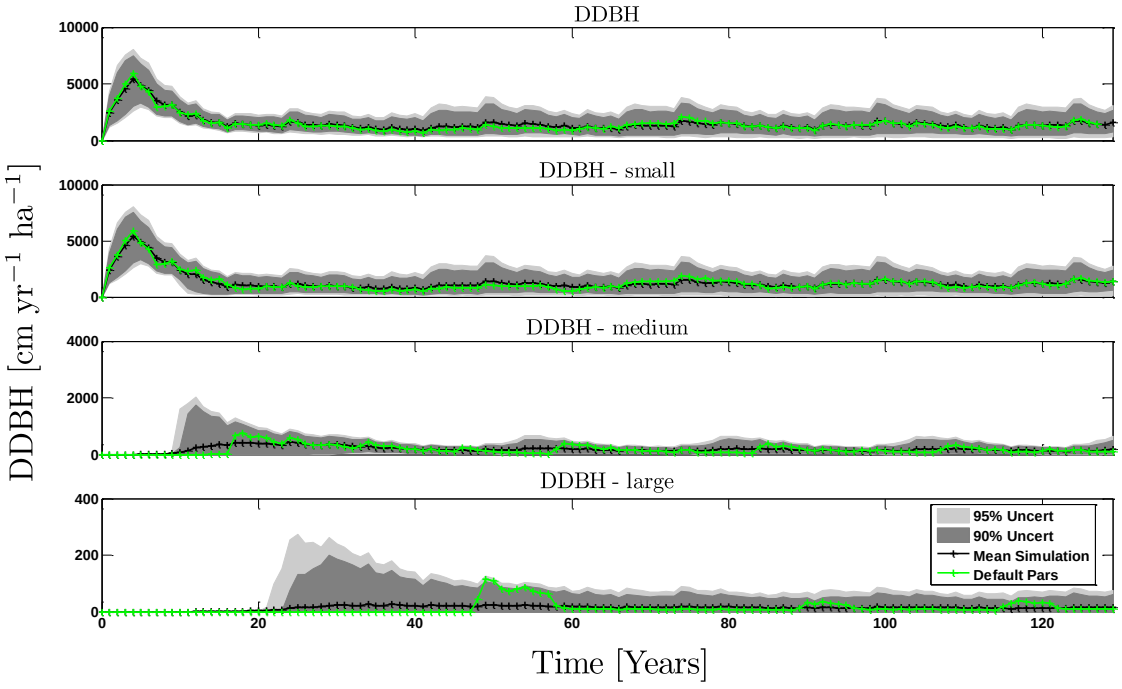

Figure 3.15: Outputs from CLM4.5(ED) for the change in diameter at breast height (dDBH). Figures show the fraction of total dDBH that is distributed to each of the tree sizes considered, including small (*diameter* $<$ 10 cm), medium (10 cm $<$ *diameter* $<$ 50 cm), and large trees (*diameter* $>$ 50 cm). Shown are the mean simulation (black line) with 95 % confidence intervals.

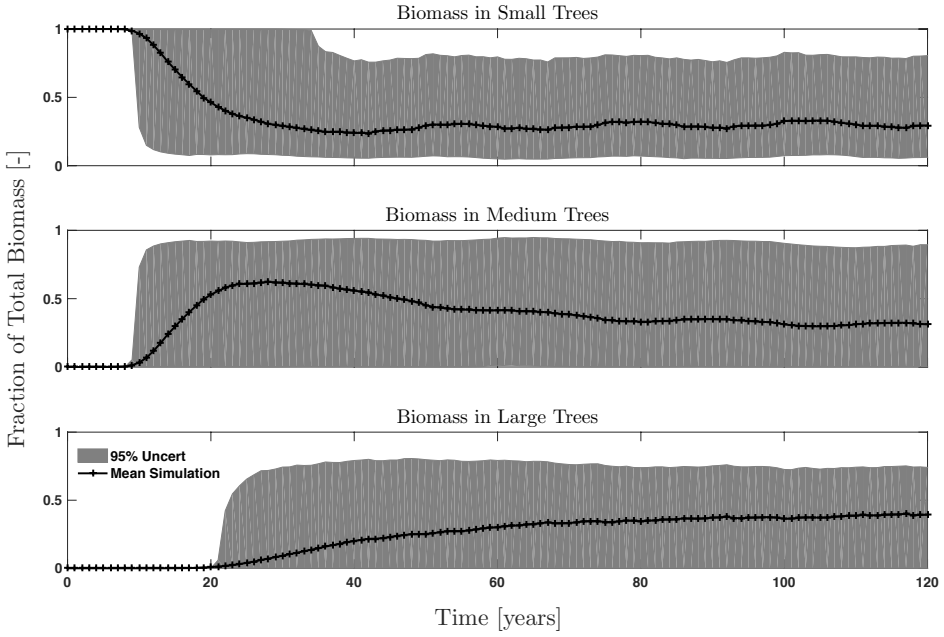

Figure 3.16: Biomass distribution for the various cohorts considered. Each figure shows the fraction of total biomass that is in each cohort.

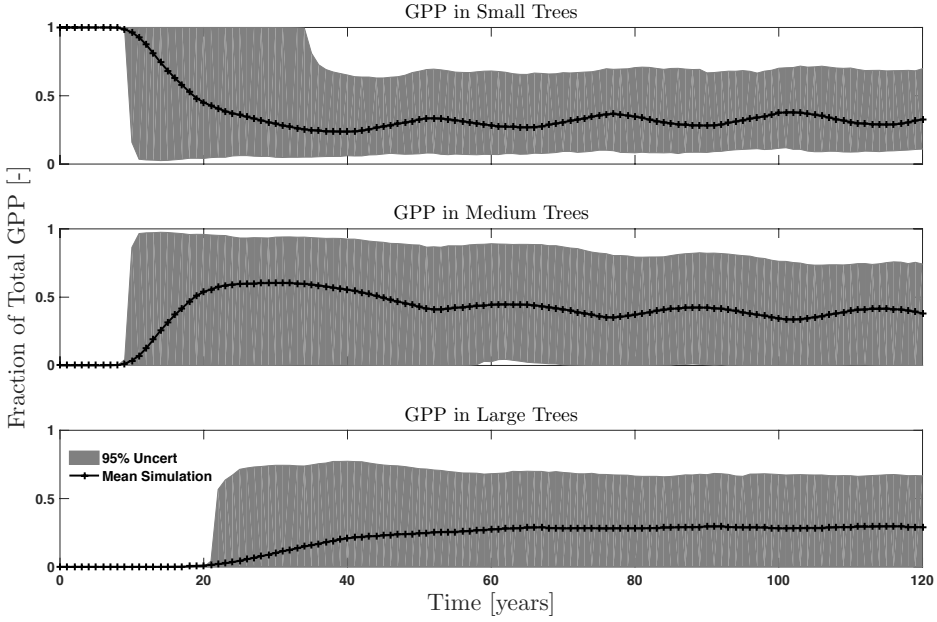

Figure 3.17: GPP distribution for the various cohorts considered. Each figure shows the fraction of total GPP that is in each cohort.

Appendix G: 2^{nd} order sensitivities

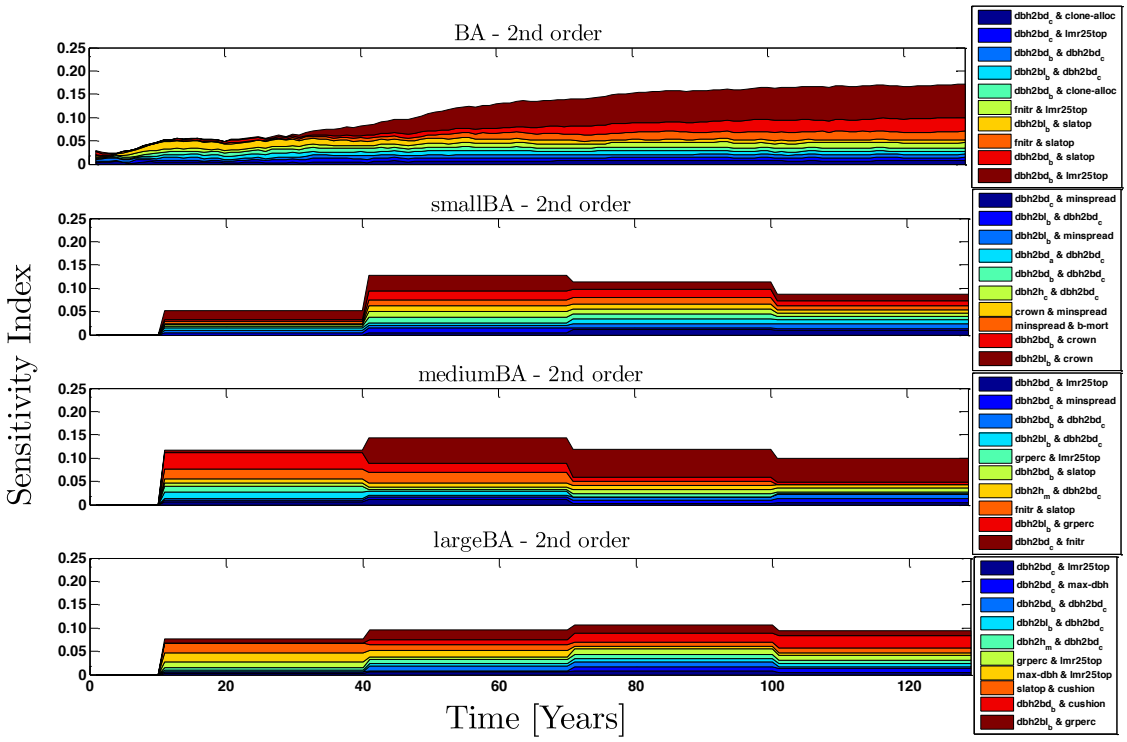

Figure 3.18: 2^{nd} order sensitivity index of the model parameters for the Basal Area (BA) outputs from CLM4.5(ED).

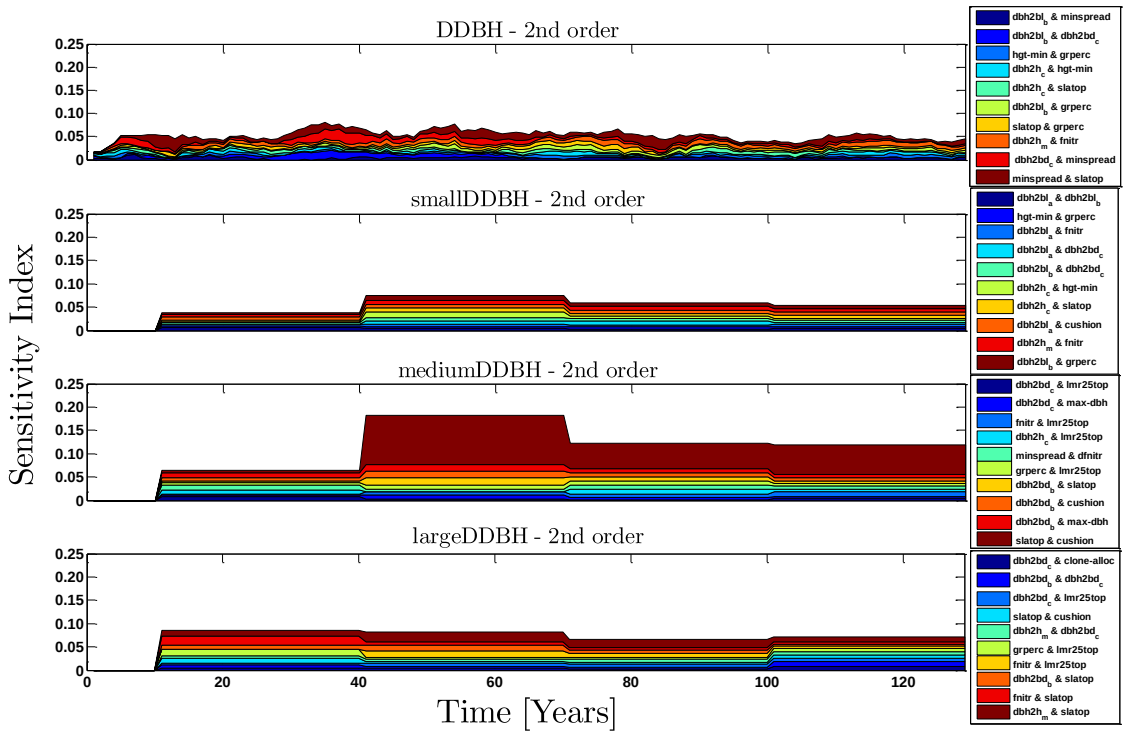

Figure 3.19: 2^{nd} order sensitivity index of the model parameters for the change in diameter at breast height (dDBH) outputs from CLM4.5(ED).

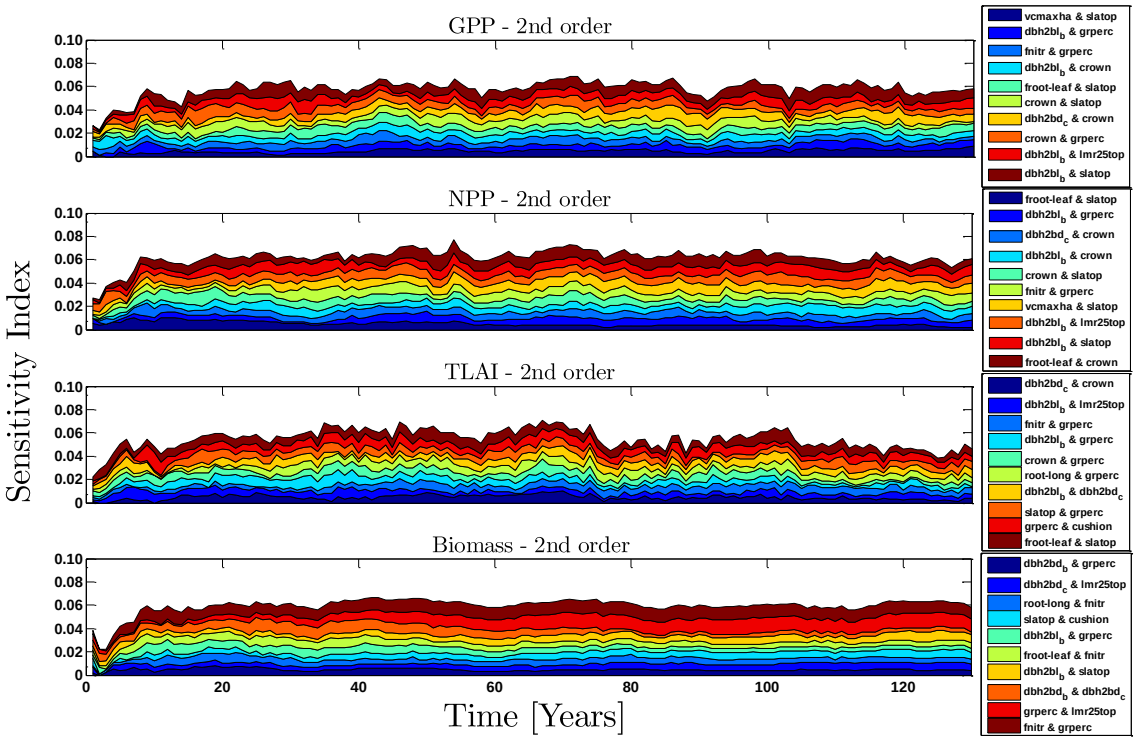

Figure 3.20: 2nd order sensitivity index of the model parameters for GPP, NPP, TLAI, and biomass outputs from CLM4.5(ED).

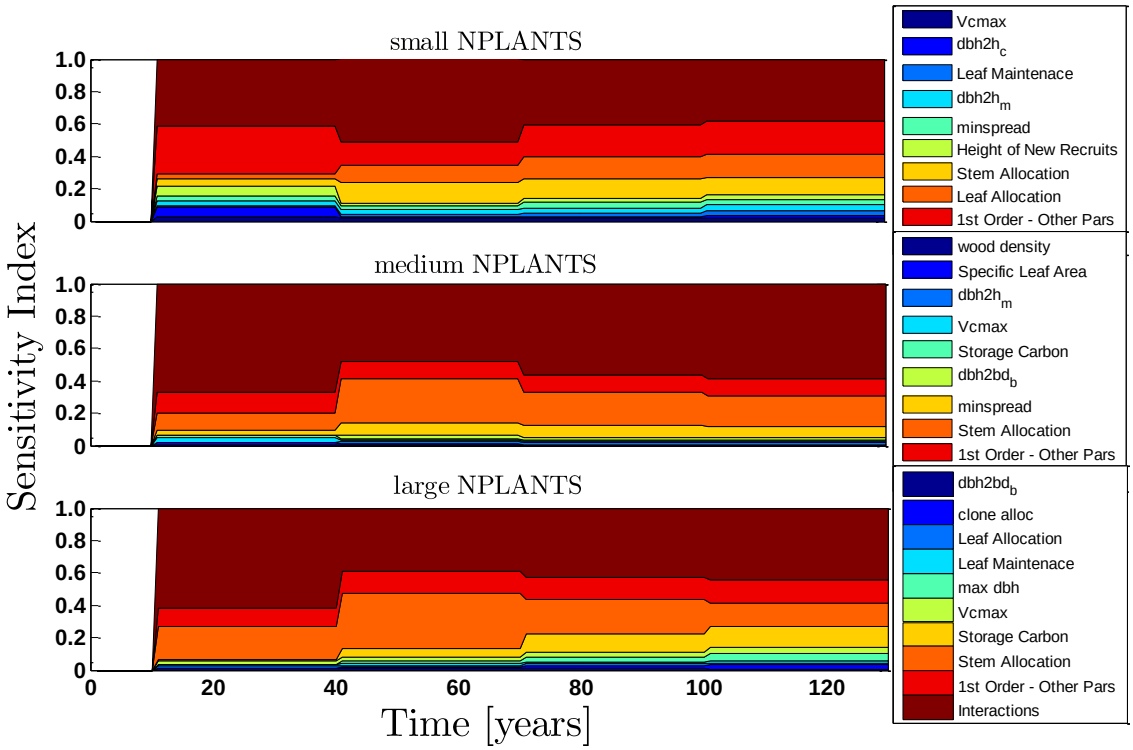

Figure 3.21: Sensitivity index of the model parameters (1st order) for the outputs of plant density (NPLANT). Also shown are sensitivities of the remaining parameters ('1st order - Other Pars') as well as the sensitivity of parameter interactions ('2nd order - All Pars').

3.7 LSM case study 2: Improved Global Nitrogen Allocation Model

The second case study in this chapter is dedicated to the improved prediction of terrestrial nitrogen and carbon fluxes. Xu et al. (2012) and other recently developed works have promoted the development of a mechanistic leaf nitrogen allocation model at a global scale. We present a dynamic carbon-nitrogen model that will incorporate recent advances in nitrogen modeling (see Figure 3.22) and uses the DREAM algorithm to rigorously calibrate and evaluate the developed model against observations. The observed data is unique, and includes soil fertilization and free air CO2 enrichment (FACE) observations across a range of different forest types from across the globe. The goal of the study is to provide reliable estimates of the relationships that leaf nitrogen allocation takes on within a leaf on a global scale, which is crucial for prediction of land carbon sinks in Earth System Models (ESMs). Currently, various forms of ESM's exist, and are today's critical tools for making climate projections. Throughout this thesis, the ESM that will be used is the Community Land Model with Ecosystem Demography option, or CLM4.5(ED), developed by the National Center of Atmospheric Research (NCAR), LANL and many other universities for prediction of various processes on the land surface and in the atmosphere.

A dynamic carbon-nitrogen allocation model within leaves was calibrated with the DREAM algorithm. The parameters of the model include J_{maxb_0} (baseline proportion of nitrogen allocated for electron transport rate), J_{maxb_1} (electron transport rate response to light), t_{c,j_0} (ratio of rubisco limited rate to light limited rate), and H (electron transport rate parameter related to relative humidity). The values of the parameters were calibrated under two conditions, $TRF1$ and $TRF2$. $TRF1$ considers temperature acclimation, while $TRF2$ does not. The formulation of $TRF2$ is same as $TRF1$ except that in $TRF2$, an entropy term is fixed across the entire data set.

The estimated mean and standard deviation of the parameter values are shown in Table 3.1.

Table 3.1 Mean values of parameters obtained by using DREAM when $TRF1$ and $TRF2$ were used. The parameters include; J_{maxb_0} (unitless) - baseline proportion of nitrogen allocated for electron transport rate, J_{maxb_1} (unitless) - electron transport rate response to light, t_{c,j_0} (unitless) ratio of rubisco limited rate to light limited rate, and H (unitless) - electron transport rate parameter related to relative humidity. The standard deviations are shown in the parentheses.

Statistics	J_{maxb_0}	J_{maxb_1}	t_{c,j_0}	H
$TRF1$	0.0311(0.0004)	0.1745(0.00025)	0.8054(0.0015)	6.0999(0.2416)
$TRF2$	0.0322(0.0002)	0.1695(0.0006)	0.7760(0.0031)	5.7139(0.0354)

The estimation of the model parameter values on a global scale is necessary for improved accuracy in climate projections, because a large uncertainty exists in the current simulation of nitrogen related processes (e.g. photosynthesis and soil carbon storage response to nitrogen addition), which substantially affects the reliability of predicted terrestrial carbon fluxes. Thus, the estimates provided will be implemented into larger scale models tested against observed global data of variables that represent processes occurring in the biosphere, such as Net Primary Production (NPP). This work with the collaborators from LANL assesses energy impacts on the global carbon cycle and future climates.

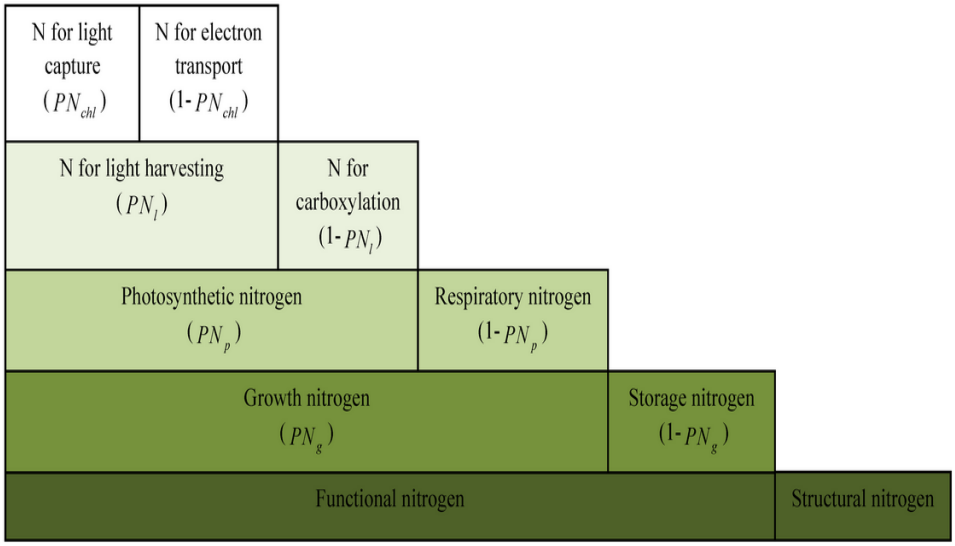

Figure 3.22: Schematic of the hierarchical plant functional nitrogen allocation for a leaf layer of a tree. This is the general view of the model structure of mechanistic leaf nitrogen allocation model. Figure drafted from Xu et al. (2012).

3.8 LSM case study 3: Determining size dependence on tree survival

Survival rates of large trees determine forest biomass dynamics. Forecasting the fate of the terrestrial carbon sink depends on correctly incorporating size-dependent tree survival. We tested patterns of tree survival across the tropics using data from 1781 species and two million individuals to assess whether life-history survival strategies can be used to characterize complex and diverse tropical forests. We resolved four size-dependent survival modes based on the relation of mortality to tree size that explain life-history strategies shaping the terrestrial forest ecosystem carbon-cycle budget. Frequently collected traits, such as wood density, leaf mass per area and seed mass were not predictive of survival mode, suggesting a disconnect between traits and survival strategy. Mean annual temperature and cumulative water deficit predicted the relative biomass of survival modes, indicating important links between evolutionary strategies, climate and carbon cycling. We were also able to use survival modes in simulations to predict biomass change over time. Our results reveal common demographic strategies useful for defining plant functional types for global dynamic vegetation models and provide an avenue for increased understanding of forest ecosystem dynamics.

Tropical forests are currently estimated to store 500-1000 Pg of C in biomass and soils. Whether intact tropical forests will be sinks or sources of carbon in the future remains a critical question, which will fundamentally depend on how forest species respond to climate change, such as storm frequency and intensity, warming and drought. Forest carbon volume depends exponentially on the annual rate of tree survival, and tree survival rates in turn depend on climate. In most forests, survival strategies range from long-lived species that retain carbon for centuries to short-live species that die within decades. Changes in the forest composition, due to differential survival responses of species to climate variation or extreme episodic events (i.e. droughts and storms), may can cause large and rapid changes in the terrestrial carbon balance that can potentially persist for centuries. Such effects on tree survival are distinct but potentially more important than climate-driven changes

in forest productivity through tree growth, which has a relatively constrained and slower influence on terrestrial carbon dynamics.

By examining the survival of more than two million trees representing 1781 species in 14 large forest plots across the tropics, we found four modes of size-dependent survival that are common across diverse forests and contain information about how allocation differs over time and across species. To understand the ecological significance of these modes we : 1) investigated how survival modes contribute to carbon fluxes through differences in growth rates and biomass turnover; 2) asked if the modes of survival that emerge from the demographic data are related to the commonly collected plant traits of wood density, leaf mass per area and seed mass; 3) compared our predictions of size-dependent survival with DGVM output to identify where improvements can be made; 4) tested whether the relative abundance of these survival modes relate to climate variables and 5) tested the predictions of our model results against the observed biomass at each site through time.

3.8.1 Results: global size-dependent tropical tree survival

Hierarchical cluster analysis of size-dependent survival of species occurring across 14 pan-tropical large area forest dynamics plots (ranging from 2 to 52 ha each with 371 ha in total) indicates four size-dependent survival modes that characterize tropical tree and shrub survival (Fig. 1). Understory species are characterized by their small maximum diameters, with an across-site mean 99th percentile diameter of 9.8 cm.

Transient species are distinguished by their very 108 low overall survival with an across-site mean maximum-survival rate of 78% yr^{-1} and an across-site mean 99th percentile diameter of 14.3 cm. There are two groups of large stature tree species. Canopy species are the group with intermediate maximum size, across-site mean 99th percentile diameter of 27.8 cm and lower small-diameter survival rates compared to Large Canopy species which have larger maximum diameter, across-site mean 99th percentile diameter of 68.4 cm and relatively higher survival at smaller diameters.

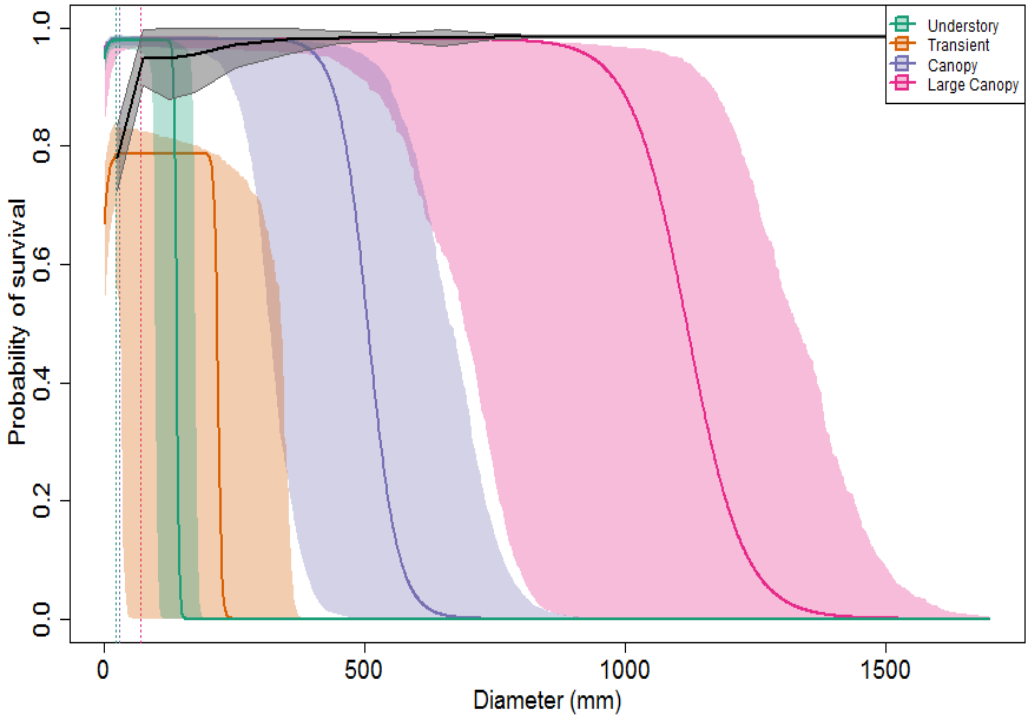

Figure 3.23: The CLM4.5(ED) model (grey envelope) over predicted large tree survival when compared with our size-based survival modes (colored envelopes). CLM4.5(ED) mortality function includes terms for carbon starvation, hydraulic failure, crushing of small trees by large tree mortality, a background rate of mortality and fire (though fire was not included in the simulations used here). We found that the model predicted higher survival in the largest size classes (greater than 70cm DBH) than our statistical model. The causes of mortality in large trees are likely to be more complex than small trees which are more vulnerable to asymmetric competition with larger trees for light and other resources.

Our analysis has an abundance threshold of 200 individuals; species with lower abundance are Unclassified, and it is possible that some of them display other survival modes that were too rare to describe statistically. To test the robustness of our survival modes, we calculated the Jaccard similarity index for all clusters which were well above the 0.75 threshold indicating stable clustering for our size-dependent survival modes based on a 1000 bootstrap simulations (Extended Data Table 2).

Our model predictions fit observed forest biomass well (Fig. 36), which is a challenge for DGVMs which rely on physiological models to drive survival outcomes. We tested how a DGVM's output

maps to survival predicted by global survival modes. The CLM4.5(ED) model over predicted large tree survival when compared with our size-based survival modes. CLM4.5(ED) mortality function includes terms for carbon starvation, hydraulic failure, crushing of small trees by large tree mortality, a background rate of mortality and fire (though fire was not included in the simulations used here). We found that the model predicted higher survival in the largest size classes (greater than 70cm DBH) than our statistical model. The causes of mortality in large trees are likely to be more complex than small trees which are more vulnerable to asymmetric competition with larger trees for light and other resources. Large trees are likely to have accumulated wounds, lianas and rot that are not currently simulated in DGVMs, though there have been calls for modeling insects and diseases in both crops and forests. Including these types of stressors could be considered in future DGVMs. Alternately, exogenous disturbance, e.g. wind, maybe a driver of large tree mortality that is not currently captured by DGVMs. A path forward may be to parameterize a size-dependent mortality term that allows for more accurate estimation of large individual survival.

3.9 Acknowledgments

This work was supported by a United States Department of Energy (US DOE) Office of Science Next Generation Ecosystem Experiment at Tropics (NGEE-T) project and the Graduate Student Researcher (SCGSR) fellowship. The first and last author are also supported by funding from the UC-Lab Fees Research Program Award 237285. Model simulations were made possible thanks to the Conejo Super Computing system at the Los Alamos National Laboratory (LANL). The FAST methodology described herein is available upon request from the second author (cxu@lanl.gov). The model simulations from all 5000 parameter combinations as well as simulation of the default parameter set are available upon request from the first author (emassoud@uci.edu).

Chapter 4

Calibration of a Multi-Output Ecohydrologic Model

by Elias C. Massoud, Brad Chrisrofferson, AJ Purdy, Chonggang Xu, Jasper A. Vrugt

References:

Massoud, E. C., Christofferson B.O., Purdy, A.J., Xu, C.: Bayesian inference of hydraulic properties in and around a Douglas White Fir using a process based ecohydrologic model. Environmental Modeling & Software, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

Trees and forests play a key role in controlling the water and energy balance at the land-air surface. One case study in this thesis involves the simulation of water flux through a tree, using a model named the Soil-Tree-Atmosphere Continuum (STAC) model. The STAC model follows fundamental biophysical principles, in which soil water is taken up by the roots and moves through the water-conducting vessels or xylem along a water potential gradient into the canopy, where it transpires into the atmosphere through leaf stomata.

Historically, models that attempt to mimic this process have been based on the electrical circuit analog, with resistance-capacitance (RC) models to characterize xyleme resistance to water flow and water storage. Now, physically-based nonlinear modeling framework, numerical flow models discretize the modeled domain into small elements and use finite-difference or finite-element methods to solve the flow equations. However, few models use the integrated soil-tree continuum approach, coupling the soil with the tree domain, simulating the soil, roots and tree trunk as a continuum, in which water flow is driven by water potential gradients along the coupled STAC with spatially distributed root water uptake and canopy transpiration sink terms. Water flow through the coupled system is described using the Richards' equation type of representation with both the soil and tree conducting domains modeled as a porous medium, defined by nonlinear soil and tree water relationships. Figure 4.1 shows the processes involved for the STAC model to simulate water flux through such a system.

4.1 Manuscript Title - Bayesian inference of hydraulic properties in and around a Douglas White Fir using a process based ecohydrologic model

Abstract: Parameters used in land surface and climate models are typically inferred from empirical relationships or extracted from arbitrary data sets that may not represent local variation or capture seasonal or ontogenetic changes. This has caused major deviations in the representation of

vegetation in this suite of models and new research has been prompted to identify parameter sets that describe vegetation properties consistently. We present here a parameter estimation study of the Soil-Tree-Atmosphere Continuum (STAC) model, a process based model that simulates water flow through an individual tree and its surrounding root zone. Parameters are fitted to observations made for a Douglas White Fir (Abies concolor) in the Sierra Nevada, California. Bayesian inference is applied with a likelihood function that explicitly considers first-order temporal correlation of the residuals. We highlight the estimation of properties such as the tree's root distribution, tolerance to drought, and hydraulic conductivity and retention functions.

4.2 Introduction

Ecohydrology is a field of study focusing on the hydrologic mechanisms that govern and explain ecologic patterns and processes (Jackson et al., 2009; Rodriguez-Iturbe, 2000). Ecohydrologic system properties depend on many interrelated links between climate, soil, and vegetation (Rodriguez-Iturbe et al., 2001). One part of this cycle includes the role that climate and soil have in controlling vegetation dynamics (Boyer, 1982; Jones, 2013; Kramer and Boyer, 1995; Lange et al., 1976; Larcher, 2003), and another part of the cycle is the important control that vegetation exerts on the water and energy balance (Kutzbach et al., 1996; Schlesinger et al., 1990; Zeng et al., 1999). Many important issues depend on the quantitative understanding of these vegetation dynamics, including environmental preservation, proper management of resources, as well as improved model representation in simulations of ecohydrologic systems (Noy-Meir, 1973; Scholes and Walker, 2004; Shmida et al., 1986; Xu et al., 2013).

Generally, the role that trees play on the overall water cycle in regards to water storage, residence time, and vegetation tolerance to drought are hardly represented in Land Surface Models (LSMs). This lack of understanding of water dynamics in vegetation is a potential cause for large discrepancies seen between various model simulations that run into the next century (McDowell

et al., 2015; Sitch et al., 2008). Much data has been collected to fill some of the gaps in our understanding of tree water relations (Zweifel et al., 2007), and can usually inform ecohydrologic models to enhance their fidelity and guide their development (Christoffersen et al., 2016). These models can then be diagnosed to help further understand the same ecohydrologic relationships that create them, providing information on where additional data is needed to re-iniform and thus re-develop the model. Future development will enhance the ability of LSMs to predict individual tree processes such as drought tolerance or storage capacity and thus enhance the overall representation of vegetation's effect on the global water cycle (Medlyn et al., 2016).

We present here a numerical model that simulates water storage and transport through a tree and its root zone, coined the Soil-Tree-Atmosphere Continuum (STAC) model. The STAC model is structured with an axi-symmetrical 2D (or quasi-3D) representation of water flow through the combined soil-tree domain. The model uses Richards' equation and Mualem-van Genuchten hydraulic functions from Van Genuchten (1980) and Mualem (1976) to characterize water storage and movement. This representation is important, as soil water flow models usually do not take into account water storage in trees. In the STAC model, root water uptake is not assumed to be equivalent to transpiration, which conceptually allows for change in the tree's water storage. Our goal here is to accurately simulate the dynamics of water flow in a mature Douglas White Fir and its root zone. In doing so, we aim to properly infer model parameters that dictate how much water can be stored (capacity) or can flow (conductance) in the tree and its surrounding soil domain.

Similar studies have been a topic of interest in the ecohydrologic community for decades (Sala and Lauenroth, 1982; Tyree, 1988). Previous works have highlighted the differences between saturated and unsaturated flow (Aumann and Ford, 2002), improved on the representation of branch junctions (Schulte and Brooks, 2003), linked tree sap flow to stem growth (Steppe et al., 2006), modeled both xylem as well as phloem water fluxes (Hölttä et al., 2009; Lacointe and Minchin, 2008), improved prediction of xylem ABA concentrations by proper accounting of sap flow (Dodd et al., 2008),

improved understanding of the affect of root system architecture for the enhancement of drought tolerance (Draye et al., 2010), accounted for hydraulic redistribution between different soil parts via plant root systems (David et al., 2013; Prieto et al., 2012), provided a computationally efficient 1-D alternative to 3-D models that includes a xylem flow model (Janott et al., 2009), among others. The purpose of this study is to investigate mechanisms such as water uptake and transpiration by a tree, spatial and temporal dynamics of soil moisture, retention and conductivity of the soil and xylem, and soil and tree water storages as functions of time.

We utilize Bayesian inference to obtain optimal parameters that allow model simulations to fit closely with observations made for a mature Douglas White Fir (Abies concolor) in the Kings River Experimental Watershed (KREW) in the Sierra Nevada, California. Applying Bayes' law also allows for estimation of the underlying parameter uncertainty (Katz, 2002). This is important since various sources of error can introduce uncertainty, such as errors from calibration data, model inputs, or model structural errors. Numerical implementation of the Bayesian paradigm requires the user to specify a prior parameter distribution as well as a likelihood function. The prior distribution should encode all the subjective knowledge about the parameters before collection of the data, whereas the likelihood function summarizes, in a probabilistic sense, the compatibility of the observed data to the simulated model outputs (Vrugt and Massoud, 2017). Likelihood functions play a key role in statistical inference, and here we utilize a specially designed likelihood function that can combine various data streams from several processes being considered.

It is generally assumed that if only one data set is used for the parameter estimation, the parameter values will be fitted to that specific process too closely. However, by considering various processes during the parameter estimation, the parameter search will be balanced by each data set and an overall more realistic representation of the system properties can be achieved (Medlyn et al., 2015). To this end, we use the likelihood function defined in Schoups and Vrugt (2010) and utilize multiple data streams in the estimation of the STAC model parameters. We show that this

does provide an overall accurate estimation of the soil and tree water properties when compared to observed data.

4.3 Material and methods

4.3.1 STAC Model

The Soil-Tree-Atmosphere Continuum (STAC) Model is a physically-based nonlinear modeling framework (Bohrer et al., 2005; Chuang et al., 2006; Kumagai, 2001; Mirfenderesgi et al., 2016), as typical for the simulation of water flow in unsaturated media (Rings et al., 2013; Siqueira et al., 2008). The STAC model discretizes the system domain and couples the soil with the tree domain, simulating the soil, roots and tree trunk as a continuum. Water flow is driven by water potential gradients along the coupled system (Bittner et al., 2012) with spatially distributed root water uptake and canopy transpiration sink terms. The STAC model utilizes the HYDRUS model (Simuunek et al., 2008), where water flow through the soil and the tree root system and stem is driven by the evaporative demand and soil-available water, leading to a gradient in soil and xylem water potentials along the STAC. We approximate both the soil and plant conducting tissues by a porous medium, with conductive and capacitive properties that are a function of water potential.

4.3.2 Domain and Boundary Conditions

HYDRUS allows the estimation of water potential, volumetric water content, and water flux density across the coupled soil-tree domain. Both the soil and tree trunk are modeled as axial-symmetrical, represented by a rectangular domain (Figure 4.2). The simulated soil domain extends to 5 m outwards; three soil layers characterize the top 2.5 m in the unsaturated soil, and the bottom 2.5 m interval represents the weathered low conductivity saprolite that can store water but is inaccessible to tree roots.

The lower boundary of the soil at the 5 m depth was described by a seepage boundary, allowing water to leave the soil domain when saturated, and allowing for both upwards and downwards flow across the whole soil domain. The upper boundary of the soil domain consists of measured values of rainfall and evapotranspiration. The lower boundary condition of the tree trunk is root water uptake from the soil domain, and the upper boundary of the tree is atmospheric demand of potential evapotranspiration. The 10 cm radius of the tree trunk was chosen so that the domain volume is approximately equal to that of the sapwood of the tree.

We use observations of soil moisture, soil water retention curves, and assume hydraulic equilibrium to initiate water potential distribution across the domain. For the soil, we converted 24 elements of soil moisture data collected on July 15, 2009, to soil water matric potential values using the laboratory-measured retention curves. Then, a 2^{nd} order polynomial interpolation scheme was applied to estimate the soil water potential across the measured soil domain, assuming hydraulic equilibrium at the domain boundaries. This completed the necessary initial and boundary conditions of the domain for the model simulations.

4.3.3 Unifying Equations

To set up the model simulations, we use the finite element HYDRUS software (Simuunek et al., 2008), which can solve unsaturated water flow across the soil-tree domain using the Richards' Equation (Richards, 1931) in a discretized system of linear equations (Equations 4.1 and 4.2). The flow in the soil domain is presented here in its axisymmetrical, two-dimensional, isotropic form:

$$\frac{\partial \theta_{\text{soil}}}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rK_{\text{r}}(h)\frac{\partial h}{\partial r}\right) + \frac{\partial}{\partial z}\left(K_{\text{z}}(h)\frac{\partial h}{\partial z}\right) - \frac{\partial K_{\text{z}}(h)}{\partial z} - W_{\text{soil}}(h, r, z) \tag{4.1}$$

where θ_{soil} (L^3L^{-3}) is the volumetric soil water content, $K(h)$ (LT^{-1}) defines the unsaturated hydraulic conductivity function (further denoted by either r- for radial direction or z- for vertical direction), h (L) is the soil water pressure head, r and z are the lateral and vertical coordinates

(positive downwards) of the soil domain respectively, t (T) is time, and W_{soil} ($L^3L^{-3}T^{-1}$) defines a sink/source term that quantifies spatially distributed root water uptake from the soil. Both K and W_{soil} are functions of θ and/or h. The subscripts r- and z- allow for the possibility of soil anisotropy, i.e., to simulate water flow with the unsaturated hydraulic conductivity function being different for the r- and z- directions.

The set up of Richards' equation for the tree domain to represent flow through the canopy is similar to that of the soil domain in Equation 4.1, but in one-dimensional form. This equation is derivable directly from Equation 4.1 by reducing to one dimension, z- only. Thus, the axi-symetrical flow through the tree canopy is represented by:

$$\frac{\partial \theta_{\text{tree}}}{\partial t} = \frac{\partial}{\partial z}\left(K_{\text{z}}(h)\frac{\partial h}{\partial z}\right) - \frac{\partial K_{\text{z}}(h)}{\partial z} - W_{\text{tree}}(h,z) \tag{4.2}$$

where θ_{tree} (L^3L^{-3}) is the volumetric tree water content, $K(h)$ (LT^{-1}) defines the unsaturated hydraulic conductivity function (further denoted by z- for vertical direction), h (L) is the tree water pressure head, z is the vertical coordinate of the tree domain (positive downwards), t (T) is time, and W_{tree} defines a sink term ($L^3L^{-3}T^{-1}$) that quantifies spatially distributed canopy transpiration.

For solution of Equations 4.1 and 4.2, unsaturated hydraulic conductivity and the water retention functions must be defined for both the soil and tree conducting matrix. The unsaturated hydraulic conductivity function (Equation 4.3) defines the relationship between the moisture content and the corresponding hydraulic conductivity of the domain, and the retention function (Equation 4.4) characterizes the ability of the domain to retain water. We define these functions using the relationships of Van Genuchten (1980) and Mualem (1976), where

$$K(h) = K_{\text{s}}\sqrt{S_{\text{eff}}}\left(1-\left(1-S_{\text{eff}}{}^{\frac{1}{m}}\right)^m\right)^2 \tag{4.3}$$

and

$$S_{\text{eff}}(h) = \frac{\theta-\theta_{\text{r}}}{\theta_{\text{s}}-\theta_{\text{r}}} = (1+|\alpha h|^n)^{-m} \tag{4.4}$$

in which $K(h)$ represents the hydraulic conductivity, and the degree of effective saturation, $S_{\text{eff}}(h)$, represents the retention function. For these equations, θ_{s} denotes the saturated water content at $h = 0$ (L^3L^{-3}), θ_{r} is the residual water content (L^3L^{-3}), α is a scale parameter inversely proportional to mean pore diameter (L^{-1}), n ($m = 1 - 1/n$) is a shape parameter of the soil water characteristic, and K_{s} (LT^{-1}) is defined as the conductivity at saturated conditions, or when $\theta = \theta_{\text{s}}$.

4.3.4 Root Water Uptake Model

The actual root water uptake term in Equation 4.1 is computed from:

$$W_{\text{soil}}(h, r, z) = \gamma(h)\,\beta(r, z)\,\pi r_{\text{m}}^2 ET_{\text{p}} \tag{4.5}$$

with $W_{\text{soil}}(h, r, z)$ representing actual water uptake of roots from the soil ($\text{L}^3\text{L}^3\text{T}^{-1}$) at each node in the soil domain, controlled by root density distribution, $\beta(r, z)$ (L^{-3}), and a soil water stress response function, $\gamma(h)$. r_{m} is a coefficient that represents the maximum radial root depth, and ET_{p} is the potential tree transpiration shown later in Equation 4.8. Both $\beta(r, z)$ and $\gamma(h)$ have functional values between 0 and 1.

The normalized root distribution, $\beta(r, z)$ for an axisymmetrical soil domain Ω, is defined by (Gardenass et al., 2005; Vrugt et al., 2001a):

$$\beta(r, z) = \frac{\beta^*}{2\pi \int_{\Omega} r\beta^*\, d\Omega} \tag{4.6}$$

with a general nonuniform root distribution, β^*, as defined by (Vrugt et al., 2001a):

$$\beta^*(r, z) = \left[\left(1 - \frac{z}{z_{\text{m}}}\right)\right]\left[\left(1 - \frac{r}{r_{\text{m}}}\right)\right] e^{-\left(\frac{p_{\text{z}}}{z_{\text{m}}}|z^* - z| + \frac{p_{\text{r}}}{r_{\text{m}}}|r^* - r|\right)} \tag{4.7}$$

where z_{m} and r_{m} define the maximum rooting extent in the vertical and radial directions (L), respectively. z^* and r^* are empirical parameters (L) that shift the maximum of the distribution in vertical and radial direction, respectively, and p_{z} and p_{r} are empirical parameters that determine the

exponential shape of the distribution. Refer to Figure 4.3 for an example of these root distribution values.

For water-stressed root conditions in the soil, $\gamma(h)$ (dimensionless) was introduced by Feddes et al. (1978), and reduces root water uptake from its maximum possible value because of soil water stress. $\gamma(h)$ is defined by four water potential values, P1 through P4. For soil water potential values between P2 and P3, $\gamma(h)$ will be optimum and equal to 1.0. For h-values between P1 and P2 (soil aeration stress) and between P3 and P4 (soil water stress), $\gamma(h)$ values will be smaller than one and zero at a minimum. Refer to Figure 4.4 for an example of these stress values.

For the estimation of the potential tree transpiration (ET_{p}) in Equation 4.5, meteorological data from the weather tower were used to estimate local hourly potential evapotranspiration, ET_0, using the Penman-Monteith equation (Allen et al., 1998). Values for the aerodynamic resistance and bulk surface stomatal resistance terms were calculated according to FAO guidelines (Allen et al., 1998). To estimate the potential tree transpiration, ET_{p}, we adopted the crop coefficient approach (Doorenbos, 1977), and multiplied ET_0 with a tree coefficient, s_{ET_0} (-) , or

$$ET_{\mathrm{p}} \quad = s_{\mathrm{ET}_0} ET_0 - E_{\mathrm{s}}. \tag{4.8}$$

We assume soil water evaporation to be negligible as canopy cover dominates the landscape and dry surface soil moisture conditions occurred throughout the study time period.

Finally, from integration of Equation 4.5 over the soil domain, the actual total root water uptake, R_{a}, (L T^{-1}) is computed from:

$$R_{\mathrm{a}} \quad = \tfrac{2\pi}{\pi r_{\mathrm{m}}^2} \textstyle\int_{\Omega} r W_{\mathrm{soil}} \, d\Omega. \tag{4.9}$$

In the presented coupled domain, the volume of water taken up by the roots must now be transported in the conducting vessels (xylem) of the sapwood in the tree trunk. For that purpose, the coupled model includes a small storage reservoir that acts as a buffer for water transport between the soil and the tree. Finally, by defining a lower flux boundary condition for the tree domain, the tree's

sapwood draws water from the buffer storage and initiates water flux through the tree. This water flux, and ultimately tree transpiration, are discussed in the next section.

4.3.5 Tree Transpiration

We use the Jarvis model to quantify plant transpiration (Waring et al., 1979). We represent the canopy transpiration sink term, $W_{\text{tree}}(h, z)$, in Equation 4.2 as follows:

$$W_{\text{tree}}(h, z) \quad = \gamma\,(h)\,\beta(z) ET_{\text{p}} \tag{4.10}$$

where $\beta(z)$ is the one-dimensional canopy density distribution function used for estimation of transpiration at different elevations of the tree domain, $\gamma(h)$ is the canopy water stress response function that represents the stomatal closure under increasing water tension and is of a similar Feddes form as used for characterizing soil water stress. ET_{p} is the potential tree transpiration from Equation 4.8.

The normalized canopy distribution, $\beta(z)$ for an axisymmetrical tree domain Ω, is defined by (Gardenass et al., 2005; Vrugt et al., 2001a):

$$\beta(z) \quad = \frac{\beta^*}{2\pi \int_\Omega \beta^*\, d\Omega} \tag{4.11}$$

with a general nonuniform tree distribution, β^* (L^{-1}), as defined by (Rings et al., 2013):

$$\beta^*\,(z) \quad = 1 - \left(\tfrac{z-6}{24}\right) \tag{4.12}$$

for z >6 m and zero below 6 m. Thus, the actual tree transpiration (ET_{a}) is computed from:

$$ET_{\text{a}} \quad = \tfrac{2\pi}{\pi r_{\text{m}}^2} \int_\Omega r W_{\text{tree}}\, d\Omega \quad = 2\pi R_{\text{a}} \int_\Omega \gamma\,(h)\,\beta\,(z)\; dz \tag{4.13}$$

In all, this approach couples root water uptake with tree transpiration.

4.3.6 Data

The model will be tested using data collected in and around a White Fir (Abies concolor) in a 99 ha subcatchment (P301) of the King's River Experimental Watershed (KREW), as part of the Critical Zone Observatory (CZO-TREE 1) project. This site is located in the rain-snow transition zone of the southern Sierra Nevada mountain range in California at an elevation of 2018 m. Data include soil water content and water potential in 3 spatial dimensions in the root zone, tree stem water content and sap flux, canopy water potential, and atmospheric variables including: net radiation, air temperature, and humidity. Undisturbed soil samples were collected to a depth of 2.5 m for the soil analysis. Corresponding measurements of saturated hydraulic conductivity were made using the constant head method (Reynolds et al., 2002).

Calibration data were selected for an 18-day rainless period in summer of 2009, starting July 15, and includes sapflow, stem potential, and soil storage. Three sap flow sensors (TransfloNZ, Palmerston North, NZ) were installed into the sapwood at a trunk height of 2.5 m. Using the compensation heat pulse technique (Green and Clothier, 1988), average sap flux flow (L/T) was estimated at 30-minute time intervals. Then, stem water potential measurements were taken from needle stems of various lower tree branches, at about 6 m from the ground. Seven measurements were taken during 24 hours on July 21-22, 2009. Finally, Echo- 5TE soil moisture sensors were installed at depths of 0.15 m, 0.30 m, 0.60 m and 0.90 m in each of 6 locations within a 5 m radius from the tree trunk. The sensors were calibrated in the laboratory (Kizito et al., 2008), from which it was determined that their accuracy is around 3% for a range of soils. Using all water content measurements, average total soil water storage (m^3) was computed during the 18-day measurement period every half hour. However, only about half of the soil water storage data was used for the parameter estimation due to incomplete measurements on some days.

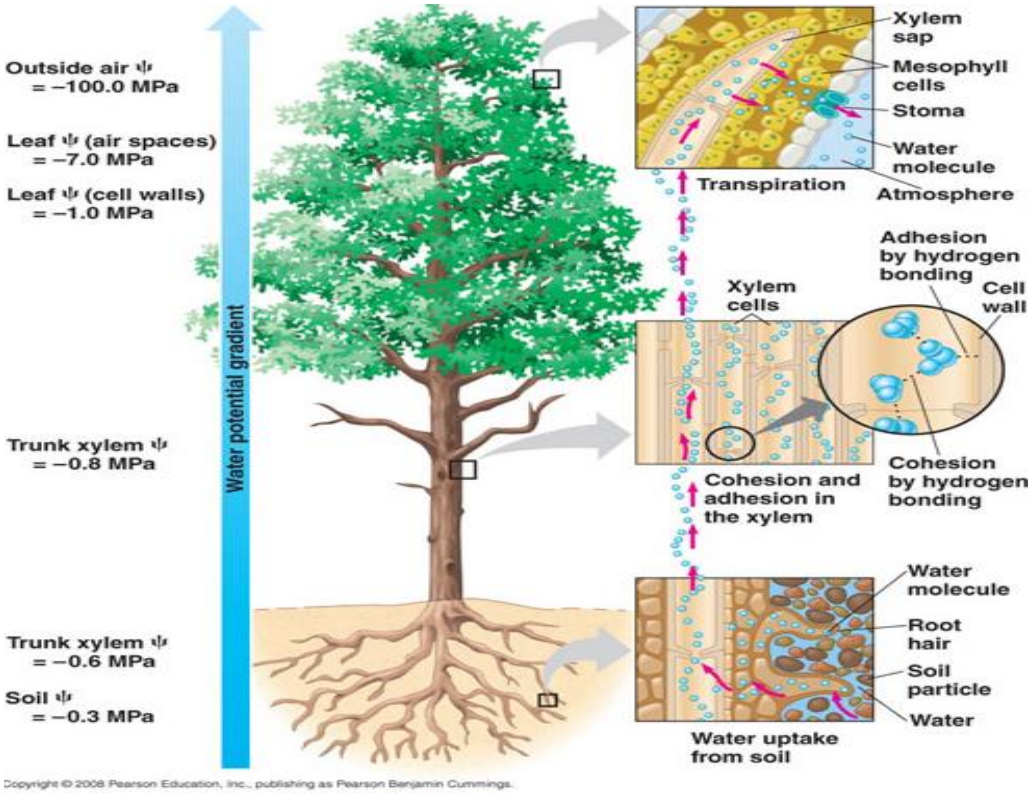

Figure 4.1: A schematic showing all of the processes involved for the simulation of water flow through the Soil-Tree-Atmosphere Continuum (STAC) model.

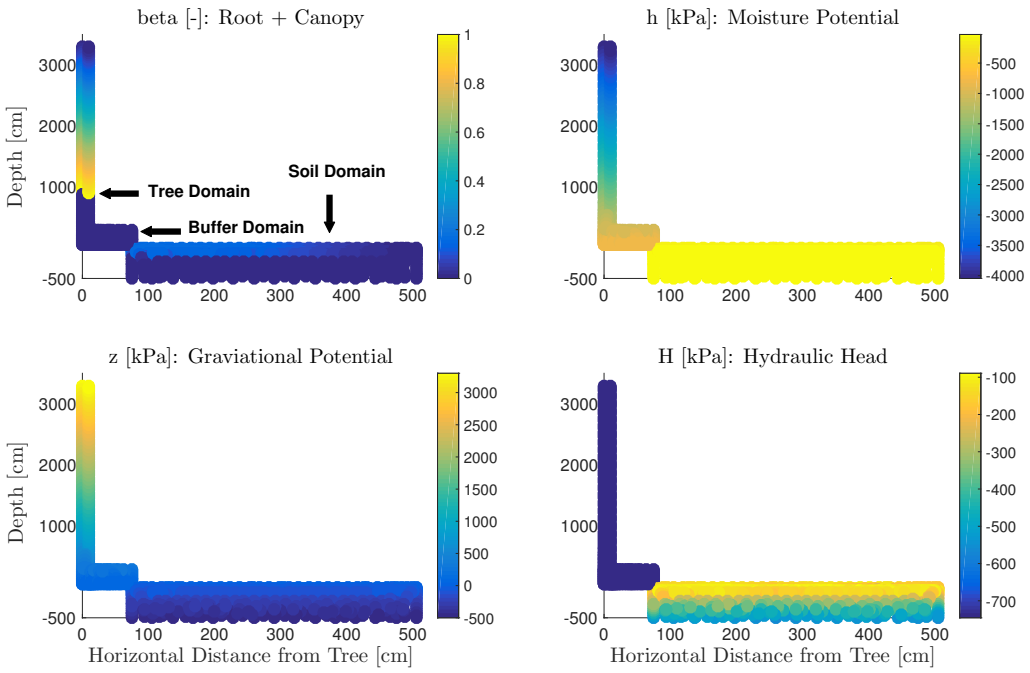

Figure 4.2: Initial domain used in the STAC model simulations. 'beta' represents the root or canopy distributions in Equations 4.7 and 4.12.

Table 4.1 Estimated STAC parameters for root distribution and Feddes stress functions, with units and prior ranges. Shown are the values when calibrating against each of the observed data sets, i.e. sapflow (SAP), stem potential (STEM), soil storage (STOR), and the combination of all 3 (FULL). The standard deviation of the posterior samples is shown in parenthesis. The subscripts z and r denote vertical or radial direction. The effect of these parameters on the root distribution and corresponding Feddes functions are represented visually in Figures 4.3 and 4.4.

Root Distribution Parameters			Estimation vs.			
Symbol	Prior Ranges	Units	SAP	STEM	STOR	FULL
p_z	50 - 250	(-)	219.47 (60.60)	60.43 (58.26)	243.64 (58.44)	184.13 (53.13)
p_r	50 - 250	(-)	145.88 (57.75)	184.72 (48.85)	101.31 (56.67)	181.80 (54.23)
z^*	50 - 250	cm	211.65 (56.75)	96.85 (59.96)	152.76 (56.56)	128.81 (58.61)
r^*	50 - 250	cm	54.29 (58.48)	230.77 (48.96)	113.97 (57.73)	123.04 (59.20)

Feddes Parameters - Calibrated			Estimation vs.			
Symbol	Prior Ranges	Units	SAP	STEM	STOR	FULL
$SoilP3$	-2e4 - -25	cm	-1.51e4 (5.6e3)	-1.04e4 (6.0e3)	-3.97e2 (6.0e3)	-1.09e4 (5.0e3)
$TreeP3$	-2e4 - -2	cm	-2.11e3 (1.8e3)	-1.01e4 (3.0e3)	-6.70e3 (5.9e3)	-1.30e2 (7.0e2)
$TreeP4$	-7.5e4 - -2e4	cm	-2.48e4 (7.3e2)	-2.27e4 (1.8e3)	-2.15e4 (1.6e4)	-2.00e4 (2.1e2)

Table 4.2 Estimated STAC parameters for the Van Genuchten functions of soil layers 1 and 2 and for the tree layer. Shown are the values when calibrating against each of the observed data sets, i.e. sapflow (SAP), stem potential (STEM), soil storage (STOR), and the combination of all 3 (FULL). The standard deviation of the posterior samples is shown in parenthesis. The resulting hydraulic conductivity and retention functions for soil layers 1 and 2 as well as for the tree layer are shown in Figure 4.5.

Soil VG Parameters			Estimation vs.			
Symbol	Prior Ranges	Units	SAP	STEM	STOR	FULL
$K_{s,Soil1}$	0.2 - 1.2	cm day^{-1}	0.55 (0.37)	0.96 (0.38)	1.07 (0.37)	0.54 (0.37)
$\theta_{s,Soil1}$	0.2 - 0.6	m^3 m^{-3}	0.31 (0.10)	0.49 (0.08)	0.21 (0.12)	0.54 (0.12)
$K_{s,Soil2}$	0.2 - 1.2	cm day^{-1}	0.48 (0.32)	0.79 (0.34)	0.25 (0.40)	1.46 (0.40)
$\theta_{s,Soil2}$	0.2 - 0.6	m^3 m^{-3}	0.39 (0.11)	0.36 (0.13)	0.21 (0.12)	0.24 (0.12)

Tree VG Parameters			Estimation vs.			
Symbol	Prior Ranges	Units	SAP	STEM	STOR	FULL
α_{Tree}	1e-6 - 3e-5	cm^{-1}	2.18e-5 (5.0e-6)	1.42e-5 (3.5e-6)	2.49e-6 (6.0e-6)	2.99e-5 (7.5e-5)
n_{Tree}	4.0 - 8.0	(-)	4.04 (1.02)	7.38 (1.25)	4.87 (1.15)	4.12 (0.19)
$K_{s,Tree}$	0.5 - 2	cm day^{-1}	1.32 (0.04)	1.93 (0.15)	1.91 (0.42)	1.38 (0.02)
$\theta_{s,Tree}$	0.05 - 0.6	m^3 m^{-3}	0.29 (0.10)	0.12 (0.18)	0.14 (0.16)	0.09 (0.06)

4.3.7 Bayesian Inference of Model Parameters

The flow simulations are coupled with observed data using a rigorous sampling algorithm, the Differential Evolution Adaptive Metropolis (DREAM) of Vrugt (2016). The STAC model contains 15 parameters that can be estimated. A total of 4 parameters are used to characterize the spatial tree root distribution (see Table 1), and these include z^* and r^* which are empirical parameters (m) that shift the maximum of the distribution in vertical and radial direction, respectively, and p_z and p_r are empirical parameters that determine the exponential shape of the distribution. Also, a total of 3 parameters are used to characterize the water stress response functions (see Table 1). Finally, a total of 8 parameters are used to characterize the hydraulic conductivity and retention of the entire coupled domain, and are listed in Table 2. These parameters include θ_s, K_s, α, and n for the tree layer, and θ_s and K_s for soil layers 1 and 2.

In recent decades, Bayes' theorem has emerged as a working paradigm for modern probability theory and hypothesis testing Vrugt (2016). This theorem (also referred to as Bayes' law or Bayes' rule) expresses mathematically the fundamental relationship between the prior, conditional, and posterior beliefs of the parameter values (a 15-dimensional vector denoted by $\mathbf{x}$). This probability equation is formalized as follows:

$$P(\mathbf{x}|\widetilde{\mathbf{Y}}) = \frac{P(\mathbf{x})P(\widetilde{\mathbf{Y}}|\mathbf{x})}{P(\widetilde{\mathbf{Y}})}, \tag{4.14}$$

where $P(\mathbf{x})$ and $P(\mathbf{x}|\widetilde{\mathbf{Y}})$ signify the prior and posterior parameter distributions, respectively, and $P(\widetilde{\mathbf{Y}}|\mathbf{x})$ represents the likelihood function, also denoted by $L(\mathbf{x}|\widetilde{\mathbf{Y}})$. The model evidence, $P(\widetilde{\mathbf{Y}})$ (or the marginal likelihood) acts as a normalizing constant here and is canceled out in this analysis. Therefore we can estimate the probability of a given parameter set, $\mathbf{x}$, when evaluated against a measured data set, $\widetilde{\mathbf{Y}}$, using the following proportionality statement,

$$P(\mathbf{x}|\widetilde{\mathbf{Y}}) \propto P(\mathbf{x})L(\mathbf{x}|\widetilde{\mathbf{Y}}). \tag{4.15}$$

Since the prior distributions of the unknown parameters are assumed to be uniform, Equation 4.15 is reduced to

$$P(\mathbf{x}|\widetilde{\mathbf{Y}}) \propto L(\mathbf{x}|\widetilde{\mathbf{Y}}). \tag{4.16}$$

4.3.8 Likelihood Function

In order to estimate the model parameters to the calibration data set, a likelihood function must be defined prior to the estimation. Generally, a single data set is used for the estimation of a model's parameters, whereas in this case we want to join various data streams, namely the sapflow, stem potential, and soil storage measurements. Therefore, a likelihood function is needed that can combine all three sources of information without being affected by the magnitude of error that is contributed from any of the individual sources. This will allow the parameter estimation algorithm to combine residuals from many predictions, in this case sapflow, stem potential, and soil storage.

This likelihood function is derived mathematically based on assumptions regarding simulation residuals, and is applied as follows:

$$L(\mathbf{x}|\widetilde{\mathbf{Y}}, \phi, \sigma^2) = -\frac{1}{2}\sum_{j=1}^{3}\sum_{t=2}^{n_j}\left\{\frac{(e_{j,t}(\mathbf{x}) - \phi e_{j,t-1}(\mathbf{x}))}{\sigma_{j,t}}\right\}^2 \tag{4.17}$$

where j distinguishes the various model outputs considered for the parameter estimation, $\mathbf{x}$ is the parameter set, $\widetilde{\mathbf{Y}}$ is the observed data, ϕ is the temporal correlation of the residuals (with 30 mins between each time step), and very importantly σ is the measurement error of the calibration data. $e_{j,t}(x)$ is the error of the model simulations for output j and a given parameter set $\mathbf{x}$. Therefore, we calculate a likelihood value from each considered output (i.e. $L(\mathbf{x}|\widetilde{\mathbf{Y}}, \phi, \sigma^2)_j$), and as Equation 4.17 indicates we sum the three likelihood values to obtain one overall 'probability' value. Each of the sapflow, stem potential, and soil storage outputs are compared with their respective observation data, and a likelihood value is assigned to each simulation. The total likelihood of our proposed

parameter vector, $\mathbf{x}$, is the sum of these three individual likelihood values. Note, the likelihood function in Equation 4.17 is actually applied as a log-likelihood for mathematical convenience, i.e. $\mathcal{L}(\mathbf{x}|\widetilde{\mathbf{Y}}, \phi, \sigma^2)$. This transformation allows us to simply add each likelihood value, rather than multiplying them as the original likelihood function would require. Please refer to Vrugt and Massoud (2017) for further information.

For this study we examine various parameter estimation strategies. First we fit the model simulations to individual data sources (i.e. sapflux, stem potential, and soil storage), and then fit the simulations to all three sources combined. In the first case, the model is calibrated to just the sapflux data, and Equation 4.17 is reduced to just one term, the likelihood obtained from the fit to the sapflux data; these results are assigned the 'SAP' acronym. For the second and third cases, the model is calibrated to the stem potential and soil storage data, respectively, and Equation 4.17 is similarly reduced to one term; these results are assigned the 'STEM' and 'STOR' acronyms, respectively. In the final case, the model is calibrated to all three data sets, and Equation 4.17 utilizes all three likelihood terms; these results are assigned the 'FULL' acronym. This final case highlights the ability of the likelihood function to combine various data streams by normalizing the prediction errors based on the observation error and combining the likelihood (or probability) from all three processes considered. For the parameter estimation runs, the measurement error values were defined as $\sigma_{Sap} = 1$ cm/day, $\sigma_{Stem} = 100$ kPa, and $\sigma_{Stor} = 0.05\ m^3$.

4.4 Results

For the remainder of the paper, the results are color coded as follows: the estimation to sapflux (SAP) is shown in blue, estimation to stem potential (STEM) is shown in light blue, estimation to soil storage (STOR) is shown in green, and estimation to all three data sets (FULL) is shown in black.

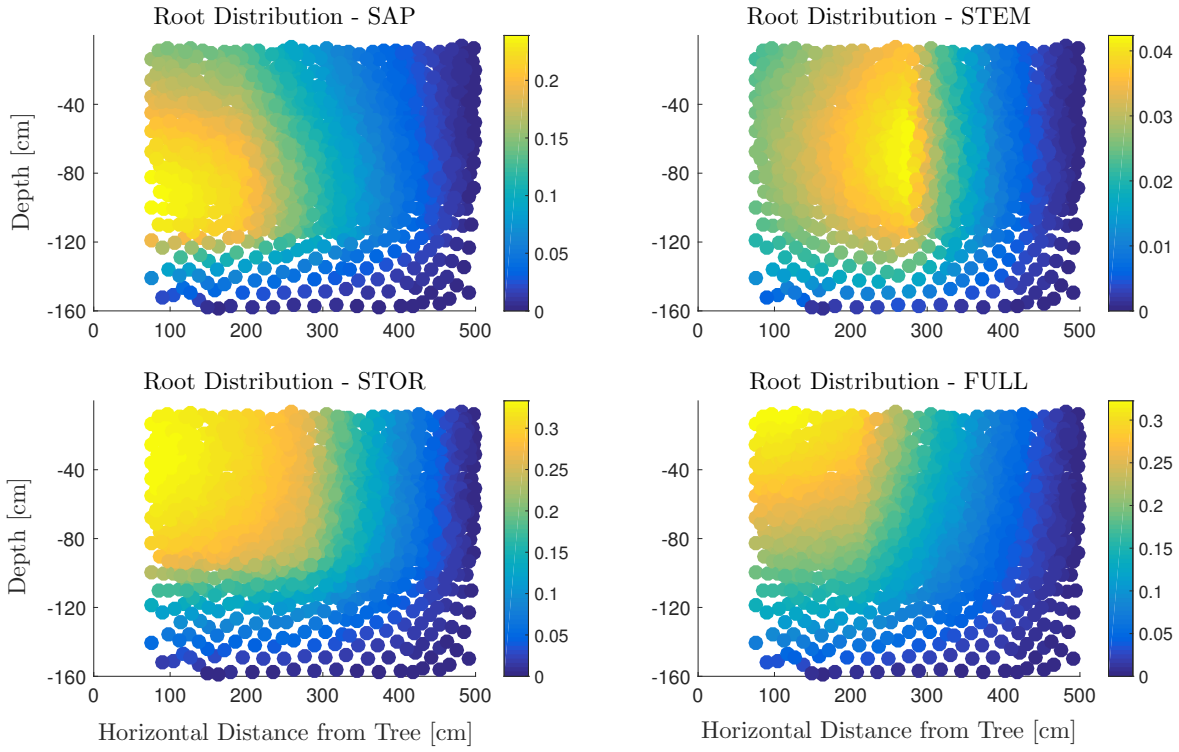

Figure 4.3: Root distributions created from each of the parameter estimation methods. These figures were created using parameter values from Table 1 and the general nonuniform root distribution formula of Equation 4.7.

4.4.1 Parameter estimates

The STAC model simulates the spatial root distribution of the tree using Equation 4.7. For the 4 estimation methods considered, the estimated root distribution parameters are shown in Table 1. The resulting root distributions shown in Figure 4.3 allow a visual comparison of all the estimation strategies.

Both the tree and its roots may experience stress from water limitation, as well as nutrient limitation, or stress from extreme vapor pressure deficits as a result of hot temperatures, among other factors. The STAC model numerically accounts for this through the stress terms, γ, in Equations 4.5 and 4.10. These stress terms are characterized using the Feddes function (Feddes et al. (1978)), and have values ranging from 0 (full stress) to 1 (no stress). Four levels of head,

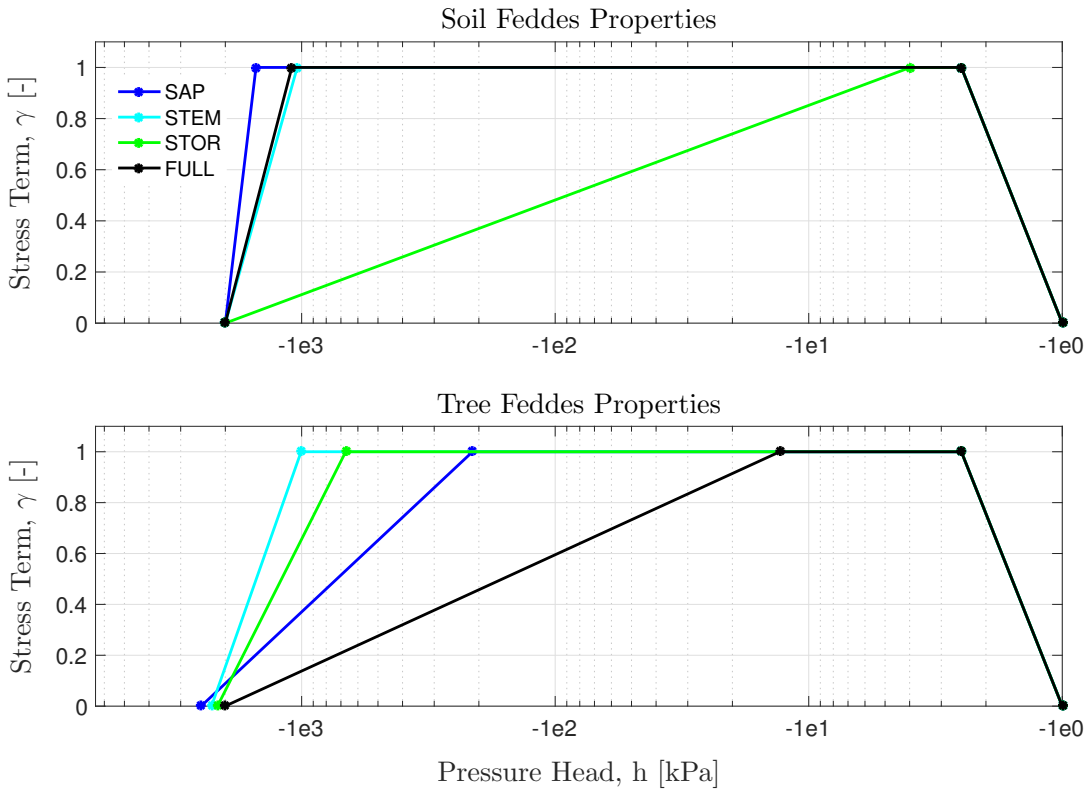

Figure 4.4: Feddes stress functions for both the soil and tree layers. These functions were created using the coefficients of Tables 1. The stress term, γ, in Equations 4.5 and 4.10 are the values on the y-axis. For pressure head values between P2 and P3, there is no stress and $\gamma = 1$. The stress occurs between P1-P2 (aeration stress, or saturated conditions) and between P3-P4 (water stress, or drought conditions). These figures show that stress occurs sooner for the tree domain than for the soil layers, which generally indicates that the tree layer cannot carry as much moisture as the soil layer is able to, otherwise the tree starts to become stressed.

or pressure, are used to express this function, P1, P2, P3, and P4. For soil water potential values between P2 and P3, there is no stress, while the stress occurs between P1-P2 (aeration stress, or saturated conditions) and between P3-P4 (water stress, or drought conditions). A few of these coefficients are parameterized, and the calibrated values are shown in Table 1. After all Feddes parameters are defined, the stress functions for each layer can be constructed. These functions are shown in Figure 4.4 and also allow a visual comparison of the various estimation strategies.

Through Equations 4.3 and 4.4, the STAC model characterizes the retention and hydraulic conductivity of each soil layer and of the tree layer. Table 2 shows the values of the model parameters that are Van Genuchten parameters used to create hydraulic relationships of the tree and soil

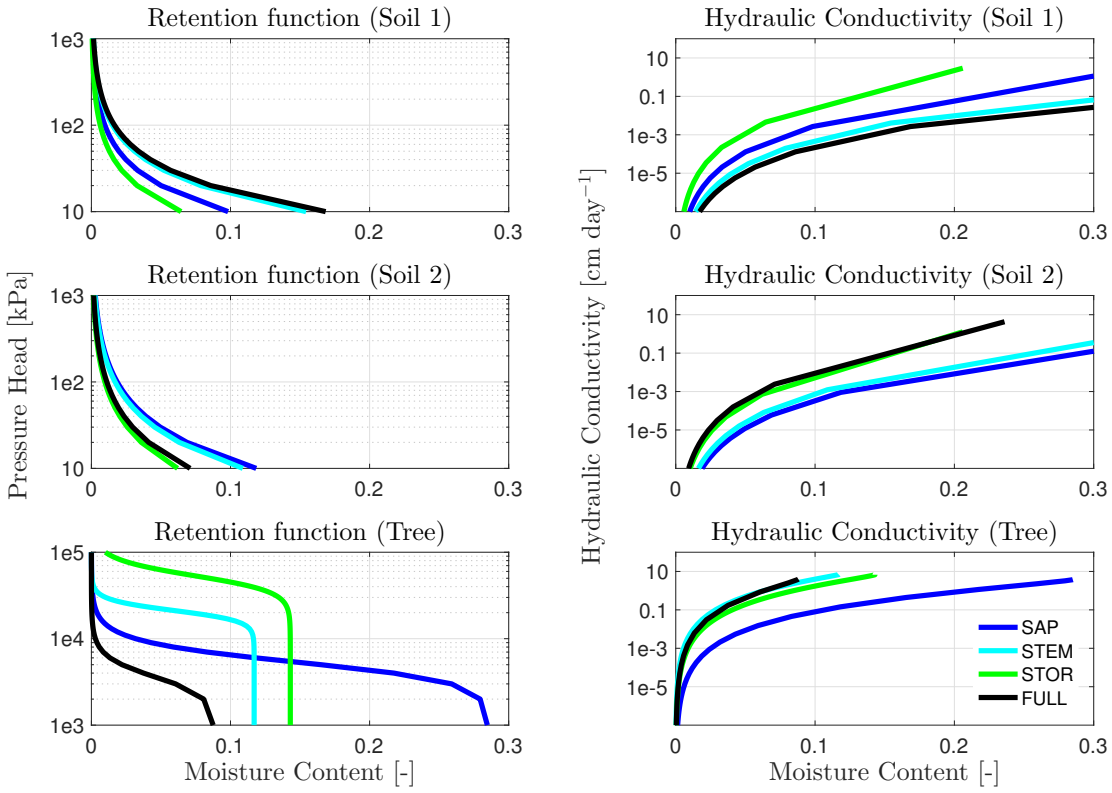

Figure 4.5: Hydraulic relationships for soil layer 1, soil layer 2, and the tree layer. Shown in the first column of figures are the retention functions for each layer, and in the second column the hydraulic conductivity of each layer is shown. The parameters used to create these relationships are presented in Tables 2.

domains. These parameters help represent the saturated conductivity and moisture contents of each layer and also contain certain shape parameters for the Van Genuchten functions. The resulting hydraulic conductivity and retention functions for soil layers 1 and 2 as well as for the tree layer are shown in Figure 4.5. In the first column of figures, the retention function of each layer is shown. One thing to note is that the retention function of the tree domain greatly differs from that of the soil layers. The figure shows that it will require higher amounts of pressure to extract the water from the tree than from the soils (10^3 kPa for soils vs. 10^5 kPa for tree), which can be expected (Black and Pritchard, 2002; North and Nobel, 1997).

4.4.2 Comparing STAC model simulations with observations

In this study we focus on three processes represented in the STAC model, which are the sapflux through the tree domain, the stem water potential in the canopy, and the water storage of the soil domain. All of these outputs are accompanied by observed data, and the model simulation results are shown in Figure 4.6 for each parameter estimation method. For the sapflux simulations, the SAP and FULL strategies performed the best according to the RMSE (Table 3). The other estimation strategies (STEM and STOR) do not fit the observations quite as well, however they do allow the simulations to capture high peaks in the observed data. For the stem potential simulations, the STEM strategy provided the lowest RMSE. Although estimation to the stem potential data was not as informative since there were only 7 data points to fit, these observations still allow us to constrain the simulations. This is shown in Table 3 since the RMSE for the stem potential simulations of the FULL strategy is lower than that of the SAP and STOR strategies, which is not surprising since the FULL strategy considers the stem potential data in its likelihood function. For the soil storage simulations, the STOR parameter set performed the best according to the RMSE, yet we see in Figure 4.6 that all the simulations behave almost identically. In all simulations, the model underestimated the observed soil storage, indicating a flaw in the model that allows too much water to be drawn from the soil domain than is observed in the data.

Table 4.3 RMSE of model outputs for each parameter estimation strategy using the MAP parameter values. Tabulated values shown here are corresponding to simulations shown in Figure 4.6.

RMSE	SAP	STEM	STOR	FULL
Sapflux (cm day^{-1})	0.684	2.27	2.60	0.681
Stem Potential (kPa)	265.3	138.8	513.0	236.9
Storage (m^3)	0.054	0.054	0.052	0.053

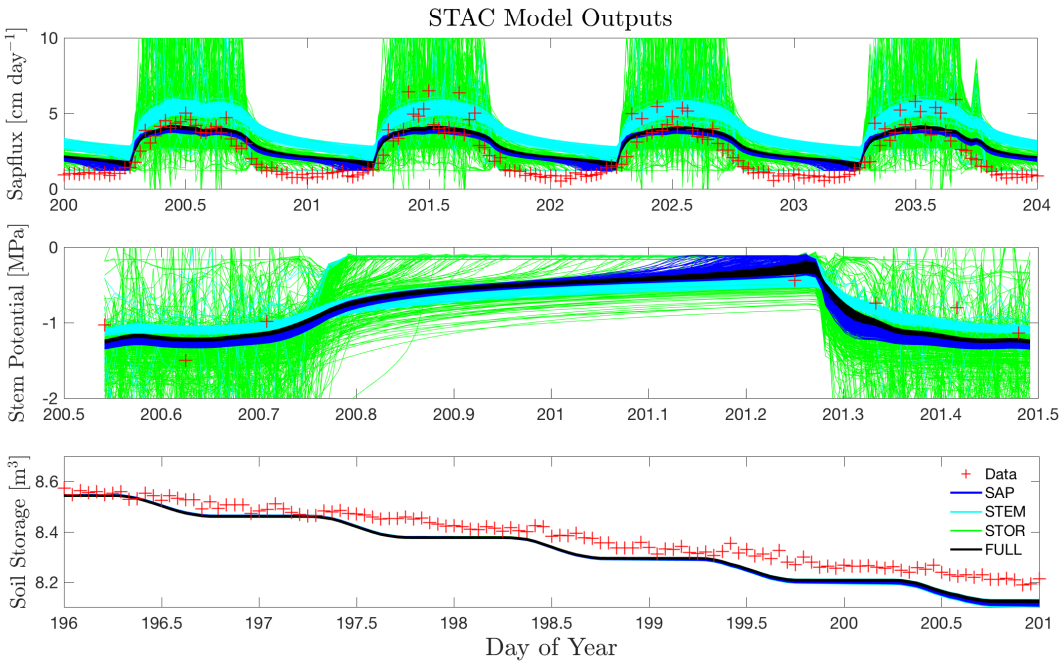

Figure 4.6: Posterior solutions of the STAC model outputs, including sapflux (cm day^{-1}), stem potential (MPa), and water storage in the soil domain (m^3). Observations are shown with red marks. The simulations with parameters calibrated to sapflux data are shown in blue (SAP), simulations with parameters calibrated to stem potential data are shown in light blue (STEM), simulations with parameters calibrated to soil storage data are shown in green (STOR), and simulations with parameters calibrated to all the data are shown in black (FULL).

4.5 Discussion

4.5.1 Including multiple data streams in likelihood function

With the advance of ecohydrologic models and increase in the corresponding data used to inform these models, estimation of hydraulic processes between the land and air interface is becoming more possible. Here, we investigated a process-based ecohydrologic model (the STAC model) at a single site. STAC was calibrated using various data streams since ideally we want the model to behave with fidelity to several processes, and not just one. Various estimation strategies were tested, most of which including only one data stream for calibration, with the exception of one case that included three different streams of data in the estimation process.

According to results in Table 3, the FULL estimation strategy performed the best overall, one

can argue since it considers all three data streams in its likelihood function. For the sapflux simulations, the FULL strategy was even better than the SAP strategy. For the stem potential simulations, the FULL strategy was second best, to the over-fitted STEM strategy. For the soil water storage, the FULL strategy again performed second best. Therefore, it seems that when multiple processes are considered in an environmental or ecologic parameter estimation problem, the obtained parameter values and corresponding model simulations are better able to represent real-life properties and processes observed in the system.

4.5.2 Hydraulic properties in trees

Numerical models that simulate environmental processes range in scale from single site (such as the STAC model) to the global, such as many Earth System Models (ESMs) in use today. Although some of these models are state-of-the-art in their development and management, few (if any) of them consider process based dynamics of coupled soil-tree water transport (Powell et al., 2013). When considering a forest or continental scale simulation, the aggregation and averaging of water behavior within and across trees causes a huge discrepancy between what the models simulate and what is actually observed in reality. Thus, it is argued that more understanding of water processes (such as water storage, residence time, resistance to drought) within trees is needed for a better representation of the water cycle between the land and air surface. We hope that methods and results presented in this study will show case the usefulness of the likelihood in Equation 4.17, in which various calibration data streams can be considered for better representation of hydraulic properties in a single tree.

4.5.3 Other parameter estimation methods - ABC

In this study, the parameter estimation algorithm sums the errors of the simulations into a single index and computes a likelihood based on this sum. In fact, most calibration studies are performed

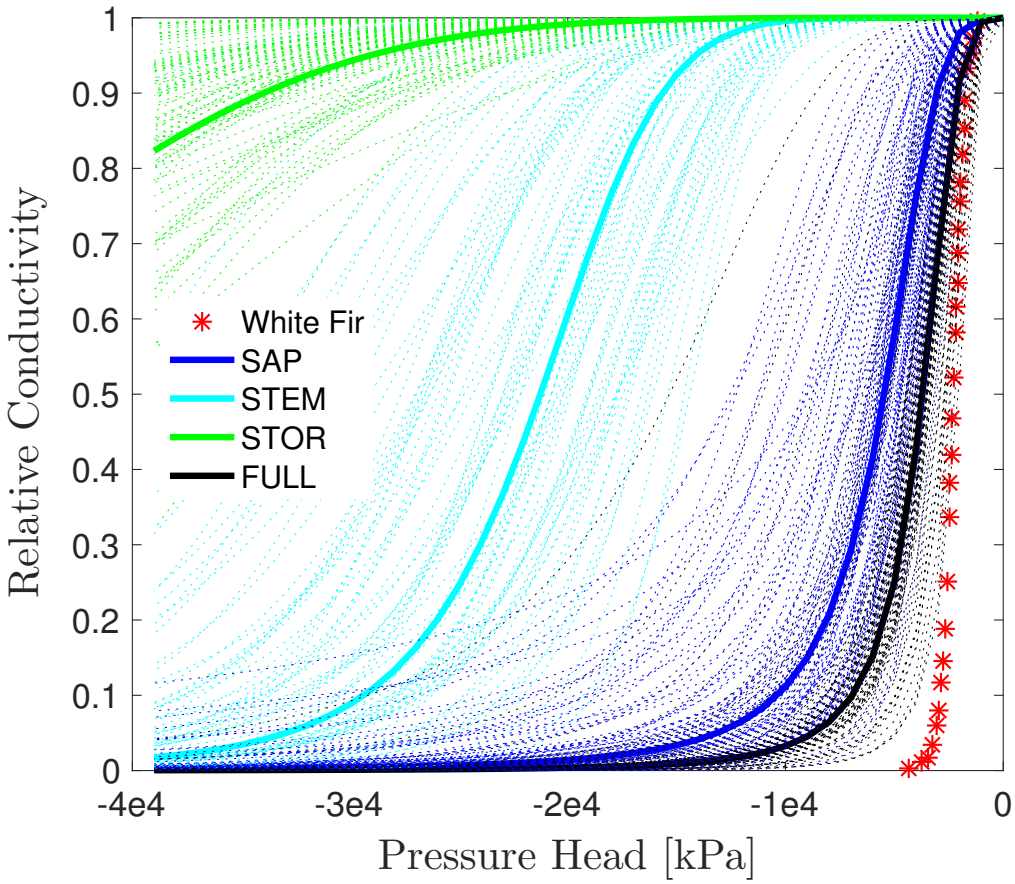

Figure 4.7: Observations of the relative hydraulic conductivity in a Douglas White Fir (Abies Concolor), compared with relationships simulated with the STAC model using calibrated parameters in this study. The mean relationships are shown with solid curves and uncertainty ranges are presented with dotted lines.

in this manner. However, likelihood-free calibration methods are new in the literature, e.g. the Approximate Bayesian Computation (ABC) method which allows parameter estimation to a set of summary metrics instead of calibrating to a set of simulation residuals (Sadegh and Vrugt, 2014; Vrugt and Sadegh, 2013). For instance, Figure 4.7 shows the relative hydraulic conductivity of a White Fir (Abies Concolor) compared with those produced from the calibrated parameter sets in this study. The FULL parameter set creates a hydraulic relationship that is most realistic according to the observed relationship. Yet, there could be a parameter set that produces a closer hydraulic relationship to the observed data. This set can be inferred with the ABC method, and may provide even more realistic parameters. Although this goes beyond the scope of this study, we

encourage readers to try various parameter estimation algorithms that are available, and with ABC we encourage the use of many different summary metrics that may better capture the hydraulic behaviors of the soil-tree-atmosphere system.

4.6 Conclusion

Major inconsistencies exist in the representation of vegetation in large scale land surface and climate models. Model parameters are typically inferred from empirical relationships or extracted from arbitrary data sets and efforts are now aimed to identify parameter sets that appropriately describe these vegetation properties. We presented simulations with the Soil-Tree-Atmosphere Continuum (STAC) model showing the hydraulic processes of a mature white fir (Abies concolor) and its surrounding root zone. The model couples both soil and tree domains, and simulates the movement of water based on different ecohydrologic processes. We used Bayesian inversion to estimate the model parameters against observations of sapflux, stem water potential, and soil water storage.

We evaluated the models ability to describe the data, with specific emphasis on the soil and tree water flow and storage properties. The calibration of the model allowed us to estimate the spatial root distribution of the tree, the Feddes stress parameters to describe aeration or water stress, and the Van Genuchten parameters that correspond to the retention and conductivity functions of soil and tree domains. After calibration, the STAC model simulated processes such as sapflow, stem potential, and soil storage, and the outputs were compared with the observed data for a full diagnosis of the model.

The results presented in this paper show that the choice of calibration data largely affects the parameter estimates and thus the model outputs. By considering the full domain of the tree and combining all the observed data in the parameter estimation process, the most realistic parameter combination was estimated and the closest fit between the model outputs and the observed data was achieved. A likelihood function that considers various streams of data by normalizing simulation

errors by the measurement errors was considered and implemented.

We conclude that the STAC model offers a physical representation of water flow in and around a vegetative medium, and can provide insight on how trees actually behave in their environments. Ecohydrologic models are evolving from having empirically based structures to more physically based ones, setting the stage for tools such as the STAC model to expand our knowledge on fundamental hydraulic processes that occur in vegetation.

4.7 Acknowledgements

This research is all in thanks to Dr. Jasper A. Vrugt for providing most of the tools necessary to conduct this study, including the STAC model source code, the calibration data from the KREW site, as well as constant discussion that inspired ideas. The authors are very appreciative of support from the SCGSR fellowship of the Department of Energy, the UC-Lab Fees Research Program Award 237285, and the DOE Office of Science Next Generation Ecosystem Experiment at Tropics (NGEE-T) project.

Chapter 5

Emulation of Environmental Models

by Elias C. Massoud, Eric Laloy,
Chonggang Xu, and Jasper A. Vrugt

Reference:

Massoud, E. C., E. Laloy, Xu, C., Vrugt, J.A.: Emulation of environmental models using polynomial chaos expansion. Environmental Modeling & Software, Submitted. X: XX-X. doi: XX.XX/x. XXX. 2017.

5.1 Manuscript Title - Emulation of environmental models using polynomial chaos expansion

Abstract: Computer models are collections of scientific laws and principles that explain real-world systems and predict their behaviors. More specifically, environmental models are applied for problems regarding the Earth system, and have been evolving in the past few decades due to extensive hypothesis testing, field and laboratory experimentation, and qualitative and quantitative interpretation. Unfortunately, a model's CPU, memory, and storage requirements may impose limits on its application, and complicate routine tasks such as sensitivity analysis, parameter estimation, uncertainty quantification, and scenario-analysis. Here, we investigate the applicability of emulation to speed-up the simulation time of CPU-intensive computer models. We illustrate the application of polynomial chaos expansion (PCE) to model emulation and parameter estimation using three different environmental case studies of increasing complexity. Our results demonstrate that PCE emulators mimic closely outputs of relatively simple, low-dimensional, simulation models, but do not approximate as well outputs of complex, parameter-rich, system models.

5.2 Introduction

Ecosystems and environmental systems constitute a complex network of living organisms, interconnected with their environment through a myriad of physical, chemical and biological processes operating at or near the Earth's surface. Many of these processes are difficult to observe directly in the field, particularly at large spatial scales. This daunting complexity has stimulated researchers in many different fields of study to explore the use of mathematical modeling to mimic the behavior of complex systems. Computer models are particularly useful to gain insights and understanding of system functioning and to predict their temporal and spatial behavior. The capabilities of

such models exceed by far traditional paper-and-pencil calculations and can involve simulations on spatial scales of individual atoms to the entire ecosystem, and temporal scales of nanoseconds to many millions of years. Examples include numerical weather prediction models, astrophysical and cosmological simulations of dark matter, computational modeling of the brain, or spatially distributed simulation of environmental systems. The CPU-time of these simulations can vary from less than a second for simple dynamic models up to many hours of calculation for spatially explicit models involving multidimensional numerical solution of differential/algebraic or ordinary/partial differential equations.

Earth system models (ESMs) are dynamic representations of the atmosphere, ocean, ice and land, described in coupled systems to analyze and predict short and long term Earth system behavior (Claussen et al., 2002; Hurrell et al., 2013; Wood et al., 2011). Typically, obtaining better accuracy in these models comes at the expense of dimensionality and complexity, and thus an enhanced computational efficiency (Arora et al., 2013; Dunne et al., 2012). State-of-the-art weather and climate prediction models contain a large number of parameters to simulate the exchange of energy, momentum and mass, between the land-surface and overlying atmosphere (Bastidas et al., 1999; Masson et al., 2003; Noilhan and Planton, 1989; Sargsyan et al., 2014). Calibration of the model parameters is a necessary step before the model can be ran for operational use (Gupta et al., 1999; Rosolem et al., 2013). Yet, there is an immense CPU and time cost associated with estimating model parameters since many thousands of runs are required to properly estimate the parameter values (e.g. using Markov Chain Monte Carlo (MCMC) methods). Certainly, computational limitations still remain a major obstacle to the effective use of ESMs in environmental decision-making (Ratto et al., 2012).

Emulation is an important and growing field of research that signifies a major achievement in the study of complex mathematical models (Ratto et al., 2012; Razavi et al., 2012). These methods define surrogate responses of the original model outputs at a minimal computational cost. The

concept of emulation dates back to Blanning (1975) and Kleijnen (1975) and has been greatly evolving in recent years. The central idea is that the original complex model has a response surface for its outputs (for a pre-defined parameter range), and this response surface has some mathematical representation that the emulation methods are built to capture. Here we present and implement polynomial chaos expansion (PCE) emulators (Ghanem and Spanos, 1991; Le Maître and Knio, 2010; Xiu and Karniadakis, 2002), which utilize orthogonal polynomials to mimic the parameter effects on model outputs, and we build surrogates of various environmental models. To our knowledge, this is the first application of PCE emulators for environmental models.

There are various algorithms and methods other than PCE used for emulation and surrogate modeling and for many different types of applications. The set of methods referred to as the design and analysis of computer experiments (DACE) is very popular (Levy and Steinberg, 2010; Sacks et al., 1989; Santner et al., 2013), and books, reviews, and journal special issues have been written on this subject (Fang et al., 2005; Kleijnen, 2010; Kuhnt and Steinberg, 2010; van Gigch, 1991). The DACE method is also known as Gaussian Process (GP) or kriging metamodeling (Di Pierro et al., 2009). For example, Conti and O'Hagan (2010) applied a GP emulator to mimic a multi-output and dynamic computer simulation model, the Sheffield Dynamic Global Vegetation Model, and Machac et al. (2016) used a similar method to create a surrogate model for an urban hydrodynamic drainage simulator. Another emulation method uses radial basis functions (RBF), such as the techniques used in Bliznyuk et al. (2012) and Regis and Shoemaker (2007a,b). Furthermore, artificial neural networks (ANN) and support vector machines (SVM) are also emulation techniques, and have been used in many studies such as the emulation of a hydrologic model (Zhang et al., 2009), the identification of controlling mechanisms in transpiration of a Pine tree (Vrugt et al., 2002a), or the estimation of regional variation in particulate organic carbon to nitrogen ratio in the surface ocean (Martiny et al., 2013). In Keating et al. (2010), prediction uncertainty of a highly complex CPU-intensive groundwater model was explored using a simplified surrogate model. Emulators

have been used for data assimilation and parameter estimation using a probabilistic collocation based Kalman filter (PCKF), which combines the Kalman filter with PCE emulators (Fan et al., 2016; Man et al., 2016; Saad and Ghanem, 2009). Additionally, Laloy et al. (2013) jointly use a PCE emulator of the original model and a dimensionality reduction (using the Karhunen-Love tranform) of the parameter space to speed up the calibration of a groundwater model. Zeng et al. (2016) used similar methods to build PCE emulators of a groundwater model using stochastic collocation and sparse grids.

A goal of the study is to investigate the applicability of PCE based emulators for the calibration of complex environmental and Earth system models. First, we consider two environmental models of low and medium complexity, and compare the calibration of the original models to that of their emulators. These two cases involve a conceptual rainfall runoff model (hmodel) (shown in Schoups and Vrugt (2010)) and a Soil-Tree-Atmosphere Continuum (STAC) model that simulates hydrodynamic processes of a single tree (Rings et al., 2013). Finally, we create an emulator for an ESM, the Community Land Model (CLM) shown in Fisher et al. (2015). For this study, we build the polynomial emulators using training parameters sampled from sparse grid approximation (Nobile et al., 2008a,b; Smolyak, 1963; Xiu and Hesthaven, 2005). Sparse grids are numerical techniques to represent, integrate or interpolate high dimensional functions, and can be useful in the application of PCE emulators to sample the locations of the training runs. The user chooses the 'level' of sparse grid approximation, where an increasing level denotes an increased density of sampling (which usually indicates a stronger emulator). We test 1^{st} and higher level sparse grids, as well as 1^{st} and higher order polynomial emulators, and investigate the level of training and order of polynomial necessary to sufficiently build accurate emulators. We analyze for these models the accuracy, efficiency, and fidelity of the emulators compared to their original counterparts. Our aim is to show that calibration of the emulator is an option for calibration of complex models.

The remainder of this paper is structured as follows. Section 2 describes the PCE method and

reviews briefly the use of sparse grids to minimize emulator training. In section 3 we introduce the models and data of each case study and report immediately the corresponding results of the PCE emulator. This is followed in section 4 by a general discussion of the main implications of our collective results. Finally, section 5 presents a summary of the most important findings and offers concluding remarks.

5.3 Materials and Methods

In this section, we will briefly review the building blocks of the PCE method that is used herein to emulate the output of different ecosystem models. Next, we will revisit the use of sparse grid approximation methods to help build polynomial emulators with the least amount of grid points, that is, values of the model parameters.

5.3.1 Model formulation

Consider a n-vector of measurements, $\widetilde{\mathbf{Y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$ observed at discrete times $t = (1, \ldots, n)$ that summarizes the response of an environmental system, $\Im$, to k temporally-variant control inputs, $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$, with column elements, $\mathbf{b}_t = (b_{t1}, \ldots, b_{tk})$. We use a computer model, $\mathcal{F}(\cdot)$, to explain the observed data

$$\widetilde{\mathbf{Y}} \leftarrow \mathcal{F}(\mathbf{x}, \widetilde{\boldsymbol{\zeta}}_0, \widetilde{\mathbf{B}}) + \mathbf{E}, \tag{5.1}$$

where $\mathbf{x} = (x_1, \ldots, x_d)$ is the $d \times 1$-vector of model parameters, $\widetilde{\boldsymbol{\zeta}}_0$ stores the values of the state variables at the start of simulation, $\widetilde{\mathbf{B}}$ signifies the control matrix with temporal measurements of the forcing variables, and $\mathbf{E} = (e_1, \ldots, e_n)$ is a vector of residuals

$$\mathbf{E}_{\mathcal{F}}(\mathbf{x}, \widetilde{\boldsymbol{\zeta}}_0, \widetilde{\mathbf{B}}) = \widetilde{\mathbf{Y}} - \mathbf{Y}_{\mathcal{F}}(\mathbf{x}, \widetilde{\boldsymbol{\zeta}}_0, \widetilde{\mathbf{B}}). \tag{5.2}$$

The index t for time takes on strictly positive integer values in the remainder of this paper, $t \in \{1, \ldots, n\} \in \mathbb{N}_+$, yet may take on real values, $t \in (0, n] \in \mathbb{R}_+$ in the actual system model, $\mathcal{F}(\cdot)$, to

resolve for continuous-time processes, wherein the simulated output at $t = 0$ is defined completely by $\widetilde{\boldsymbol{\zeta}}_0$.

We make the convenient assumption that the forcing data are observed without measurement error, and that errors in the initial states, $\widetilde{\boldsymbol{\zeta}}_0$, pose no harm as their impact on the simulated output, $\mathbf{Y}$, diminishes rapidly with advancing time. This latter assumption is certainly appropriate for environmental systems controlled by negative (or degenerative) feedback. The assumptions of perfect input data, and initial states (due to spin-up period) are common to environmental modeling. This ideal case leaves as our only "unknowns" the model parameters. Without further loss of generality, we restrict the model parameters to a closed space, $\boldsymbol{\chi}$, equivalent to a d-dimensional hypercube, $\mathbf{x} \in \boldsymbol{\chi} \in \mathbb{R}^d$, called the feasible parameter space. Note, that the tilde operator is used for the initial state and forcing data to signify explicit use of measured values.

The large computational requirements of complex system models complicates tremendously tasks such as parameter estimation via nonlinear optimization or statistical inference as the resulting inverse problem may require many successive executions of the model. This becomes particularly cumbersome for parameter-rich models. Emulation methods may help simplify parameter estimation by using a fast-running surrogate model of the original complex CPU-intensive process model. The next section reviews one of such emulation methods, namely PCE. Note, curly brackets are used to differentiate between random variables and their actual sampled values. Thus, $(\{\mathbf{x}_1\}, \ldots, \{\mathbf{x}_T\})$ stores a sequence of T different realizations (draws) of the model parameters, $\mathbf{x}$.

5.4 Model Emulation Using Generalized Polynomial Chaos Expansion

The PCE method emulates the output, $\mathbf{Y}$, of the original model in equation (5.1) via polynomial approximation. We can write this emulator as follows

$$\widehat{\mathbf{Y}} = \mathcal{F}_p(\mathbf{x}), \tag{5.3}$$

where p signifies the order of the polynomial, e.g. $p = 1$ for a first-order polynomial emulator. As detailed in the previous section, we ignore measurement uncertainty of the initial states and forcing variables, and thus assume that the model's ability to describe the observed data, $\widetilde{\mathbf{Y}}$, is determined only by the d-values of the model parameters, $\mathbf{x}$. Consequently, only the parameter values are needed as input by the emulator to replicate the response of the original model.

The PCE emulator approximates the model output in the following manner

$$\widehat{\mathbf{Y}} = \mathcal{F}_p(\mathbf{x}) = \sum_{j=1}^{m} a_j \Psi_j(\mathbf{x}), \tag{5.4}$$

where $\mathbf{a} = (a_1, \ldots, a_m)$ is a m-vector of, yet unknown, deterministic expansion coefficients, m signifies the number of d-dimensional orthogonal polynomials, and each $\Psi_j(\mathbf{x}) = \psi_{j,1}(x_1) \times \ldots \times \psi_{j,d}(x_d)$ is the product of one-dimensional polynomials for expansion terms $j = (1, \ldots, m)$. Thus, each model parameter, x_i, impacts Ψ_j in their own specific way. The number of orthogonal polynomials, m, depends on the polynomial order, p, of the emulator, and the dimensionality, d, of the parameter space, $\boldsymbol{\chi}$, according to $m = (p+d)!/(p!d!)$, where the symbol ! denotes factorial.

In this study, the values of the coefficients $\mathbf{a}$ are estimated via spectral projection of the original model response, $\mathbf{Y}(\mathbf{x})$, against each individual basis function [e.g see Xiu (2007)]

$$\widehat{\mathbf{a}}_j = \frac{\sum_{k=1}^{T} w_k \Psi_j(\{\mathbf{x}_k\}) \mathcal{F}(\{\mathbf{x}_k\}, \widetilde{\zeta}_0, \widetilde{\mathbf{B}})}{\langle \Psi_j^2(\{\mathbf{X}\}) \rangle}, \tag{5.5}$$

where $\{\mathbf{X}\} = (\{\mathbf{x}_1\}, \ldots, \{\mathbf{x}_T\})$ is a $T \times d$ matrix of T different realizations of the parameter vector, $\mathbf{x}$, and the w_k's denote the corresponding weights of an integration (cubature) rule on the feasible parameter space, $\boldsymbol{\chi}$, on $\mathbb{R}^d$. The denominator, $\langle \Psi_j^2(\{\mathbf{X}\})\rangle$ in equation (5.5), is equivalent to the inner product of the jth multivariate orthogonal polynomial, and can be computed as follows

$$\langle \Psi_j^2(\mathbf{X})\rangle = \prod_{i=1}^{d} \langle \psi_{j,i}^2 \rangle, \tag{5.6}$$

where $j = (1, \ldots, m)$. The d-univariate inner products, $\langle \psi_{j,i}^2 \rangle$, of each jth polynomial where $i = (1, \ldots, d)$ (right hand side of equation (5.6)), have a simple closed-form analytic solution. Therefore, the computational cost of PCE is determined by the time it takes to evaluate the forward model, $\mathcal{F}(\{\mathbf{x}_k\}, \widetilde{\zeta}_0, \widetilde{\mathbf{B}})$ for each of the T training parameter vectors, $\{\mathbf{x}_k\}$, of $\{\mathbf{X}\}$ in equation (5.5), where $k = (1, \ldots, T)$. If we now combine equations (5.5) and (5.6) then the approximate PCE emulator reads as follows

$$\widehat{\mathcal{F}}_p(\mathbf{x}) = \sum_{j=1}^{m} \widehat{a}_j \Psi_j(\mathbf{x}). \tag{5.7}$$

We refer interested readers to Xiu (2007) and Laloy et al. (2013) for a more detailed description of spectral projection methods within the context of PCE.

The use of an cubature integration rule in equation (5.5) causes $\mathcal{F}_p(\mathbf{x})$ in equation (5.4) to differ from $\widehat{\mathcal{F}}_p(\mathbf{x})$ in equation (5.7). This is also referred to as the *aliasing* error (Xiu, 2007) and decreases with T in equation (5.5). In theory, the larger the values of p and T, the better the emulator output, $\widehat{\mathbf{Y}} = \widehat{\mathcal{F}}_p(\mathbf{x})$, should approximate the simulations, $\mathbf{Y}(\mathbf{x})$, of the original forward simulator, $\mathcal{F}(\mathbf{x}, \widetilde{\boldsymbol{\zeta}}_0, \widetilde{\mathbf{B}})$ for random parameter vector, $\mathbf{x}$. The rate of convergence of $\widehat{\mathbf{Y}}(\mathbf{x})$ to $\mathbf{Y}(\mathbf{x})$ depends on the regularity of $\mathcal{F}(\mathbf{x}, \widetilde{\boldsymbol{\zeta}}_0, \widetilde{\mathbf{B}})$. The smoother the response of the forward model, the faster $\widehat{\mathbf{Y}}(\mathbf{x})$ will approximate $\mathbf{Y}(\mathbf{x})$. In practice, however computational requirements impose restrictions on the values of p and T. Thus, the available computational budget dictates the accuracy of the

surrogate model, $\widehat{\mathcal{F}}_p$.

To maximize the capabilities of the emulator, it would be desirable to select wisely, $\{\mathbf{X}\}$, the matrix of T training samples. Spectral projection has the advantage that it requires far fewer parameter vectors than linear regression for a well-posed solution of the PCE expansion coefficients in equation (5.7). The next section reviews briefly sparse grid methods to generate the matrix of training realizations.

5.4.1 Sparse grid approximation

Sparse grid approximation methods reduce drastically the required number of parameter vectors to construct an emulator, while preserving a high accuracy for moderately large dimensional parameter spaces (Nobile et al., 2008a,b; Smolyak, 1963; Xiu and Hesthaven, 2005). A pth order PCE emulator can be built with sparse grids of different levels, L, and thus variable number of training data points. Thus, a sparse grid of level L does not necessarily equate to a $p = L$th order PCE emulator. In general, the larger the order of the polynomial emulator the more training points are required for its calibration. Yet, this does not guarantee that the emulator will closely approximate the output of the original forward model. Thus, we build, test, explore and contrast emulators of different polynomial orders using a range of sparse grid levels.

Figure 5.1 depicts graphically training data pairs, $(\{x_1\}, \{x_2\})$ for a hypothetical two-parameter model. We used a Gauss-Patterson sparse grid and separately present the results for levels $L = 1$ up to $L = 6$. The higher the order, L of the sparse grid, the more training data points. Indeed, with $L = 1$ only $TR = 5$ samples are required to build the PCE emulator, and this number increases to $TR = 769$ for $L = 6$.

Table 1 lists the number of training samples associated with sparse grids of levels $L = 1$, $L = 2$, and $L = 3$ for the three forward models considered in this paper. Details of these models appear in the case study section, and readers are referred to this part of the paper for further information.

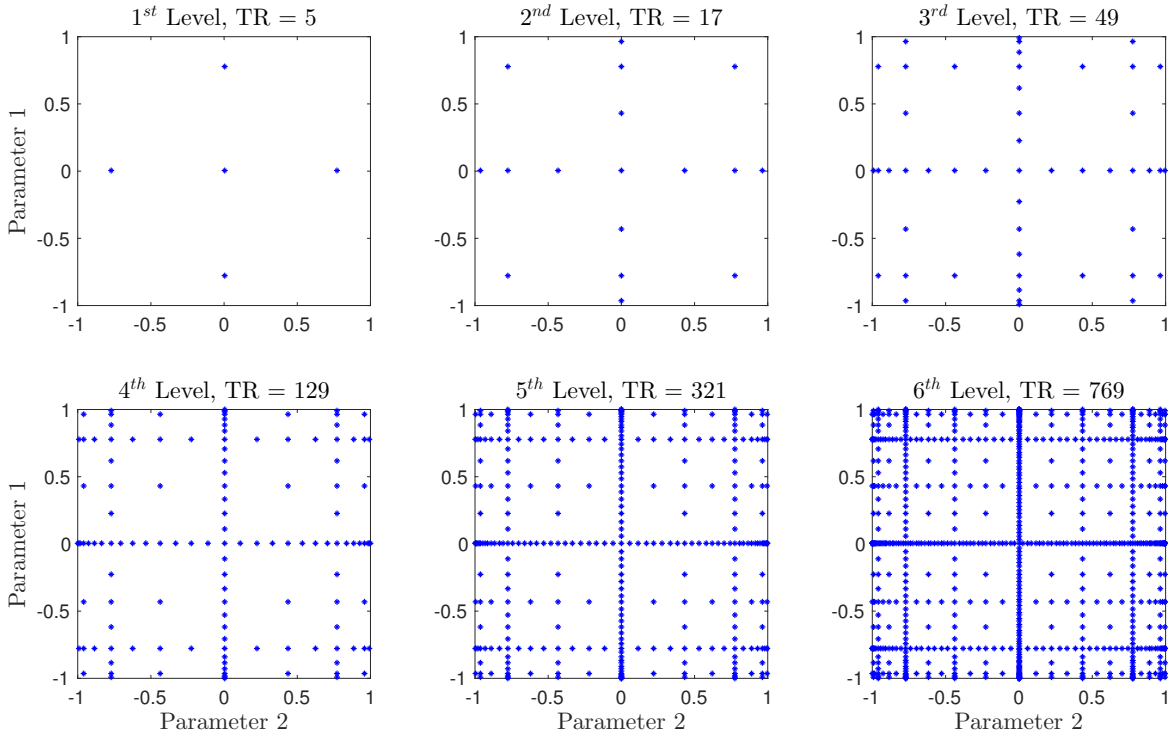

Figure 5.1: Training data points of a hypothetical two-parameter model using a Gauss-Patterson sparse grid of level. Top: Left to Right - $L = 1$, $L = 2$, $L = 3$. Bottom: Left to Right - $L = 4$, $L = 5$, and $L = 6$. We list separately in each graph the total number of training data points, also coined TR.

The first case study considers a $d = 7$-parameter conceptual watershed model, named hmodel. The first level sparse grid of this model corresponds to $T = 15$ TR's, whereas a second and third level grid increases the number of training points to $T = 127$ and $T = 799$, respectively. The CPU-efficiency of the model warrants emulation using all three different grid levels.

The second case study involves simulation of water transport using a soil-tree-atmosphere continuum, or STAC model. This model has $d = 15$ different parameters. A first level sparse grid would require only $T = 31$ TR's, and this number increases rapidly to $T = 511$ and $T = 5,925$ for a second and third order level. Due to the somewhat large CPU-demands of STAC, we limit construction of the emulator to a first and second level sparse grid.

Lastly, the third case study considers CLM. This model is rather complex, yet we consider herein only the $d = 87$ parameters of interest to dynamic vegetation modeling. A first order

sparse grid equates to $T = 175$ TR's, whereas the second and third level grids necessitate $15,487$ and $T = 2,044,416$ executions of CLM. We deliberately use $L = 1$, that is, a sparse grid of level one, as the second and third grid levels demand too much time. Indeed, a single forward run of CLM takes about 8 hours using serial processing. We also consider separately a refined CLM parameterization, in which only the ten most sensitive parameters are allowed to vary. This alternative parameterization, coined CLM-10, is desirable as it requires far fewer model evaluations for each of the three sparse grids.

In summary, we consider all three levels of the sparse grid for the hmodel, use $L = 1$ and $L = 2$ for STAC, and $L = 1$ for CLM. Furthermore, for the hmodel and STAC model, we examine polynomial emulators of first, second, and third order $p \in (1, 2, 3)$. We coin these emulators "Lxpyy", where "x" denotes the sparse grid level, and "yy" signifies the polynomial order of the emulator. Thus, L1p1 signifies a first-order polynomial emulator built using a sparse grid of level one. L2p3 thus is equivalent to a third-order polynomial emulation using a second order sparse grid.

Table 5.1 Total number of training runs (TR's) of a Gauss-Patterson sparse grid of level L for each of the three system models considered in the case study section.

	# of TR			
Sparse Grid	hmodel 7 pars	STAC 15 pars	CLM-87 87 pars	CLM-10 10 pars
1^{st} Level	15	31	175	21
2^{nd} Level	127	511	15,487	241
3^{rd} Level	799	5,925	2,044,416	2,001

5.4.2 Summary statistics

We use the root-mean-square-error (RMSE) and the correlation coefficient (CORR) to quantify the level of agreement between the emulator output, $\widehat{\mathbf{Y}}(\mathbf{x}) = (\widehat{y}_1(\mathbf{x}), \ldots, \widehat{y}_n(\mathbf{x}))$, and the output, $\mathbf{Y}(\mathbf{x}) = (y_1(\mathbf{x}), \ldots, y_n(\mathbf{x}))$, of the original forward model. These two summary statistics are also used to determine the goodness-of-fit between the emulator or model output and the observed data,

$\widetilde{\mathbf{Y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$, and can be calculated as follows

$$\begin{aligned} \text{RMSE}(\mathbf{a}, \mathbf{b}) &= \sqrt{\frac{1}{n}\sum_{t=1}^{n}(a_t - b_t)^2} \\ \text{CORR}(\mathbf{a}, \mathbf{b}) &= \frac{\frac{1}{n}\sum_{t=1}^{n}(a_t b_t) - m_{\mathbf{a}} m_{\mathbf{b}}}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(a_t - m_{\mathbf{a}})^2}\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(b_t - m_{\mathbf{b}})^2}}, \end{aligned} \tag{5.8}$$

where $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ are n-vectors with the simulated, emulated or observed data, and $m_{\mathbf{a}}$ and $m_{\mathbf{b}}$ store the mean value of $\mathbf{a}$ and $\mathbf{b}$, respectively.

5.5 Results

We present the results of our case studies in three sections, each discussing separately the results of a specific model. These sections follow a similar format. First, we discuss the model of interest, followed by a description of its parameters. We then present the emulation results using different polynomial emulators and sparse grid levels.

We present the results of our case studies in three sections, each discussing separately the results of a specific model. These sections follow a similar format. First, we discuss the model of interest, followed by a description of its parameters. We then present the emulation results using different polynomial emulators and sparse grid levels.

5.5.1 Case study 1: The rainfall-runoff transformation

The first case study simulates the rainfall-runoff transformation of the Guadalupe River at Spring Branch in Texas using the hmodel conceptual watershed model. The model transforms rainfall into discharge at the watershed outlet using explicit process descriptions of interception, throughfall, evaporation, surface runoff, percolation, and surface and subsurface routing. A detailed description of the hmodel structure and processes representations can be found in Schoups and Vrugt (2010).

This manuscript also summarizes the $d = 7$ parameters of the hmodel and their prior uncertainty ranges.

A six-year record of daily discharge (mm/day), mean areal precipitation (mm/day), and mean areal potential evapotranspiration (mm/day) of the Guadalupe River was used for hmodel parameter estimation using Markov chain Monte Carlo (MCMC) simulation with the DREAM algorithm (Vrugt, 2016). This data was derived from the MOPEX data set Duan et al. (2006). We assumed a uniform prior distribution for the hmodel parameters, and implemented a Gaussian likelihood function. A 365-day spin-up period was used to reduce sensitivity to state-value initialization. Details about the basin, experimental data, prior distribution and likelihood function can be found in Schoups and Vrugt (2010).

Table 5.2 summarizes the results of our analysis and lists the CPU-time and RMSE of the calibration and evaluation data period using first, second, and third order polynomial emulators with sparse grids of level one, two, and three. We separately also report the average run time of a single hmodel evaluation.

The calibration of the original model cost a total of 8,000 secs (2.2 hrs), which is not too expensive considering the accuracy of the results. For the emulators, we must examine accuracy under both the 'Calibration' and 'Evaluation' runs to assure that calibrated parameter values are reasonable. We see in Table 2 that the 'Calibration' runs all performed very well, since the emulators were specifically calibrated to the observed data and thus achieved a good fit. However, for the 'Evaluation' runs we see discrepancies between the original model simulations and the observed data. The emulators with higher polynomial orders (i.e. L1p3, L2p3, and L3p3) provided parameter sets that performed the best with the original model, with RMSE's of 0.68, 1.03, and 0.64 respectively, indicating that the calibration of the hmodel emulators obtain better parameter values when a higher order polynomial is used. Interestingly, the L2p1 emulator performed better than the L2p3 emulator in the 'Evaluation' run, indicating that the calibration of the hmodel emulator

with sparse grid of level 2 obtained the best parameter values when using a linear polynomial of order 1 instead of a polynomial of order 3.

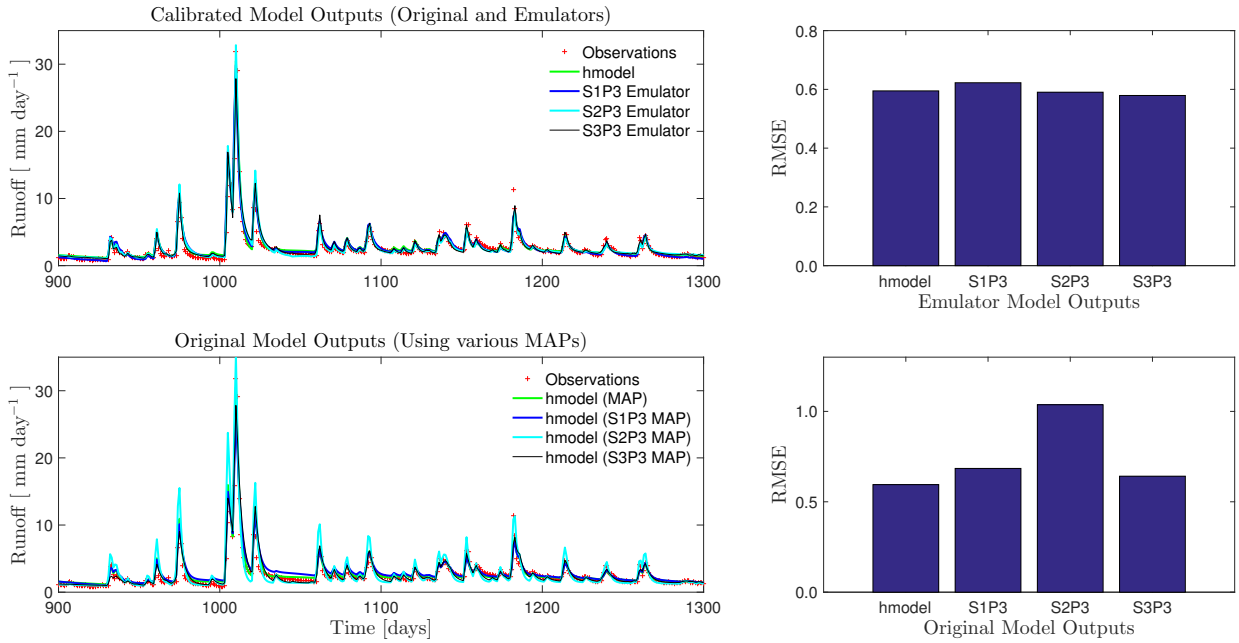

Figure 5.2: Simulation results for the hmodel and a selection of its emulators. The panels on the left show the model simulations compared with the observations. The top panels shows the direct calibration result, and the bottom panels shows the original hmodel output when simulated with each respective parameter set. RMSE values on the right are also shown in Table 2.

The model outputs for various runs of the hydrologic model and the p3 emulators are shown in Figure 2. The top panel shows simulations of each calibration, meaning the results of calibrating the original model and the emulators. The bottom panel shows the simulations of the original model, but when using the parameter sets obtained with each calibration. This is an important test to check, because in theory an emulator can be used for calibration and obtain accurate simulation results, but when that corresponding parameter set is ran with the original model the simulation results maybe totally unrealistic. Therefore, this test allows a secure assessment of the parameter values obtained from the calibrated emulators.

The calibration run time cost for the hmodel emulators were all 'cheaper' than the original model calibration (see column under 'Calibration' in the 'Run-Time' section of Table 2), with the exception of the L3p3 emulator which cost 12,400 sec (3.4 hrs). However when considering the training cost of the emulators, the L3p2 emulator calibration also had a higher CPU cost than the

Table 5.2: Total run time and accuracy results for the hmodel. The listed CPU-time (under 'Total') of the emulator includes overall cost of training (under 'Training') and calibration (under 'Calibration'). The Root-Mean-Square Error (RMSE) for each calibration strategy compared to the observed data is shown. Results of calibrating the emulator are shown under 'Calibration' and the results of running the original model with the parameter set obtained with each calibration is shown under 'Evaluation'. Each MCMC run spanned 1,000 generations and utilized a total of 8 chains. Therefore a total of 8,000 simulations were made for calibrating each of the hmodel and its emulators.

	CPU time (sec)				RMSE (mm/day)	
Model	Single Run	Training	Calibration	Total	Calibration	Evaluation
Original hmodel	1	N/A	8,000	8,000 (2.2 hrs)	0.59	0.59
Sparse Grid Level 1						
L1p1 hmodel	0.2	15	1,600	1,615 (27 min)	0.60	2.50
L1p2 hmodel	0.4	15	3,200	3,215 (54 min)	0.61	1.70
L1p3 hmodel	0.7	15	5,600	5,615 (1.5 hrs)	0.62	0.68
Sparse Grid Level 2						
L2p1 hmodel	0.25	127	2,000	2,127 (35 min)	0.64	0.86
L2p2 hmodel	0.45	127	3,600	3,727 (1 hr)	0.62	1.40
L2p3 hmodel	0.75	127	6,000	6,127 (1.7 hrs)	0.59	1.03
Sparse Grid Level 3						
L3p1 hmodel	0.65	799	5,200	5,999 (1.6 hrs)	0.67	1.00
L3p2 hmodel	0.95	799	7,600	8,399 (2.3 hrs)	0.62	0.73
L3p3 hmodel	1.55	799	12,400	13,199 (3.6 hrs)	0.58	0.64

original calibration (see column under 'Total' in the 'Run-Time' section of Table 2).

5.5.2 Case study 2: Soil-Tree-Atmosphere Continuum (STAC) model

The second case study involves the STAC model, which is a physically-based nonlinear modeling framework (Bohrer et al., 2005; Chuang et al., 2006; Kumagai, 2001) that simulates water flow in the combined soil and tree systems (Rings et al., 2013; Siqueira et al., 2008). The STAC model discretizes the system domain and couples the soil with the tree domain, simulating the soil, roots and tree trunk as a continuum. Water flow is driven by water potential gradients along the coupled system (Bittner et al., 2012) with spatially distributed root water uptake and canopy transpiration sink terms. The STAC model utilizes the HYDRUS model (Simuunek et al., 2008), where water flow through the soil and the tree root system and stem is driven by the evaporative demand and soil-available water, leading to a gradient in soil and xylem water potentials along the STAC. The STAC model contains 15 parameters, and we build emulators with 1^{st} and 2^{nd} level sparse grid sampling (Table 1).

The STAC model uses Richards' equation (Richards, 1931) to solve the unsaturated water flow across the soil-tree domain. Each layer in the model domain can be represented by its own set of equations in a axisymmetrical, two-dimensional, isotropic form. The Richards' equation for the soil domain is as follows:

$$\frac{\partial \theta_{\text{soil}}}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rK_r(h)\frac{\partial h}{\partial r}\right) + \frac{\partial}{\partial z}\left(K_z(h)\frac{\partial h}{\partial z}\right) - \frac{\partial K_z(h)}{\partial z} - W_{\text{soil}}(h, r, z) \tag{5.9}$$

where θ_{soil} (L^3 L^{-3}) is the volumetric soil water content, $K(h)$ (L T^{-1}) defines the unsaturated hydraulic conductivity function (further denoted by either $r-$ for radial direction or $z-$ for vertical direction), h (L) is the soil water pressure head, r and z are the lateral and vertical coordinates

(positive downwards) of the soil domain respectively, t (T) is time, and W_{soil} (L^3 L^{-3} T^{-1}) defines a sink/source term that quantifies spatially distributed root water uptake from the soil. Both K and W_{soil} are functions of θ and/or h. The subscripts r and z allows for the possibility to include soil anisotropy, i.e., to simulate water flow with the unsaturated hydraulic conductivity function being different for the $r-$ and $z-$direction.

Additionally, the set up of Richards' equation for the tree domain to represent flow through the canopy is similar to that of the soil domain in Equation 5.9, but in one-dimensional form. This equation is derivable directly from Equation 5.9 by reducing to one dimension, z only. Thus, the axi-symetrical flow through the canopy is represented by:

$$\frac{\partial \theta_{\text{tree}}}{\partial t} = \frac{\partial}{\partial z}\left(K_z(h)\frac{\partial h}{\partial z}\right) - \frac{\partial K_z(h)}{\partial z} - W_{\text{tree}}(h, z) \tag{5.10}$$

where θ_{tree} (L^3 L^{-3}) is the volumetric tree water content , $K(h)$ (L T^{-1}) defines the unsaturated hydraulic conductivity function (further denoted by $z-$ for vertical direction), h (L) is the tree water pressure head, z is the vertical coordinate of the tree domain (positive downwards), t (T) is time, and W_{tree} defines a sink/source term (L^3 L^{-3} T^{-1}) that quantifies spatially distributed canopy transpiration.

Calibration data include soil water content and water potential in three spatial dimensions in the root zone, tree stem water content and sapflux, canopy water potential, and atmospheric variables such as net radiation, air temperature and humidity providing the necessary information for potential tree evapotranspiration. Three sapflux sensors (TransfloNZ, Palmerston North, NZ) were installed into the sapwood at a trunk height of 2.5 m. Then, stem water potential measurements were taken from needle stems of various lower tree branches, at about 6 m from the ground. Additionally, Echo- 5TE soil moisture sensors were installed at depths of 0.15 m, 0.30 m, 0.60 m and 0.90 m in each of six locations within a 5 m radius from the tree trunk. The data was selected for an 18-day rainless period in summer of 2009, starting July 15, and were collected in and around

a mature Douglas white fir (Abies concolor).

There are at least m=4 outputs for the STAC model, including sapflux, stem potential, soil storage, and tree storage, however for the purposes of this study we focus solely on the sapflux output (in [cm day^{-1}]), and we attempt to emulate the response for n=817 time steps representing 30-min intervals during an 18-day period.

In this second case study, several emulators are built for the STAC model, requiring a total of 31 and 511 training sets for the sparse grid level 1 and 2 emulators, respectively. The original and emulator models are simulated and calibrated to an observed data set spanning 18+ days. The calibration of the original model cost a total of 600,000 secs (or roughly 1 week), which is rather expensive since the calibration results cannot be obtained until the following week.

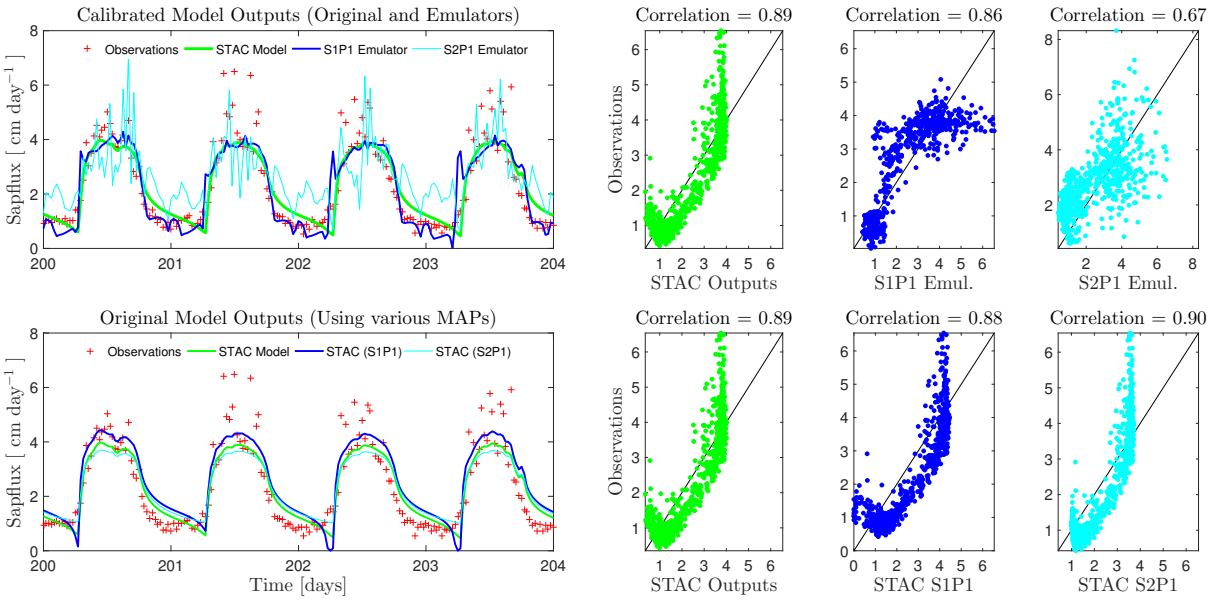

Figure 5.3: Simulation results for the STAC model and a selection of its emulators. The panels on the left show the model simulations compared with the observations. The top panel shows the direct calibration result, and the bottom panel shows the original STAC model output when simulated with each respective parameter set. Correlation figures on the right show how closely the simulation dynamics track the observations, and the scatter plots allow for the comparison between each model and respective parameter set.

Table 3 shows the results of calibrating the STAC model and its emulators. Under 'Calibration' we see that the accuracy of calibrating the emulators gets progressively worse with increasing polynomial order. For example, the L1p1 emulator calibration performed better than the L1p2 and L1p3 emulators (RMSE of 0.77 vs 0.83 and 1.59 respectively), and the L2p1 emulator calibration performed better than the L2p2 and L2p3 emulators (RMSE of 1.20 vs 1.54 and 1.56 respectively). This is also shown for the 'Evaluation' runs; the L1p1 and L2p1 emulator calibrations provided the best overall parameter sets when ran through the original STAC model (RMSE's of 0.80 vs

Table 5.3: Total run time and accuracy results for the STAC model. The listed CPU-time (under 'Total') of the emulator includes overall cost of training (under 'Training') and calibration (under 'Calibration'). The RMSE for each calibration strategy compared to the observed data is shown. Results of calibrating the emulator are shown under 'Calibration' and the results of running the original model with the parameter set obtained with each calibration is shown under 'Evaluation'. Each MCMC run spanned 2,500 generations and utilized a total of 8 chains. Therefore a total of 20,000 simulations were made for calibrating each of the STAC and its emulators.

	CPU cost (sec)				RMSE	
Model	Single Run	Training	Calibration	Total	Calibration	Evaluation
Original STAC	30	N/A	600,000	600,000 (1 week)	0.68	0.68
Sparse Grid Level 1						
L1p1 STAC	0.20	930	4,000	4,930 (3 hrs)	0.77	0.80
L1p2 STAC	0.45	930	9,000	9,930 (4.6 hrs)	0.83	0.93
L1p3 STAC	1.6	930	32,000	32,930 (10.9 hrs)	1.59	2.34
Sparse Grid Level 2						
L2p1 STAC	0.6	15,330	12,000	27,330 (36 hrs)	1.20	0.74
L2p2 STAC	1.05	15,330	21,000	36,330 (40 hrs)	1.54	0.88
L2p3 STAC	3.05	15,330	61,000	76,330 (52 hrs)	1.56	1.23

0.93 and 2.34 for the L1 emulators, and 0.74 vs 0.88 and 1.23 for the L2 emulators). These results indicate that the calibration of the STAC model emulators obtain better parameter values when a first order polynomial is used for the two considered sparse grid levels.

The model outputs for various runs of the STAC model and its emulators of polynomial order 1 (P1 emulators) are shown in Figure 3. The top and bottom panels are shown similarly to Figure 2, where results of calibrating the original model and the two emulators are shown in the top panel, and simulations of the original model when using the parameter sets obtained with each calibration are shown in the bottom panel. Again, this test is important to check for realistic model outputs given the parameter values obtained from calibration of the emulator. On the right, correlation scatter plots show how closely the simulation dynamics track the observations, and the correlation with the observed data is included with each run. Apparently, the L2p1 emulator (one of the more expensive emulators built for the STAC model) provided a parameter set that produced simulations from the original model that had the highest correlation with the observed data (= 0.9), which was even higher than the correlation of the original model calibration (= 0.89). Yet, even a much 'cheaper' emulator such as the L1p1 provided a parameter set that produced simulations from the original model with a very high correlation (= 0.88) with the observed sapflux data.

The calibration run time cost for the STAC model emulators were all at least an order of magnitude 'cheaper' than the original model calibration (see column under 'Calibration' in the 'Run-Time' section of Table 3). Interestingly, all the emulator calibrations had a much lower CPU cost than the original calibration, even when considering the training cost of the emulators which significantly increases the overall CPU cost (see column under 'Total' in the 'Run-Time' section of Table 2). These results indicate that calibration of the STAC model can be achieved in a much shorter time if an emulator is used for the calibration process, since the original model calibration cost over 1 week of CPU cost and the calibration of the emulators were anywhere between a few hours to a few days.

5.5.3 Case study 3: the Community Land Model with Ecosystem Demography, CLM4.5(ED)

The ESM considered in this study is the CLM4.5(ED). CLM is a community-based open-source model that is widely used for understanding climate-vegetation interactions. CLM is the land surface model used within various ESM's, including the Community Earth System Model (CESM) and the Norwegian Earth System Model (NorESM) (Bonan et al., 2011; Lawrence et al., 2011). The Ecosystem Demography (ED) concept is a method for scaling the behavior of forest ecosystems by aggregating individual trees into representative 'cohorts' based on their size and plant type or PFT, and by aggregating groups of cohorts into representative 'patches' (conceptually similar to a forest plot) which explicitly tracks the time between disturbances (Moorcroft et al., 2001). The ED component is the most advanced DGVM incorporated into the CLM framework. The main property of the ED concept that differs from most commonly used 'big-leaf' models is the capacity to predict distributions and compositions of plants directly from their given physiological traits described by the model parameterization (Fisher et al., 2015).

CLM(ED) simulates growth by integrating photosynthesis across different leaf layers for each cohort, and mechanistic mortality is simulated based on plant carbon starvation and hydraulic failure in addition to a background mortality rate, mortality from tree-fall impacts, and fire. The model allocates photosynthetic carbon to different tissues such as leaf, root and stem based on the allometry of different tree species. CLM(ED) can be simulated at different modes including point mode for sites, regional mode for watershed or regional scale, and global mode for continental and global scale. See supplementary model description in Fisher et al. (2015) for details on specific components of the model structure. The CLM(ED) version considered here contains 87 parameters, making it only acceptable to build an emulator with 1^{st} level sparse grid sampling, but for a 2^{nd} level sparse grid emulator the number of training runs already become too large.

For this analysis, the runs are initialized (i.e. $\widetilde{\zeta}_0$) with a bare ground, or a state with no vegetation. The climate conditions (i.e. $\widetilde{\mathbf{B}}$) for this site are from Qian et al. (2006) representative of data from 1948-1972. The model is simulated for a site in the State of Pará, The Amazon, Brazil (7° S, 55° W). There are dozens of possible outputs for the CLM, and here we focus on m=3 outputs, including Gross Primary Production (GPP [kgC m^{-2} yr^{-1}]), Leaf Area Index (LAI [-]), and overall biomass (Biomass [kgC m^{-2}]). We attempt to emulate the response for n=960 time steps representing 1-month intervals during an 80-year period.

5.5.4 87-parameter emulator

We build an L1p1 emulator for the 87-parameter CLM, requiring a total of 175 training sets. The simulations spanned 80 years and the CPU cost of the original model is roughly 8 hrs, therefore the emulator training cost a total of 1,400 CPU hrs (performed in parallel simulations on the Conejo supercomputer at the Los Alamos National Laboratory). After obtaining the training runs, a separate emulator is created for each output since any model output ordinarily has its own response surface. To test the emulator, we compared the original model simulation with its surrogate counterpart using a parameter set representing the default values for Broadleaf Evergreen Tropical trees currently implemented in CLM. The various emulators built for GPP, LAI, and biomass are shown with the original model simulations in Figure 4.

The black dotted line in Figure 4 shows the emulation results for the 87-parameter emulator, which significantly deviate from the original model outputs shown with a red line. The 87-parameter emulator of the CLM model does not capture the dynamics of the original model, and it seems that at many time steps the emulator outputs are largely affected by noise. Figure 4 also shows scatter plots of 100 various simulations with randomly drawn parameter values, using both the original CLM and the 87-parameter emulator. The fit between the two models is not good for any of the outputs, shown in the scatter plots with black dots. These simulations have correlations of 0.50

for GPP, 0.29 for LAI, and 0.47 for biomass. Clearly, this emulator will not serve the purposes of calibrating the CLM.

Table 5.4 Total run time results for the CLM model. The listed CPU-time of the emulator includes overall cost of training (under 'Training') and simulation (under 'Single Simulation').

	CPU cost (time)	
Model	Single Simulation	Training
Original CLM	8 hrs	N/A
87-Par Emulator		
L1p1	2.4 secs	1,400 hrs
10-Par Emulator		
L1p1	0.2 secs	168 hrs
L2p1	0.4 secs	1,928 hrs

5.5.5 Reducing to a 10-Parameter Emulator

Since the bad emulator results shown with the black dotted line in Figure 4 are caused by applying PCE emulators for a 87-parameter model, we hypothesize that reducing the dimensionality of the emulator will reduce the noise at each projection which might allow the emulator to better mimc the original model. According to Massoud et al. (2017, In Review), only a handful of parameters generally control the outputs of this CLM version, and shown in their study are the 10 most sensitive parameters for various outputs. In this study, we re-construct the emulator for CLM focusing on the 10 most influential parameters (rather than the complete set of 87 parameters). Thus, L1p1 and L2p1 emulators are built using training samples from the 10 most sensitive parameters, requiring a total of 21 and 241 training runs, respectively.

The dimensionality reduction applied to the emulators significantly improves their outputs when compared to the original model. The results of the 10-parameter L1p1 (blue) and L2p1 (green) emulators for CLM are shown in Figure 4. Again, these simulations use the default parameter set for the Broadleaf Evergreen Tropical tree, and the results of these emulators follow the original model

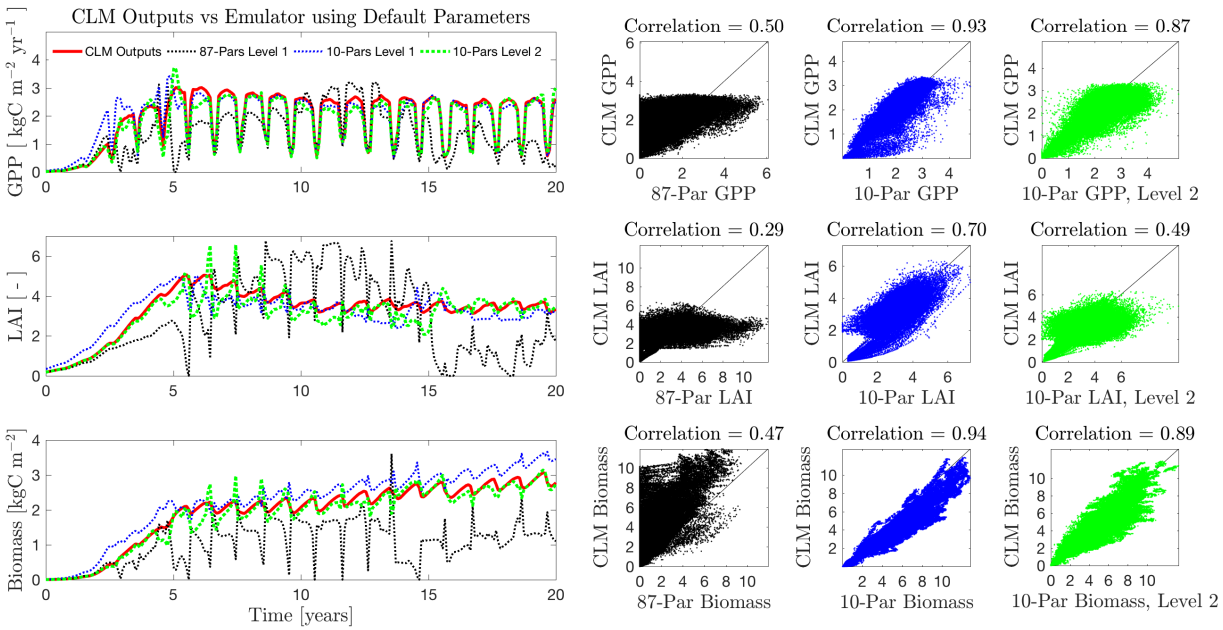

Figure 5.4: Simulation results for the CLM model and its emulators. The panels on the left show the original model outputs compared with the emulator simulations. The top panel shows results for GPP, the middle panel shows results for LAI, and the bottom panel are the results for biomass outputs. Correlation figures on the right show how closely the simulation dynamics of each emulator track the original model outputs using 100 randomly drawn parameter sets.

outputs much closer than the 87-parameter emulator. The right panel of Figure 4 shows scatter plots of 100 various simulations with randomly drawn parameter values, using both the original CLM and the 10-parameter emulators. Note, these 100 parameter vectors draw randomly the 10 influential parameters and keeps the remaining 77 "non"-influential parameters at their default value. The fit between the two models is fairly good for all of the outputs, with correlations of 0.93 for GPP, 0.70 for LAI, and 0.94 for biomass for the L1p1 emulator, and with correlations of 0.87 for GPP, 0.49 for LAI, and 0.89 for biomass for the L2p1 emulator. Clearly, these emulators perform much better than the 87-parameter emulator, and can possibly serve the purposes of calibrating the CLM, but only for the sensitive parameters that are considered in building the emulator.

Calibration of a model such as CLM requires thousands of runs, resulting in a CPU cost that is not feasible. However, if the 10-parameter emulators can produce similar output response as the original CLM, then the less complex version of the original model can be used for calibration,

which would cost a fraction of the total run time.

5.6 Discussion

In this paper, we investigate the applicability of emulation for the calibration of complex models, which are often impossible to calibrate in practice. To address this, we build emulators of two environmental models (hmodel and STAC model) and compare the calibration of these emulators with their original counterparts. Results build confidence in the calibration of model emulators as a replacement for the original model, and this also might be true for a more complex and higher dimensional model such as an ESM (here we focus on the CLM).

In the following sections, we address the accuracy of the simulations, where we compare the RMSE of the original model and emulator simulations with observed data. Then we compare the efficiency of each method, where we compare CPU costs of each attempt. Furthermore, we check for the fidelity of the emulator to the original model, where we compare the sensitivity of the model parameters in the emulators to those of the original model. We finish with a brief discussion on how the method presented in this paper can be applied for the calibration of ESMs.

5.6.1 Accuracy

To check the accuracy of each method, we must check two metrics. First, we check the RMSE of the calibration run, which shows how well each calibration method performed (under 'Calibration' in Tables 2 and 3). Then, we check the RMSE of the Evaluation run, which shows how well the original model performs with each calibrated parameter set (under 'Evaluation' in Tables 2 and 3), which again is a test to check for realistic model outputs given the parameter values obtained from calibration of the emulators.

For the hmodel, calibrating the original model resulted in a similar fit to the data as the

calibration of the emulators, with RMSE's that are nearly identical. However, when the parameter sets obtained with the emulator calibrations are used as input to the original model (i.e. Evaluation runs), the model simulations deviate a bit from the observed data (RMSE's ranging from 0.64 to 2.50, compared with RMSE of original model calibration of 0.59). Overall, it seemed that the higher the polynomial order used for the hmodel emulators the better the accuracy of the Evaluation runs. In other words, the L1p3, L2p3, and L3p3 hmodel emulators generally obtained the best parameter values.

For the STAC model, calibrating the original model resulted in the closest fit to the data, with an RMSE of 0.68. After calibration, the first order polynomial emulators (L1p1 and L2p1) had RMSE's of 0.77 and 1.20, respectively. Moreover, the parameter sets obtained with the L1p1 and L2p1 calibration fit the observed data rather well when validated with the original model. In the Evaluation runs, the L1p1 and L2p1 parameters result in RMSE's of 0.80 and 0.74, respectively, which is quite similar when compared to the original model run. Additionally, the correlation of the original model runs (= 0.89) to the observed data was very similar to the L1p1 and L2p2 Evaluation runs (correlations of 0.88 and 0.90 respectively), indicating that the simulation dynamics of the emulators track quite closely the dynamics seen in the observations. Therefore, it seems that first order (P1) emulators are the best choice for calibration of the STAC model.

For CLM, the 87-parameter emulator did not perform well, whereas simulations using the reduced 10-parameter emulators matched most of the outputs from the original model. Figure 4 shows these simulations and we argue that, overall, the 10-parameter emulators for CLM can be trusted to mimic the original model outputs with high accuracy, especially for the L1p1 emulator.

5.6.2 Efficiency

The main goal of using emulators is to save computation cost. Therefore, we define efficiency here in the context of run time efficiency. Tables 2 and 3 highlight the CPU costs of training

and calibrating the emulators for both the hmodel and STAC models, compared to calibrating the original model.

For the hmodel, calibration of the original model takes 8,000 secs. Although the total time to train and calibrate the emulators is generally lower than calibrating the original model (with the exception of the L3 emulators), the fit to the observed data is almost the same as the original model in the 'Calibration' runs. However, for the 'Evaluation' runs simulations start to deviate from the observations. The higher order polynomial emulators (L1p3, L2p3, and L3p3) achieve better fits than the lower order polynomial emulators, but with a slightly higher CPU cost. So we conclude that for a simple model with non-linear dynamics in its outputs such as the hmodel, calibration of the original model will suffice (and in fact is the most accurate), yet a third order polynomial emulator for this model produces similar results in a shorter time period than the original model, and thus can still provide benefits to the calibration problem.

For the STAC model, calibration of the original model takes over 1 week, and results in the best RMSE compared to the emulators, but not the highest correlation with the observed data which was achieved by the L2p1 emulator Evaluation run. The total time to train and calibrate the emulators, however, is much lower than calibrating the original model, ranging from a few hours to a few days for the emulators compared with 1 week for the original model. Additionally, the fit to the observed data is almost as good as the original model (see RMSE's in Table 3). So we conclude that for a medium-complexity model such as the STAC model, calibration of the original model might take too long and would be rather expensive in regards to CPU costs. The Evaluation runs of the first order polynomial emulators for the STAC model (i.e. L1p1 and L2p1) produce accurate results in a fraction of the time. Therefore, we assume that for the STAC model the calibration of a L1p1 or L2p1 emulator might be more efficient than calibrating the original model, since similar results are achieved but with a much lower CPU cost.

For the CLM, the comparison of CPU-cost is simple. Running the original model for an 80 year

simulation requires roughly 8 hrs of runtime on a super computing machine with parallel computing capabilities, whereas the emulator can produce similar outputs in just a few seconds on a personal laptop. This is indeed quite beneficial.

5.6.3 Fidelity

Lastly, we investigate the fidelity of the emulators to their original model counterpart, where we compare the sensitivity of the model parameters in the emulators to those of the original models. In theory, if the performance of the emulator is identical to that of the original model, then the parameters of the emulator should have the same sensitivity index as those of the original model. Many studies assume that emulators provide efficient sensitivity analysis for large and expensive models (Crestaux et al., 2009; Ratto et al., 2012), however this may not always be true as emulators may provide anomaluous parameter sensitivity estimates. In one study, the authors investigated the joint application of emulators and their respective parameter sensitivities, and showed that with enough training the emulator will match the original model in the parameter sensitivity rank and magnitudes (Borgonovo et al., 2012). However, the environmental model used in Borgonovo et al. (2012) only contained 12 parameters, and such a test for a higher-dimensional model such as CLM does not exist in the literature and is warranted. To this end, we assess parameter sensitivities of the original STAC and CLM models and compare results with the parameter sensitivities of the emulators. In this study we investigate the claim that it is safe to assume the sensitivity index obtained from the emulator represents the true importance of that parameter in the original model. This will indicate that the processes represented in the original model are being replicated properly in the emulator.

We use the Fourier Amplitude Sensitivity Testing (FAST) method of Xu and Gertner (2007, 2011) to evaluate the parameter sensitivities of the original model and the emulators. We test for the 5 most sensitive parameters of the L1p1 and L2p1 emulators of the STAC model and compare

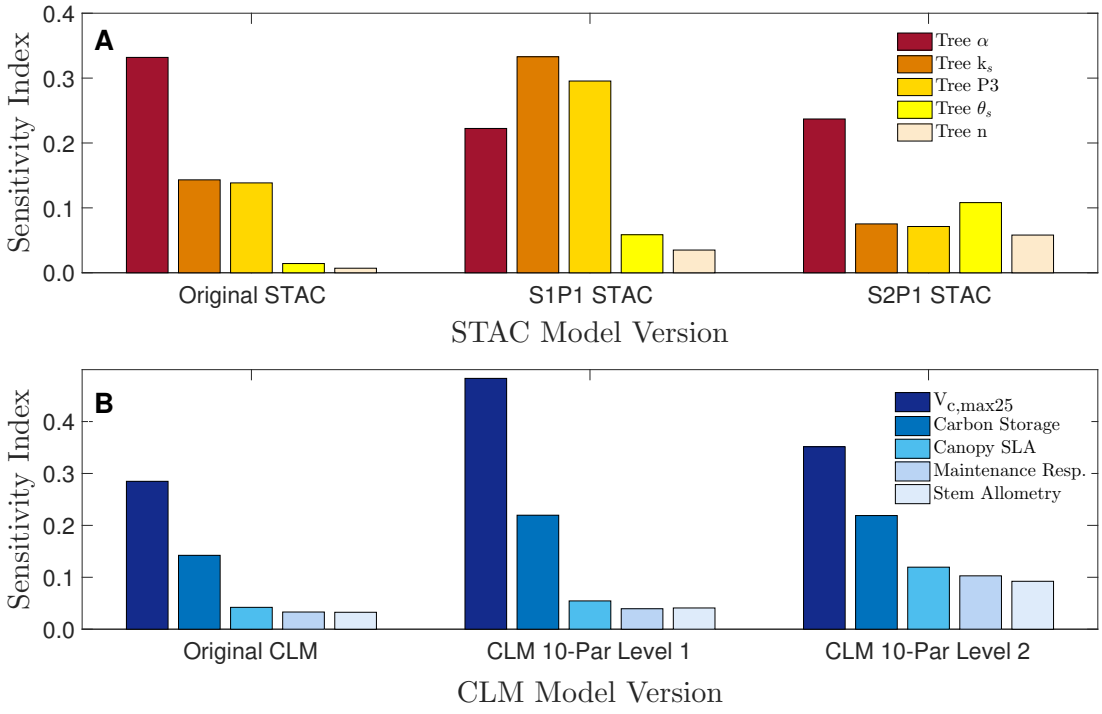

Figure 5.5: Panel A: Top 5 sensitive parameters for the sapflux output of the original STAC model and its L1p1 and L2p1 emulators. Panel B: Top 5 sensitive parameters for the GPP outputs of the original CLM model and its 10-parameter emulators (Levels 1 and 2).

with those of the original model (Figure 5A). The parameters included are several Van Genuchten parameters (Van Genuchten, 1980) that represent the hydraulic properties of the tree layer in the Soil-Tree-Atmosphere continuum domain (i.e. Tree α, Tree K_s, Tree θ_s, and Tree n) as well as a Feddes hydraulic stress parameter (Feddes et al., 1978) (i.e. Tree P3). The most sensitive parameter for the STAC model sapflux output is the Tree α parameter which matches the most sensitive parameter for the L2p1 emulator, however for the L1p1 emulator this parameter is the third most sensitive after the Tree K_s and Tree P3 parameters. Also, the fourth most sensitive parameter for the STAC model is the Tree θ_s, which is the second most sensitive for the L2p1 emulator. Overall, sensitivity of the model parameters have a different rank and magnitude for the L1p1 emulator compared to the original model, however the L2p1 emulator almost matches the original model. This is possibly an indication that the fidelity of the emulators to the original STAC model becomes more ideal when there is a higher level of sparse grid sampling (and thus

training points).

We then check the fidelity of the L1p1 and L2p1 10-parameter emulators of the CLM model for the GPP outputs to that of the original counterpart by again checking the 5 most sensitive parameters. The parameters included are several vegetation dynamics parameters, such as $V_{\mathrm{c,max25}}$ (or the parameter that controls the photosynthetic capactiy), Target Carbon Storage, Top of Canopy Specific Leaf Area (SLA), Maintenance Respiration, and Stem Allometry (or carbon allocation to stem). We see in Figure 5B that the most sensitive parameters for both the emulator and the original CLM are in the same order, but with totally different magnitudes. In fact, the L2p1 emulator sensitivities are shaped almost identically like the original model's, yet with somewhat of a bias. This bias could be attributed to the fact that the original model has 87 parameters and the emulator only considers 10 parameters, which allows each respective parameter to have more influence (and thus sensitivity) on the model outputs. However, we can assume there is high fidelity between the original model and the emulator since the rank of the parameter sensitivities is identical between the two.

5.6.4 Calibration of ESM's

We have now checked the accuracy, efficiency, and fidelity of the emulators, and assume that an emulator with reduced dimensions may be sufficient to calibrate an ESM, such as CLM4.5(ED). The accuracy of the 10-parameter emulator for CLM is very good, since it closely mimics the output response of the original model. The efficiency of the emulator is desirable, since the run time is a fraction of the original model's. The fidelity of the emulator is acceptable and quite impressive for the L2p1 emulator, since it captures most of the main parameter sensitivity rank and magnitudes. Overall, the application of PCE emulators for the calibration of ESMs seems possible and most probably will allow inference of model parameters with realistic values.

5.7 Conclusion

Environmental and Earth System Models (ESMs) are useful for analyzing the Earth and all its processes. Many of these models are CPU-intensive and contain many dozens of parameters that control the model output, making them rather difficult for calibration against observations. Here, we created emulators of various models using the PCE method in search of a strategy for efficient calibration of ESMs, a problem that the scientific community has struggled with for decades.

We first built emulators for the hmodel using sparse grid levels 1-3 and polynomial orders 1-3, and built emulators for the STAC model using sparse grid levels 1-2 and polynomial orders 1-3. The hmodel and STAC model were rigorously calibrated to measured data, and their calibration was compared to that of their surrogate models. These results provided confidence in the calibration of model emulators as a replacement for the original model, and we assume this to be true for a more complex and higher dimensional model such as the CLM. To this end, we built an L1p1 emulator for the 87-parameter CLM4.5(ED) for outputs of GPP, LAI, and biomass. The 87-parameter emulator for CLM did not perform well, therefore we reduced the dimensionality of the emulator to 10 parameters and built L1p1 and L2p1 emulators for this reduced version. This dimensionality reduction allowed the emulators to achieve a much stronger fit, requiring only 21 (L1p1) and 241 (L2p1) forward runs from the original model to train the emulators.

To investigate this further, we examined the accuracy of the simulations, the efficiency of each method, and the fidelity of each emulator to its original counterpart. In regards to accuracy and efficiency, all models were able to produce emulators that performed well compared to the original model and with much lower runtime costs. In regards to fidelity, the sensitivity of the STAC model parameters had a different rank and magnitude for its emulators than for the original model. However for CLM, the fidelity of the reduced 10-parameter emulator to the original CLM was very good, with parameter sensitivities of similar rank and magnitude.

Overall, we argue that results shown here were promising and give hope for efficient calibration

of ESMs. We look forward to the calibration and advancement of environmental models and ESMs, and promote the use of this method for efficient and accurate exploration of the parameter space for calibration problems of high CPU cost.

5.8 Acknowledgments

The authors gratefully acknowledge support from the SCGSR fellowship of the Department of Energy, the UC-Lab Fees Research Program Award 237285, and the DOE Office of Science Next Generation Ecosystem Experiment at Tropics (NGEE-T) project. The source code of the PCE method is written in MATLAB and available upon request from (elaloy@sckcen.be).

Chapter 6

Sustainable Groundwater Management in the Central Valley

by Elias C. Massoud, AJ Purdy, Michelle Miro
Jay Famiglietti, and Jasper A. Vrugt

Reference:

Massoud, E. C., Purdy, A.J., Miro, M., Famiglietti, J.S., Vrugt, J.A., Groundwater Sustainability in the Central Valley: An empirical approach to estimate groundwater depletion and recharge. Journal of Water Resources Planning and Management, Revision in Process. X: XX-X. doi: XX.XX/x. XXX. 2017.

6.1 Groundwater sustainability in California's Central Valley

California's Central Valley, the most productive agricultural area in the United States, has already experienced significant depletion in much of its aquifer systems. Although much of the Central Valley is dependent on groundwater, the state lacks data regarding how much water is being used and how much this amount changes year to year. Groundwater monitoring networks do not exist at the same scope and scale for those that track surface water (Scanlon et al., 2012a). In fact, no comprehensive framework for monitoring the world's groundwater resources currently exists (Famiglietti, 2014).

Lack of current effective management is already apparent in regions where wells have run dry, subsidence is impacting infrastructure (Sneed et al., 2013), groundwater water quality is degrading (Scanlon et al., 2007), and environmental impacts are becoming more visible. In this study, we develop an alternative method for estimating groundwater levels in the Central Valley. The model results are validated with historic groundwater level data provided by the USGS for the years 1981-2003 and from NASA's Gravity Recovery And Climate Experiment (GRACE) satellite for the years 2004-2014. The model is then used to predict future conditions of groundwater levels for the years 2015-2050 under various water management scenarios in order to understand how different practices can impact groundwater availability into the future.

To bypass the lack of publically available well monitoring data, various methods for estimating groundwater levels have been developed. Of these, the most prominent are geostatistical methods, computer model simulations (Harbaugh et al., 2000), and the use of remote sensing data for groundwater monitoring has also increased in recent years (Famiglietti et al., 2011; Rodell et al.,

2009). However, many of these methods provide estimates with high uncertainty, and there is a need for new ways to estimate groundwater behavior at the scales at which water is managed (Richey et al., 2015a). We believe that our model provides an alternative look at the groundwater picture for the Central Valley, and can provide awareness as well as knowledge for decision making in the upcoming decades.

6.2 An empirical method to estimate and project groundwater depletion and recharge

Abstract: Detailed knowledge of California's groundwater stockpile is of paramount importance for statewide planning and management of water resources, and to promote sustainable agriculture during periods of prolonged drought. Here, we use a thirteen year record (1998-2010) of water supply and demand information from California's Department of Water Resources (DWR) to quantify groundwater levels in the Central Valley. A water balance model with simple empirical relationships between annual precipitation, supply, and demand is used to explain historic groundwater levels (1981- 2014) derived from the United States Geological Survey (USGS) and NASA's Gravity Recovery and Climate Experiment (GRACE). After calibration of a recharge parameter, our empirical groundwater depletion (GWD) model mimics accurately the observed groundwater levels. The calibrated GWD model is then used to predict future groundwater levels for the years 2015-2050. During this period, we evaluate the impact of different management scenarios on future simulated groundwater availability. One of these scenarios confirms that lack of management action leads to continued depletion of California's groundwater resources. Other scenarios incorporate demand reduction and supply augmentation strategies and demonstrate sustainable groundwater levels for California's Central Valley.

6.3 Introduction and Scope

California has a complex and storied history of water management. A statewide plan of water storage, infrastructure, and conveyance has guaranteed for many decades a steady supply of surface water to satisfy demands of the more arid Central and Southern parts of the state. However, in recent years, urban, agricultural, and environmental demands in California have exceeded the natural renewable supply. To date, this gap between the available statewide surface water supply and the ever-increasing water demand is being met primarily by the extraction of groundwater. This pragmatic solution can have dire consequences as continued groundwater extraction depletes subsurface reservoirs, particularly in semi-arid regions with highly-variable precipitation amounts (Famiglietti, 2014; Famiglietti et al., 2011; McGuire, 2009; Richey et al., 2015b; Rodell et al., 2009; Wada et al., 2010). Indeed, California's Central Valley, the most productive agricultural area in the United States, has already witnessed significant groundwater depletion. Annually, at least 40 percent or more of the Central Valley's water supply comes from groundwater, which is primarily used to meet agricultural demand (Lo and Famiglietti, 2013). The groundwater extracted for irrigation more often than not exceeds the natural recharge, leading to rapid declines in the groundwater table (Faunt, 2009; Scanlon et al., 2012a; Siebert et al., 2010). This impact has been even more pronounced during prolonged dry periods when groundwater reliance increases.

The protection of California's groundwater resources is critical for sustaining the state's livelihood, ecology, and agricultural production, and key to preventing potentially harmful regional economic impacts that severe water shortages can cause. According to the United States Department of Agriculture's 2012 Census of Agriculture, the agricultural industry in California's Central Valley was valued at $42.6 billion, based on the market value of agricultural products sold. The region has also been termed the 'fruit and vegetable basket' of the United States supporting the diverse cultivation of 250 different crops (Scanlon et al., 2012a).

In addition to agriculture, the Central Valley has a growing population that increases demands

on the region's resources. By 2030, the population is projected to reach 10 million people. Data from the United States Geological Survey (USGS) shows that groundwater use has already increased from 0.6 million acre-feet per year (MAF/yr) in 1962 to nearly 2 MAF/yr in 2003 (equivalent to an increase from nearly 0.75 to 2 km3/yr) (Faunt, 2009). This ever increasing reliance on groundwater occurred despite the implementation of various urban conservation measures, and an increasing use of surface water (Faunt, 2009).

Although life and work in the Central Valley depend in large part on the availability of groundwater, detailed year-to-year data of how much water is being extracted and used is largely lacking, particularly when viewed in comparison to surface water resources. Groundwater monitoring networks do not exist at the same scope and scale for those that track surface water (Scanlon et al., 2012b). In fact, no comprehensive framework for monitoring the world's groundwater resources currently exists Famiglietti (2014). To bypass the lack of publically available well monitoring data, various different methods have been used and developed to estimate groundwater levels. This includes geostatistical interpolation methods, and computer model simulations (Harbaugh et al., 2000). What is more, in recent years, much effort has focused on groundwater monitoring using remote sensing data Alley et al. (2002); Famiglietti et al. (2011); McGuire (2009); Rodell et al. (2009); Scanlon et al. (2012a); Yeh et al. (2006). Yet, the groundwater levels predicted by these methods are subject to large uncertainties. Therefore, there is an urgent need for new monitoring techniques and/or computational methods to estimate groundwater resources at the scale of management (Richey et al., 2015b).

In this study, we develop an alternative method to estimate the groundwater levels in the Central Valley of California using a historic record of annual water use and supply from California's Department of Water Resources (DWR). Our Groundwater Depletion (GWD) model builds on the water balance and uses empirical relationships between annual precipitation, supply, and demand to simulate (predict) annual groundwater levels and recharge rates at the aggregated level of the

Central Valley. The simulated groundwater levels of the GWD model are evaluated against historic data of groundwater levels (1981-2014) measured by the USGS (1981-2003) and NASA's Gravity Recovery And Climate Experiment (GRACE) satellite (2004-2014). The GWD model is then used to predict groundwater levels for the years 2015-2050. During this period, we evaluate the impact of different management scenarios on future simulated groundwater availability.

6.4 Methods

In this section, we review briefly the experimental site under investigation, the Central Valley in California, and introduce the experimental data, empirical model and calibration approach used herein.

6.4.1 Study Site: Central Valley, CA

Our study region is the Central Valley, a 58,000 km^2 flat and elongated area located in the central heart of California, inland from and parallel to the Pacific Ocean coast (see Figure 6.1). This valley is 60-100 wide and stretches approximately 720 km from the city of Redding in the north-northwest to the city of Bakersfield in the south-southeast. The Central Valley watershed comprises 160,000 km^2 and is made up of three different main drainage systems, the relatively wet Sacramento Valley in the north, the drier San Joaquin Valley in the south, and the Tulare basin and semi-desert climate at the southernmost end. The climate is arid to semiarid with an average annual precipitation between 13-26 inches (33-66 cm) in the Sacramento Valley and between 5-18 inches (13-45 cm) in the San Joaquin Valley (Faunt, 2009). High inter-annual variability in precipitation has led the occurrence of dry and wet years (Savtchenko et al., 2015). Persistent droughts caused by recurrent dry years with insufficient rainfall and snowfall accumulation in the Sierra Nevada have enhanced considerably groundwater reliance in the Central Valley with (unsustainably) high rates of water depletion from its underlying reservoir. Evidence of this over-dependence on groundwater can be

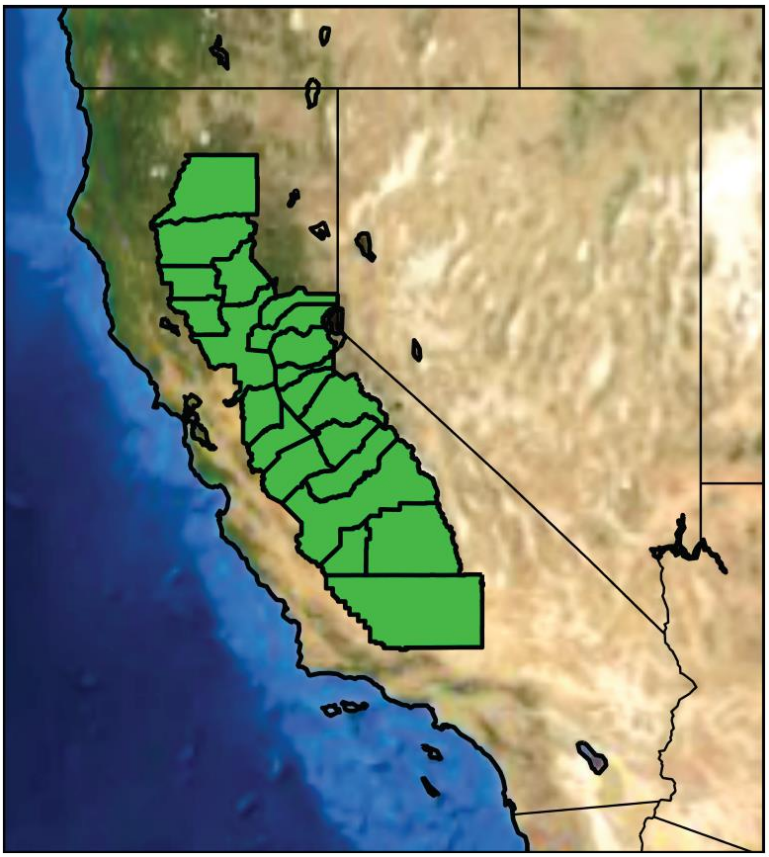

Figure 6.1: California's Central Valley (green).

found in land subsidence rates which equate to 25 cm/year in certain areas of the San Joaquin Valley (Faunt et al., 2016).

6.4.2 Experimental Data

Annual supply and demand data from DWR's California Water Plan was used to construct the calculation rules of our empirical Groundwater Depletion (GWD) model. The cited report includes data on the various fresh water supply and demand variables for regions across California. The most recent draft of this plan, the Water Plan Update 2013, provides annual constituents of water supply in the period from 1998-2010, including surface water (SS), groundwater (GW), and recycle and reuse (RR), as well as annual components of demand, including urban (URB), agricultural (AGR),

wild and scenic flows (WS), and environmental managed releases (EnvM) (California Water Plan Update, 2013). Sub-basin scale data for the Sacramento, San Joaquin and Tulare sub-basins were obtained upon request from DWR, and values for each variable were aggregated to represent the Central Valley as a whole. These values are shown in Table 1.

Precipitation data from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) were used for the years 1981-2014. For our future groundwater projections in the period from 2015-2050, annual precipitation estimates were derived from CAL-Adapt using the arithmetic mean of bias-corrected and downscaled annual rainfall predictions of four different climate models. This data was used in the Intergovernmental Panel on Climate Change Fourth Assessment Report under scenario A2 with medium-high carbon emissions (Maurer and Hidalgo, 2008; Pachauri et al., 2014), and includes output of the National Center for Atmospheric Research's Parallel Climate Model and Community Climate System Model, the NOAA Geophysical Fluid Dynamics Laboratory Model, and the Centre National de Recherches Mtorologiques Model. The precipitation data are made available through Cal-Adapt (http://cal-adapt.org), and detailed model descriptions can be found at http://ipcc-data.org. All data were aggregated to reflect anomalies of the Central Valley (Figure 6.1).

To evaluate the GWD model we use a 30-year record of 'measured' groundwater levels. This historic record includes data from the USGS (1981-2003) and from NASA's GRACE satellite mission (2004-2014) (Famiglietti et al., 2011; Faunt, 2009). Anthropogenic recharge, conceptualized as recharge from return flow from agriculture and percolation from surface water reservoirs, was obtained from the DWR California Water Plan.

Table 1: Department of Water Resources freshwater supply and demand data for the Central Valley, CA, for the years 1998-2010. Values were used to create relationships with precipitation (or PP), see Fig. 2. Shown are the supply variables, including surface supplies (SS), groundwater supplies (GW), and recycle and reuse supplies (RR), as well as demand variables, including urban (URB), agricultural (AGR), wild and scenic flows (WS), and environmental managed releases (EnvM). In blue are the. All values have units of Million Acre-Feet (MAF).

*SS includes all surface water supplies and storage, including instream environmental supplies and "Local deliveries".

			Demands (km^3)				Supplies (km^3)		
YEAR	PP	Wet/Dry	URB	AGR	WSC	ENVM	SS*	GW	RR
1998	1.71	Wet	1.45	17.55	13.01	1.17	30.64	9.57	0.0023
1999	0.92		1.18	23.49	11.64	0.00	29.01	16.12	0.16
2000	0.97		1.68	23.25	10.34	0.96	30.25	15.73	0.0023
2001	0.72	Dry	1.74	23.86	6.96	0.39	23.55	19.49	0.0023
2002	0.81	Dry	1.40	23.89	7.88	0.98	25.07	19.13	0.0073
2003	0.93		1.44	22.11	9.59	0.00	25.10	17.07	0.0
2004	0.94		1.45	24.89	9.93	1.38	27.72	19.95	0.07
2005	1.27	Wet	1.23	21.08	11.17	3.14	33.02	12.56	0.0
2006	1.27	Wet	1.35	22.83	18.02	2.55	39.40	14.44	0.03
2007	0.62	Dry	1.36	25.28	9.52	0.65	25.28	21.08	0.74
2008	0.77	Dry	1.47	26.02	7.10	0.87	22.09	23.19	0.50
2009	0.77	Dry	1.40	25.32	9.41	1.28	23.77	23.42	0.71
2010	1.04		1.35	22.13	10.61	1.46	27.67	16.39	0.40

Figure 6.2: Department of Water Resources freshwater supply and demand data for the Central Valley, CA, for the years 1998-2010. Values were used to create relationships with precipitation (or PP), see Fig. 2. Shown are the supply variables, including surface supplies (SS), groundwater supplies (GW), and recycle and reuse supplies (RR), as well as demand variables, including urban (URB), agricultural (AGR), wild and scenic flows (WS), and environmental managed releases (EnvM). In blue are the. All values have units of Million Acre-Feet (MAF). *SS includes all surface water supplies and storage, including instream environmental supplies and 'Local deliveries'.

6.5 The Empirical GWD Model

In this section we introduce the GWD model that is used to simulate (predict), and project groundwater levels in the Central Valley in the period 1981-2050. This model uses as main building block a regional water balance equation, augmented with simple empirical relationships between annual precipitation, supply, and demand. We next discuss these different relationships. To close the water balance, we include a simple loss function which scales the supplies to meet overall demands.

6.5.1 Supply-Demand Relationships with Precipitation

The data in Table 1 summarizes the total supply and demand for the years between 1998 and 2010, and is used to construct our GWD model. In essence, we develop empirical relationships between each supply and demand variable and the percent of precipitation, PP (Figure 6.3). These plots indicate how a given variable responds to different amounts of precipitation in a given year. For example, agricultural demand is greater during dry years and causes a higher reliance on groundwater supply. The opposite is true for wet years. We now use the empirical relationships depicted graphically in Figure 6.3, to setup a regional water balance for the Central Valley in California. Of course, groundwater is expected to play a major role in our model. The following equation portrays an adequate fit between the 1998-2010 trends of groundwater use and precipitation

$$GW_{\text{Pump}}(t) = -10 \cdot PP(t) + 24 \tag{6.1}$$

where GW_{Pump} (t) (km3) signifies the groundwater supply for year t, and $PP(t)$is the corresponding percent of average precipitation. For an average year, PP would equal 100% or 1.0. For dry years this value is less than 1.0, while for wet years this value is greater than 1.0.

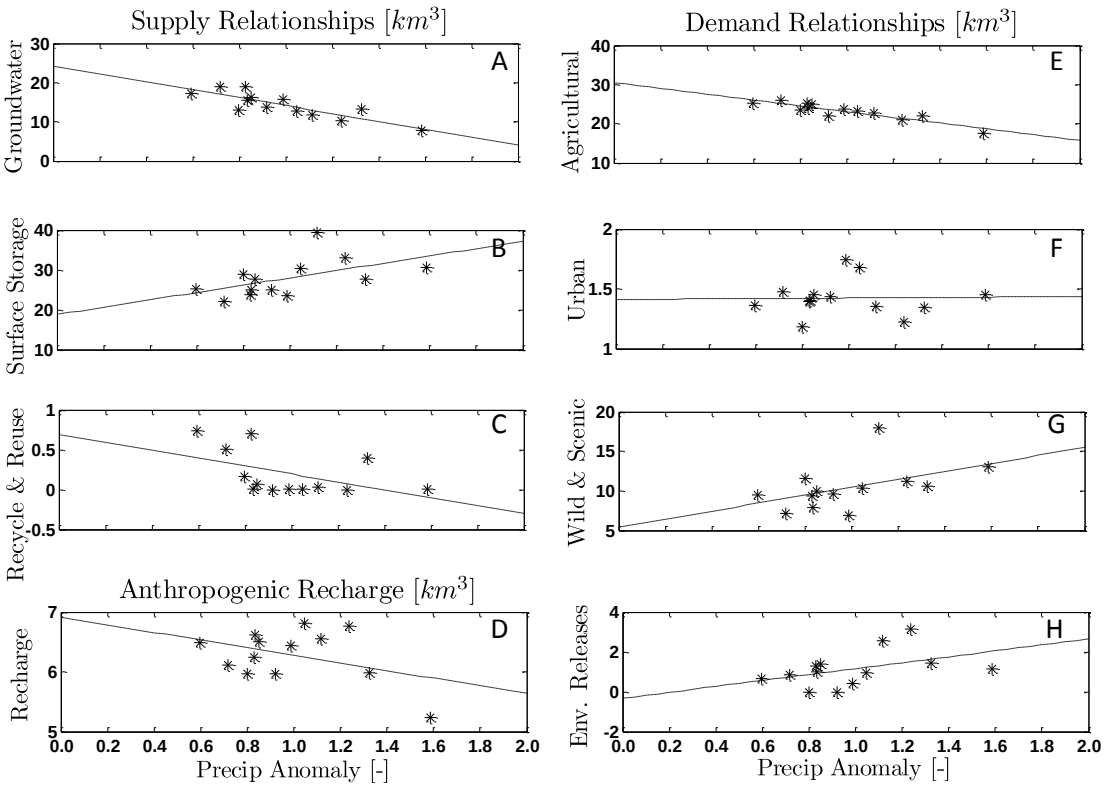

Figure 6.3: Empirical relationships that describe how each DWR supply and demand variable changes with precipitation. Also shown is the relationship between anthropogenic recharge and precipitation. DWR information for supply and demand and for anthropogenic recharge can be found in Table 1, along with the annual precipitation values (PP values in table).

6.5.2 Scaling Supplies to Meet Demands

To make sure that the supplies meet the estimated demands in each year, we include a 'loss' function that scales the supplies to meet overall demands. In other words, we assume that some amount of freshwater supply was unaccounted for in the estimates that the relationships in Figure 6.3 produce. Specifically, for each year, we calculate the difference between total demand and total supply and, if necessary, adjust the supply. Throughout the remainder of this paper, demand signifies actual use of water (e.g. agriculture or urban uses) and supply represents the type of water resource (e.g. groundwater or surface water) that was used to meet usage. Thus, a scaling factor, $X(t)$, is introduced in the model to scale supplies in order to meet the deficit, as follows

$$X(t) = \frac{TotalSupplies(t)}{TotalDemands(t)} \tag{6.2}$$

and this scaling factor is then applied to the estimated supplies

$$ActualSuplies(t) = \frac{TotalSupplies(t)}{X(t)} \tag{6.3}$$

For example, if the supply estimate in the t^{th} is 50 km^3 and the demand equates to 55 km^3, the scale factor $X(t)$ will scale up the supplies to satisfy the demand. See Figure 6.4 for further details on how the scale factor affects the estimation of groundwater supplies.

6.5.3 Recharge Formulation

Recharge from precipitation is difficult to measure in practice, particularly at large spatial scales, and was therefore not included in the DWR dataset. As a consequence, we have to parameterize this component of the regional water balance. Groundwater recharge, the percolation of water from the surface to subsurface aquifers, can come from a variety of sources, including precipitation, leakage from streams and surface-water bodies, and return flow from irrigated agriculture. For

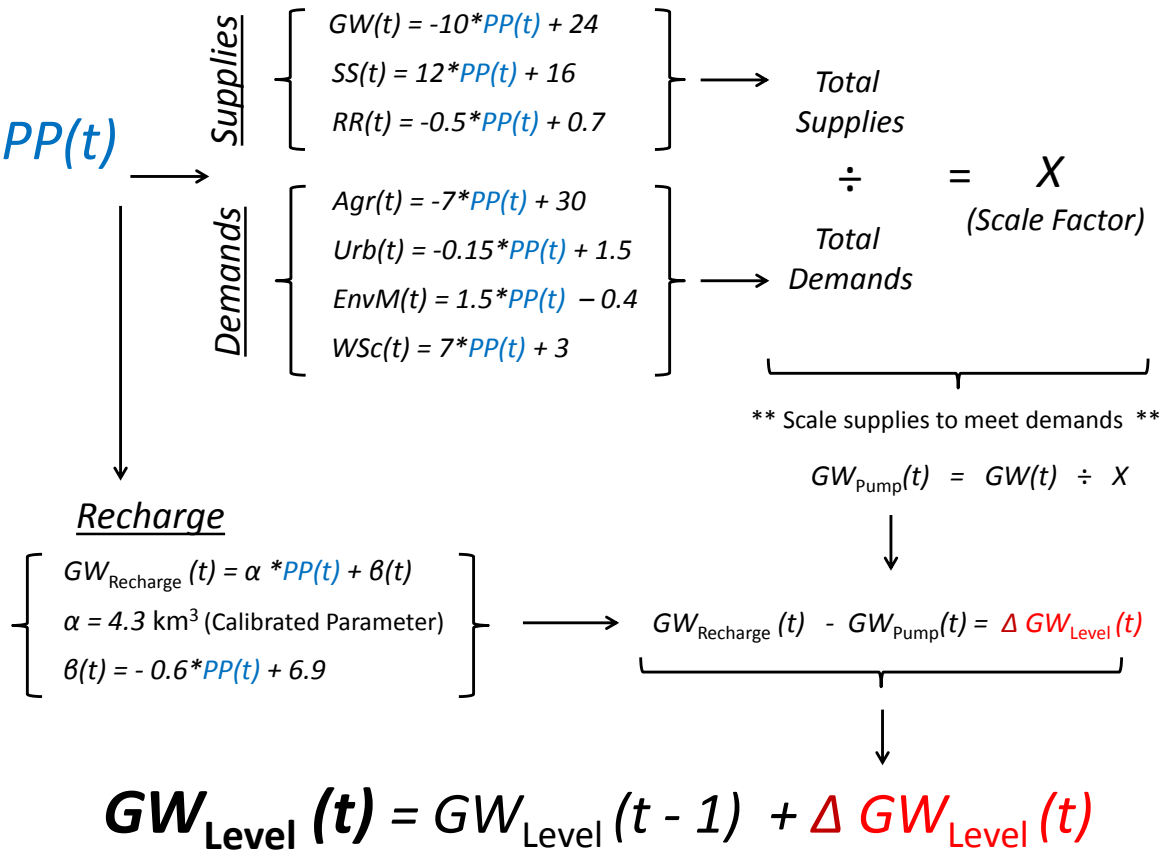

Figure 6.4: Schematic showing flow of model structure.

this reason, it is not particularly easy to accurately model recharge in a region. Many different studies have appeared in the literature that have developed methods for estimating groundwater recharge(Crosbie et al., 2015; Doble et al., 2009; Scanlon et al., 2012b). Yet, all their recharge estimates are subject to considerable uncertainty. Here, we use a much simpler approach, and relate recharge directly to precipitation

$$GW_{\text{Recharge}}(t) = \alpha \cdot PP(t) + \beta(t) \tag{6.4}$$

where $GW_{\text{Recharge}}(t)$ signifies the t^{th} year recharge (in km^3), α denotes the unknown recharge parameter (km^3) associated with precipitation, and β is another recharge term (km^3) associated with anthropogenic effects.

As will be shown in the results section, the value of α is equivalent to about 4.3 km^3.The yearly values of β can be estimated from the DWR data set using the following linear relationship with percent of average precipitation (see Figure 6.3 D)

$$\beta(t) = -0.6 \cdot PP(t) + 6.9 \tag{6.5}$$

The final recharge equation of the model now becomes

$$GW_{\text{Recharge}}(t) = 4.3 \cdot PP(t) + [-0.6 \cdot PP(t) + 0.69] = 3.7 \cdot PP(t) + 6.9 \tag{6.6}$$

Figure 6.4 provides further explanation on how the recharge estimate is calculated in each year.

6.5.4 Simulating Groundwater Storage

To estimate groundwater storage in a given year, the model considers a virtual volume of groundwater, then subtracts the supplied groundwater in that year and adds the volume of recharge calculated for that year. The temporal evolution of the storage in the reservoir is calculated using the following regional water balance equation for the groundwater compartment (all terms in km^3)

$$GW_{\text{Level}}(t) = GW_{\text{Level}}(t-1) + GW_{\text{Recharge}}(t) - GW_{\text{Pump}}(t) \tag{6.7}$$

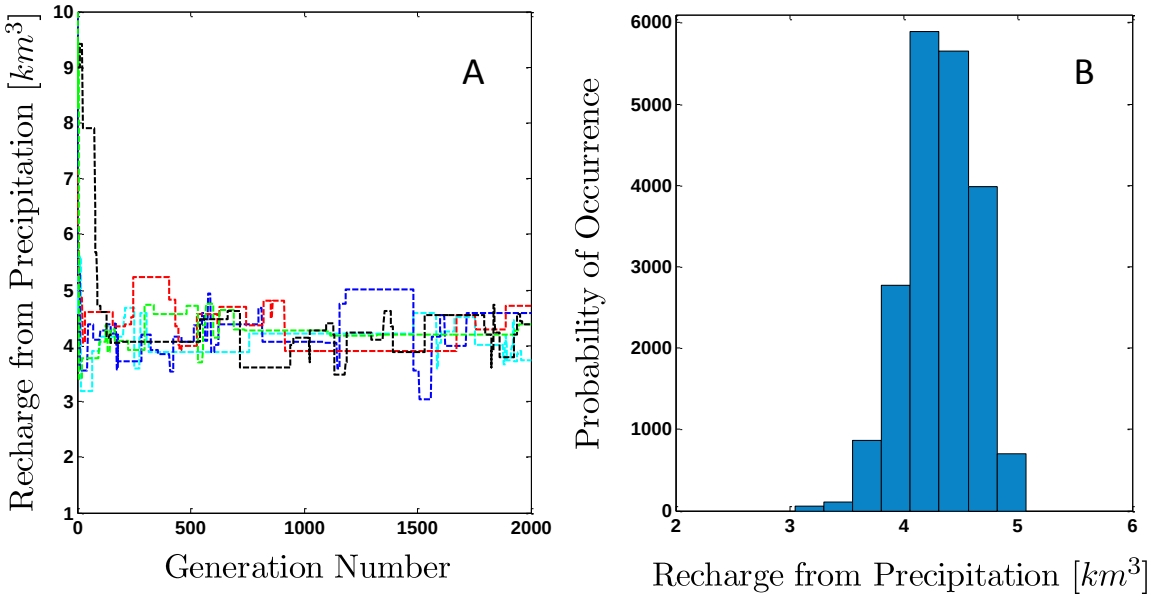

Figure 6.5: Calibration of the parameter, α , which represents the change in recharge per unit change in precipitation. Fig. 4A. shows the trajectory of each of the chains, and provides information on how the parameter value was estimated. In Fig. 4B., the final distribution of possible parameter values is shown, which indicates the level of uncertainty for the parameter estimate.

where $GW_{\text{Level}}(t)$ and $GW_{\text{Level}}(t-1)$ denote the groundwater level of the Central Valley reservoir for some given year and the immediately predating year, respectively, and $GW_{\text{Recharge}}(t)$ and $GW_{\text{Pump}}(t)$ signify the present year's recharge (estimated using Equation 6.6) and groundwater supply (estimated using Equation 6.1). The initial storage of the reservoir, the groundwater level in 1981, was derived from the USGS data set and set equivalent to -36 km^3. It is evident from Equation 6.7 that the GWD model uses an annual integration time step. The GWD model is illustrated schematically, with its underlying equations, in Figure 6.4.

6.5.5 Calibration of Recharge from Precipitation: An MCMC Approach

The parameter that represents recharge from precipitation α, is estimated using an enhanced Markov Chain Monte Carlo (MCMC) method - the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2009, 2008). This algorithm exhaustively searches the feasible space of α in pursuit of a stationary distribution. This distribution contains the optimum value of α and characterizes the underlying statistical uncertainty of this parameter. A detailed description of the DREAM algorithm appears in the cited references (Vrugt, 2016) and interested readers are referred to these publications for further details.

We use the MATLAB toolbox of DREAM described in Vrugt (2016) and assume a uniform prior distribution for the recharge parameter, α, and a classical Gaussian likelihood function to compare the observed and simulated groundwater table levels. This likelihood function,

$$L(\alpha|\widetilde{\mathbf{Y}}) \propto \sum_{t=1}^{n} |e_t(\alpha)|^{-n} \tag{6.8}$$

where $n = 34$ constitutes the number of historic groundwater level observations, $\widetilde{\mathbf{Y}} = [\widetilde{y_1}, \cdots, \widetilde{y_n}]$, in the period from 1981-2014, and the variable, $e_t(\alpha) = \widetilde{y_t} - y_t(\alpha)$, signifies the annual residuals of the observed and GWD simulated groundwater levels. A detailed derivation of the Gaussian likelihood function of Equation 6.8 appears in Vrugt and Massoud (2016) and interested readers are referred to this publication or Vrugt (2016). In essence, the recharge parameter will be tuned so that the sum of squared residuals of the simulated and observed groundwater levels is minimized.

Note, that the residuals between the observed and simulated groundwater levels express some temporal correlation. This is in disagreement with our assumption of uncorrelated residuals in our Gaussian likelihood function. Without becoming too technical for the purpose of the present paper and audience, we have tried alternative likelihood functions that account explicitly for

first-order residual correlation. The DREAM results confirm presence of serial dependence among the residuals, yet this temporal correlation of about 0.4 is rather weak and hardly affects the optimal value of the recharge parameter, α.

Figure 6.5 A provides a trace plot of the simulated chain trajectories of the DREAM algorithm. During the initial generations, the different Markov chains search different parts of the parameter space. Then, the chains collapse to a small region - the so-called stationary distribution of α - and rapidly reach convergence according to different built-in diagnostics. Figure 6.5 B, displays a histogram of the samples produced by the different Markov chains. The peak of this distribution coincides with the 'best' estimate of the recharge parameter, whereas the dispersion of the bars provides an estimate of the uncertainty of the calibrated parameter value. The value of α ranges between 3.8 - 4.8 km^3, and we select herein $\alpha = 4.3$ km^3 as a reasonable optimal estimate. It is important to note that α has a distribution that is very well defined (small uncertainty), which inspires confidence in the structure of the model. This is particularly beneficial for model simulation as we can ignore the effect of parameter uncertainty and resort to deterministic projections of future aquifer storage.

6.6 GWD Model Simulations

We now discuss the results of the GWD model in two different time frames, the past (for model evaluation) and the future (for assessing sustainability of groundwater under various management scenarios).

6.6.1 Model Evaluation

A major concern of California's water resource management is the storage of the groundwater aquifer of the Central Valley. This study aims to quantify the reliance on groundwater and highlights the increase of this reliance during dry years. To this end, Figure 6.6 A shows precipitation

anomalies from each year used to drive the model. PRISM data is used for the years 1981-2014 (dark colors), and CAL-ADAPT data is used for the years 2015-2050 (light colors). The dry years are shown in red and wet years are shown in blue. Figure 6.6 B shows the change in simulated GW levels, calculated as the difference between supply from and recharge to groundwater, i.e. $GW_{\mathrm{Pump}}(t)$ and $GW_{\mathrm{Recharge}}(t)$, respectively, from Equation 6.7. Also evident in Figure 6.6 B are large decreases in GW levels during dry years, which are presumably the result of increased groundwater reliance and decreased natural recharge. Figure 6.6 C shows the observed changes in groundwater levels and allows a comparison with model results. The simulated changes in groundwater (Figure 6.6 B) match the observed changes (Figure 6.6 C) quite well and are particularly representative of depletion during dry years (e.g. drought in 1987-1992) as well as of the high amounts of recharge in wet years (e.g. wet years in 1982-1983).

The simulated groundwater levels are shown in Figure 6.7 (black line) and are compared with observations (red and blue dots) for the years 1980-2014. The model fits the data relatively well, even though the model is structured to aggregate processes that occur at annual time-steps and over the spatial scale of the Central Valley. Errors in the model may occur if time-delayed processes such as recharge from base flow or extensive anthropogenic recharge from irrigation occur at longer timescales. One reason for mismatches between the GWD modeled data and the GRACE-based groundwater observations could be from GRACE uncertainty (Rodell et al., 2009; Yeh et al., 2006). GRACE-based groundwater changes rely on models to account for the non-groundwater components (i.e. soil moisture and surface water) of GRACE total water storage estimates. These models often fail to account for anthropogenic impacts on water storage changes and have their own degree of uncertainty, leading to error propagation in the derived groundwater change values from GRACE (Famiglietti et al., 2011).

Overall, we contend that the model is able to accurately assess impacts on groundwater levels given different water management scenarios in California to a level that is necessary for examining

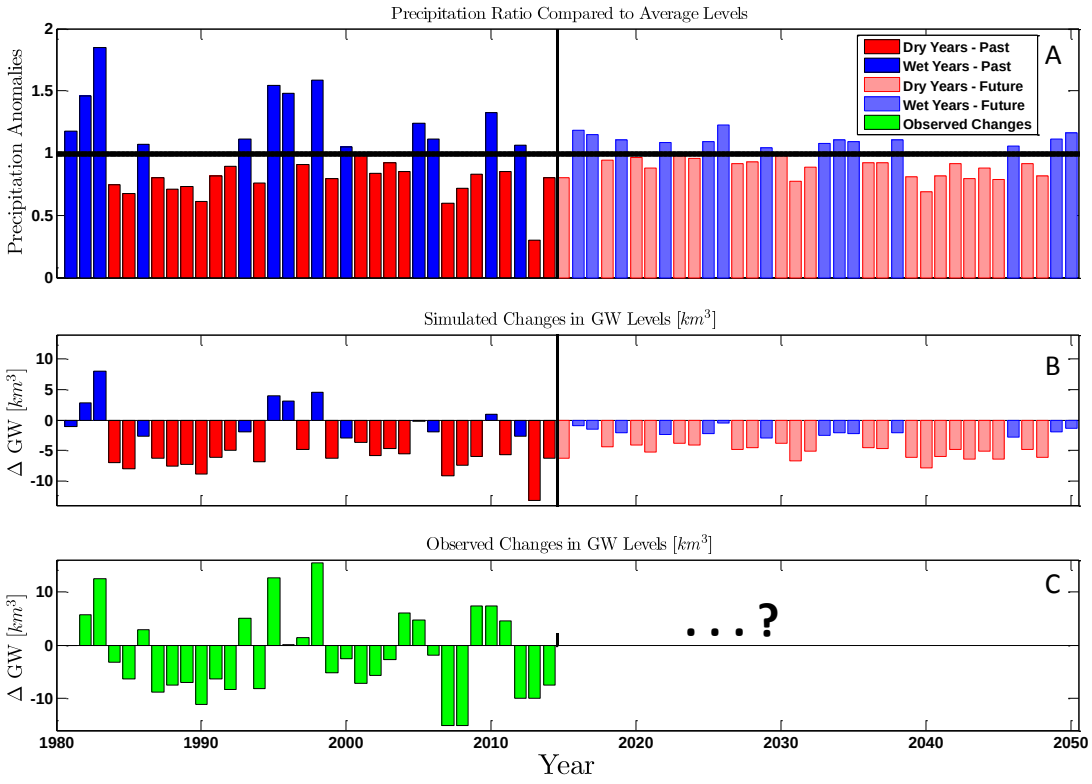

Figure 6.6: Precipitation anomalies from each year are shown in Panel A, where PRISM data was used for the years 1981-2014 (dark colors) and CAL-ADAPT data was used for the years 2015-2050 (light colors); dry years are shown in red and wet years are shown in blue. In Panel B., the change in simulated GW levels is shown, which is calculated as the difference between the volume of water extracted from groundwater and the recharge that replenishes it Panel C. highlights the observed changes in groundwater levels in order to facilitate a direct comparison with the model's estimated values in Panel B.

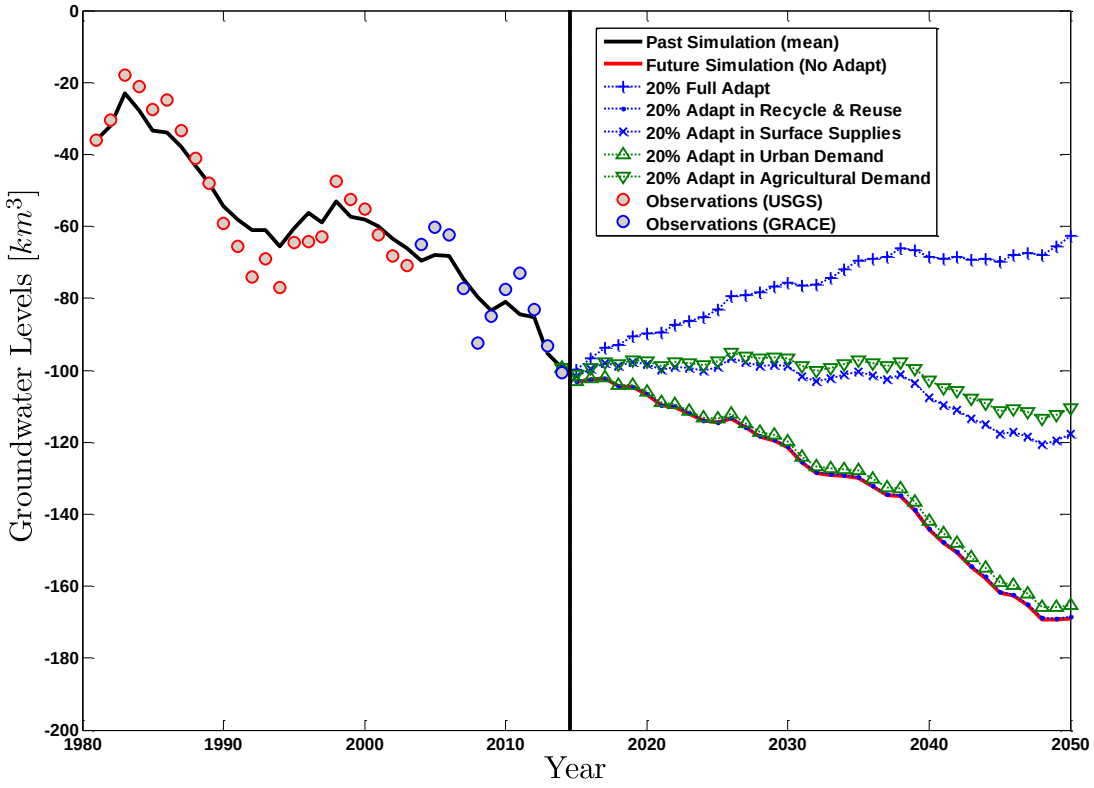

Figure 6.7: The GWD Model simulations for both the past and future years. The simulated groundwater levels for the years 1980-2014 (black line) are compared with observations (red icons). In future years, several adaptation scenarios are examined. The case of no adaptation is shown with a red line and the case with a full adaptation (i.e. 20% augmentation in RR and SS supplies and 20% reduction in URB and AGR demands) is shown with blue crosses ('+'). Then, each supply/demand variable's sensitivity to groundwater depletion is examined by assessing various cases with 20% changes in each variable individually. In these cases, changes in demands are shown in green and changes in supplies are shown in blue.

large, regional trends. Thus, we run the GWD model for future years under various adaptation scenarios to provide general insights into the rates of depletion of the aquifer based on different water management scenarios.

6.6.2 Future Projections: Business as Usual vs Adaptation

Several adaptation scenarios are examined for future years (2015-2050) using precipitation from CAL-Adapt to drive the model. In Figure 6.7, the case of no adaptation (i.e. business as usual)

is shown with a red line, and the case with significant management intervention is shown with blue crosses ('+'). The management strategy we modeled for this study is a simultaneous demand reduction and supply augmentation approach that reduces agricultural and urban demands by 20% and at the same time increases surface water supplies and supplies from recycle and reuse by 20%. To get an idea of how sensitive groundwater levels are to each variable individually, we also run the model with individual reductions by 20% in each demand or 20% increase in each supply variable. In these cases, demand reduction strategies are shown in green and supply augmentation approaches are shown in blue. Model results show that efficiencies in surface water supply and agriculture efficiency can have a stronger impact on groundwater levels compared to efficiencies in urban water use and recycle and reuse. However, from these results it is clear that a comprehensive approach that looks at both supply and demand side management strategies may be necessary to sustain groundwater levels in the future.

Agricultural efficiency encompasses more than on-farm water use efficiency, and 'improvements' can also reflect changes in total farmed acreage. As a result, management of groundwater at the basin scale will need to focus on structural water demands (farmed acreage) in addition to on-farm efficiency improvements over multiple years.

From our simulations it is clear that if no changes are made in water management (red line), groundwater depletion will continue due to projected decreases in precipitation and an unbalanced water management portfolio. Under the adaptive management strategy, groundwater storage can be sustained, showing potential that the aquifer can indeed be managed sustainably and secured for future generations.

6.7 Discussion

6.7.1 What Does the Future Hold?

Groundwater extraction in the Central Valley is and will continue to be driven by the region's fresh water supply-demand gap. This study addresses various water management portfolios that will impact the magnitude of this gap and groundwater depletion. The results of the GWD model indicate that a no change scenario will cause the aquifer to deplete further in the future. We considered several adaptation scenarios, including the reduction of agricultural or urban demand by 20%, the increase in surface supplies or recycle and reuse supplies by 20%, and the final case that considered a 20% change in all four of these variables simultaneously.

Because urban demand and supplies from recycled and reused water are such a small fraction of the current uses and supplies, the results indicate that 20% improvement in these variables only incrementally change the behavior of groundwater depletion in the Central Valley. However, according to our model, implementing strategies that work towards both a 20% reduction in agricultural demand and a 20% augmentation in the use of surface water supplies can significantly mitigate groundwater depletion.

The current rate of depletion raises alarming questions, including how long the aquifer will remain a useful source of fresh water for the region. This is a critical question for water managers, as groundwater constitutes a significant portion of the supply portfolio. If aquifer depletion continues at unsustainable rates, the gap between water demands and available supplies will eventually not be satisfied by groundwater extraction and therefore create larger, and potentially severe, social and economic problems for the state.

6.7.2 Future Work

The GWD model is useful to assess the water balance in the California Central Valley, yet is rather simplistic and empirical and thus has much room for improvement. The model aggregates spatially processes occurring over the entire Central Valley and temporally aggregates activity over an entire year. As a result it is susceptible to large uncertainties and errors. Building from the current structure to model at the sub-basin, the scale at which groundwater is now managed in California, and seasonal timescale can further the impact of this tool to help identify management scenarios that incorporate processes that occur at spatiotemporal scales not captured by the current version of the GWD model.

One result of the management strategies scenario testing shows that groundwater and surface water use are strongly linked. Therefore, incorporating data on water availability in surface water reservoirs into the model structure while simultaneously considering groundwater changes can provide a more detailed picture on the dynamics of water resource management in the Central Valley. This will also assist with multi-annual impacts not currently represented (i.e. consecutive dry years).

An additional portion of the model that could be improved is that of urban demand. Population growth and urban demand change over time and may not necessarily reflect changes in climate or precipitation, despite policies. However, as stronger urban policies take hold we expect a stronger relationship to emerge. The current structure of the model does not consider population growth, something that would need to be considered if the model advances towards smaller spatial scales where urban demand may have a stronger impact on a particular basin.

6.8 Conclusion

Due to the high degree of uncertainty associated with inter-annual precipitation and surface water availability in semi-arid regions, groundwater is often relied upon to meet water needs. Lack of

current effective management is already apparent in regions where wells have run dry, subsidence is impacting infrastructure (Sneed et al., 2013), groundwater quality is degrading (Scanlon et al., 2007), salinization of soil and groundwater resources is taking place (Schoups et al., 2005), and environmental impacts are becoming more visible. Allowing unsustainable use of groundwater resources will only exacerbate these problems. Costs associated with water extraction would also increase due to the need to drill deeper wells, treat degraded groundwater quality and pump from deeper depths (Konikow and Kendy, 2005). Moreover, 21 groundwater basins within California are classified as critically over drafted, and recent groundwater management regulations require implementation of groundwater sustainability plans over the next few decades.

Our simple GWD model provides an easily employable and interpretable tool to help analyze the interconnected components that complicate system management. In this study, we present the GWD model, a new empirical method to estimate groundwater depletion. We simulated groundwater levels from 1981-2050 using historical and projected precipitation data. We explore different future scenarios using projected precipitation and changes in management practices. Our results show that a 20% augmentation of supplies other than groundwater coupled with 20% reductions in demands for agriculture and urban water use could sustain groundwater storage at its present level. An interactive web application was developed as a template for users to explore and assess the sensitivity of groundwater sustainability in this region under a variety of management strategies using our simple model. However, groundwater is a finite resource when rates of use surpass natural replenishment, and the amount of groundwater in many areas is unknown and/or highly uncertain (Richey et al., 2015b). Effective groundwater management requires accurate knowledge of how strongly surface water and groundwater are linked and how much groundwater is changing every year. Water management strategies must consider both surface and groundwater supply within a given area in order to prevent unsustainable dependence on these resources (McNutt, 2014).

The relatively constant gap between renewable surface water supplies and statewide demands

indicates that supply mitigation and demand reduction strategies implemented today could have lasting impacts into the future. In particular, the development of recycled and reused water systems and the improvement of agricultural and urban water efficiencies could help close the gap even more and reduce reliance on our non-renewable resources in times of drought. In addition, research and innovation on alternative management strategies and new ideas to close the gap could also help to ensure a steady supply as well as minimize future exploitation of groundwater resources.

6.9 Acknowledgements

The authors would like to acknowledge the California Department of Water Resources (DWR) for the data used to derive this model. Also, the UC Center for Hydrologic Modeling for Groundwater Depletion data (USGS and GRACE). The first and last author are supported by funding from the UC-Lab Fees Research Program Award 237285.

Chapter 7

Summary and Conclusions

In 2013, the World Meteorological Organization (WMO) urged the global community for coordinated international action against accelerating and potentially devastating climate change. Preliminary data indicated that carbon levels increased more between 2012 and 2013 than during any other year since 1984, and this was possibly related to reduced uptake by the Earth's biosphere in addition to the steadily increasing emissions from the Earth's surface. In the upcoming decades, it will be critical for scientists and policy makers to not only resolve the problem of carbon emissions by assessing human behavior, but also to understand as thoroughly as possible the underlying coupled processes of the Earth's atmosphere and biosphere in order to adequately measure and estimate the fluxes of carbon, water, and energy that are dictating the climatic trends we observe today. Fortunately, our ability to understand Earth's processes and predict climate change is improving. This is becoming possible due to increased volumes of data, increases in computational power, advances in climate models and related statistical and optimization methodologies, and most importantly, societal needs to develop better strategies for sustainable natural resource management.

This thesis covered a suite of environmental models and numerical methods to disentangled information found both in observed data as well as model simulations. Various methods were applied such as parameter estimation with Markov Chain Monte Carlo (MCMC), state estimation

with data assimilation using the Ensemble Kalman Filter (EnKF), and sensitivity analysis of model parameters using the Fourier Amplitude Sensitivity Test (FAST), which all in one way or another offer treatments to predictive uncertainty. Furthermore, applying these methods on more sophisticated and complex models was impossible sometimes due to their high CPU costs; in this thesis model emulators were built using Polynomial Chaos Expansion (PCE) to reduce the computational burden for various environmental models. Overall, our goal in this dissertation was to present what tools are currently available for making predictions of environmental systems, with emphasis on maintaining accuracy of model simulations when compared to observed data, optimizing the efficiency of computationally heavy models to minimize their run time costs, and obtaining fidelity of model structures to properly represent the underlying hydrologic, biophysical, and biogeochemical processes occurring on our Earth.

Bibliography

Aanensen, D. M., Huntley, D. M., Feil, E. J., Spratt, B. G., et al. (2009). Epicollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS One*, 4(9):e6968.

Ali, A., Xu, C., Rogers, A., Fisher, R., Wullschleger, S., Massoud, E., Vrugt, J., Muss, J., McDowell, N., Fisher, J., et al. (2016). A global scale mechanistic model of photosynthetic capacity (luna v1. 0). *Geoscientific Model Development*, 9(2):587–606.

Ali, A., Xu, C., Rogers, A., Fisher, R., Wullschleger, S., McDowell, N., et al. (2015). A global scale mechanistic model of the photosynthetic capacity. *Geoscientific Model Development Discussions*, 8:6217–6266.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *FAO, Rome*, 300(9):D05109.

Alley, W. M., Healy, R. W., LaBaugh, J. W., and Reilly, T. E. (2002). Flow and storage in groundwater systems. *science*, 296(5575):1985–1990.

Archer, G., Saltelli, A., and Sobol, I. (1997). Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120.

Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G.,

Bopp, L., Brovkin, V., Cadule, P., et al. (2013). Carbon–concentration and carbon–climate feedbacks in cmip5 earth system models. *Journal of Climate*, 26(15):5289–5314.

Aumann, C. A. and Ford, E. D. (2002). Modeling tree water flow as an unsaturated flow through a porous medium. *Journal of theoretical biology*, 219(4):415–429.

Bastidas, L., Gupta, H. V., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. (1999). Sensitivity analysis of a land surface scheme using multicriteria methods. *Journal of Geophysical Research: Atmospheres*, 104(D16):19481–19490.

Bayes, M. and Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs. *Philosophical Transactions (1683-1775)*, pages 370–418.

Becks, L., Hilker, F. M., Malchow, H., Jürgens, K., and Arndt, H. (2005). Experimental demonstration of chaos in a microbial food web. *Nature*, 435(7046):1226–1229.

Benincà, E., Ballantine, B., Ellner, S. P., and Huisman, J. (2015). Species fluctuations sustained by a cyclic succession at the edge of chaos. *Proceedings of the National Academy of Sciences*, 112(20):6389–6394.

Benincà, E., Dakos, V., Van Nes, E. H., Huisman, J., and Scheffer, M. (2011). Resonance of plankton communities with temperature fluctuations. *The American Naturalist*, 178(4):E85–E95.

Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., et al. (2008). Chaos in a long-term experiment with a plankton community. *Nature*, 451(7180):822–825.

Benincà, E., Jöhnk, K. D., Heerkloss, R., and Huisman, J. (2009). Coupled predator–prey oscillations in a chaotic food web. *Ecology Letters*, 12(12):1367–1378.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berger, J. O., Bernardo, J. M., and Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, pages 905–938.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122.

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 113–147.

Bertino, L., Evensen, G., and Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2):223–241.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, 320(1):18–36.

Bittner, S., Legner, N., Beese, F., and Priesack, E. (2012). Individual tree branch-level simulation of light attenuation and water flow of threef. sylvatica l. trees. *Journal of Geophysical Research: Biogeosciences*, 117(G1).

Black, M. and Pritchard, H. W. (2002). *Desiccation and survival in plants: drying without dying*. Cabi.

Blanning, R. W. (1975). Response to michel, kleijnen and permut. *Interfaces*, 5(3):24–25.

Bliznyuk, N., Ruppert, D., Shoemaker, C., Regis, R., Wild, S., and Mugunthan, P. (2012). Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. *Journal of Computational and Graphical Statistics*, 17(2):270–294.

Bohrer, G., Mourad, H., Laursen, T. A., Drewry, D., Avissar, R., Poggi, D., Oren, R., and Katul, G. G. (2005). Finite element tree crown hydrodynamics model (fetch) using porous media flow within branching elements: A new representation of tree hydrodynamics. *Water Resources Research*, 41(11):1–17.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C. (2011). Improving canopy processes in the community land model version 4 (clm4) using global flux fields empirically inferred from fluxnet data. *Journal of Geophysical Research: Biogeosciences*, 116(G2).

Borgonovo, E., Castaings, W., and Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. *Environmental Modelling & Software*, 34:105–115.

Boyer, J. S. (1982). Plant productivity and environment. *Science*, 218(4571):443–448.

Brubaker, K. L., Entekhabi, D., and Eagleson, P. (1993). Estimation of continental precipitation recycling. *Journal of Climate*, 6(6):1077–1089.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2015). Shiny: web application framework for r. *R package version 0.11*, 1.

Chen, M., Liu, S., Tieszen, L., and Hollinger, D. (2008). An improved state-parameter analysis of ecosystem models using data assimilation. *Ecological Modelling*, 219(3):317–326.

Christoffersen, B. O., Gloor, M., Fauset, S., Fyllas, N. M., Galbraith, D. R., Baker, T. R., Kruijt, B., Rowland, L., Fisher, R. A., Binks, O. J., et al. (2016). Linking hydraulic traits to tropical forest function in a size-structured and trait-driven model (tfs v. 1-hydro). *Geoscientific Model Development*, 9(11):4227.

Chuang, Y.-L., Oren, R., Bertozzi, A. L., Phillips, N., and Katul, G. G. (2006). The porous media model for the hydraulic system of a conifer tree: Linking sap flux data to transpiration rate. *Ecological Modelling*, 191(3):447–468.

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., et al. (2001). Ecological forecasts: an emerging imperative. *Science*, 293(5530):657–660.

Clark, M. P., Kavetski, D., and Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47(9).

Claussen, M., Mysak, L., Weaver, A., Crucifix, M., Fichefet, T., Loutre, M.-F., Weber, S., Alcamo, J., Alexeev, V., Berger, A., et al. (2002). Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate Dynamics*, 18(7):579–586.

Conroy, M. J., Cohen, Y., James, F. C., Matsinos, Y. G., and Maurer, B. A. (1995). Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. *Ecological Applications*, 5(1):17–19.

Conti, S. and O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of statistical planning and inference*, 140(3):640–651.

Conway, G. R., Glass, N. R., and Wilcox, J. C. (1970). Fitting nonlinear models to biological data by marquardt's algorithm. *Ecology*, 51(3):503–507.

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

Crestaux, T., Le Maı?tre, O., and Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172.

Crosbie, R. S., Davies, P., Harrington, N., and Lamontagne, S. (2015). Ground truthing groundwater-recharge estimates derived from remotely sensed evapotranspiration: a case in south australia. *Hydrogeology Journal*, 23(2):335–350.

Crosetto, M. and Tarantola, S. (2001). Uncertainty and sensitivity analysis: tools for gis-based model implementation. *International Journal of Geographical Information Science*, 15(5):415–437.

Crossley, J., Polcher, J., Cox, P., Gedney, N., and Planton, S. (2000). Uncertainties linked to land-surface processes in climate change simulations. *Climate Dynamics*, 16(12):949–961.

Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A., and Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of chemical physics*, 59(8):3873–3878.

Cukier, R., Levine, H., and Shuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics*, 26(1):1–42.

Da Rocha, H. R., Nobre, C. A., Bonatti, J. P., Wright, I. R., and Sellers, P. J. (1996). A vegetation-atmosphere interaction study for amazonia deforestation using field data and a ?single column?model. *Quarterly Journal of the Royal Meteorological Society*, 122(531):567–594.

David, T. S., Pinto, C. A., Nadezhdina, N., Kurz-Besson, C., Henriques, M. O., Quilhó, T., Cermak, J., Chaves, M. M., Pereira, J. S., and David, J. S. (2013). Root functioning, tree water use and hydraulic redistribution in quercus suber trees: A modeling approach based on root sap flow. *Forest Ecology and Management*, 307:136–146.

De'ath, G. and Fabricius, K. E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11):3178–3192.

Dekker, S. C., Vrugt, J. A., and Elkington, R. J. (2012). Significant variation in vegetation characteristics and dynamics from ecohydrological optimality of net carbon profit. *Ecohydrology*, 5(1):1–18.

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C., and Martin, P. R. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*, 105(18):6668–6672.

Di Pierro, F., Khu, S.-T., Savić, D., and Berardi, L. (2009). Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms. *Environmental Modelling & Software*, 24(2):202–213.

Dietze, M. C., Sala, A., Carbone, M. S., Czimczik, C. I., Mantooth, J. A., Richardson, A. D., and Vargas, R. (2014a). Nonstructural carbon in woody plants. *Annual review of plant biology*, 65:667–687.

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., Kooper, R., LeBauer, D., Mantooth, J., McHenry, K., et al. (2014b). A quantitative assessment of a terrestrial biosphere model's data needs across north american biomes. *Journal of Geophysical Research: Biogeosciences*, 119(3):286–300.

Dirac, P. (1953). The lorentz transformation and absolute time. *Physica*, 19(1-12):888–896.

Doak, D. F., Estes, J. A., Halpern, B. S., Jacob, U., Lindberg, D. R., Lovvorn, J., Monson, D. H., Tinker, M. T., Williams, T. M., Wootton, J. T., et al. (2008). Understanding and predicting ecological dynamics: are major surprises inevitable. *Ecology*, 89(4):952–961.

Doble, R. C., Simmons, C. T., and Walker, G. R. (2009). Using modflow 2000 to model et and recharge for shallow ground water problems. *Groundwater*, 47(1):129–135.

Dodd, I. C., Egea, G., and Davies, W. J. (2008). Accounting for sap flow from different parts of the root system improves the prediction of xylem aba concentration in plants grown with heterogeneous soil moisture. *Journal of Experimental botany*, 59(15):4083–4093.

Doorenbos, J. (1977). Guidelines for predicting crop water requirements. *FAO irrigation and drainage paper*, 24:15–29.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J., Brazile, J., Singh, R., and Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for

enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2):165–193.

Draye, X., Kim, Y., Lobet, G., and Javaux, M. (2010). Model-assisted integration of physiological and environmental constraints affecting the dynamic and spatial patterns of root water uptake from soils. *Journal of Experimental Botany*, 61(8):2145–2155.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y., Habets, F., Hall, A., Hay, L., et al. (2006). Model parameter estimation experiment (mopex): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, 320(1):3–17.

Duan, Q., Sorooshian, S., and Gupta, V. K. (1994). Optimal use of the sce-ua global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(3):265–284.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., et al. (2012). Gfdl's esm2 global coupled climate-carbon earth system models. part i: Physical formulation and baseline simulation characteristics. *Journal of Climate*, 25(19):6646–6665.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of maxent for ecologists. *Diversity and Distributions*, 17(1):43–57.

Eltahir, E. A. and Bras, R. L. (1994). Precipitation recycling in the amazon basin. *Quarterly Journal of the Royal Meteorological Society*, 120(518):861–880.

Entekhabi, D. and Eagleson, P. S. (1989). Land surface hydrology parameterization for atmospheric general circulation models including subgrid scale spatial variability. *Journal of climate*, 2(8):816–831.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Famiglietti, J. (2014). The global groundwater crisis. *Nature Climate Change*, 4(11):945–948.

Famiglietti, J., Lo, M., Ho, S., Bethune, J., Anderson, K., Syed, T., Swenson, S., De Linage, C., and Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in california's central valley. *Geophysical Research Letters*, 38(3).

Fan, Y., Huang, G., Baetz, B., Li, Y., Huang, K., Li, Z., Chen, X., and Xiong, L. (2016). Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method. *Environmental Modelling & Software*, 86:30–49.

Fang, K.-T., Li, R., and Sudjianto, A. (2005). *Design and modeling for computer experiments*. CRC Press.

Fasham, M., Ducklow, H., and McKelvie, S. (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, 48(3):591–639.

Faunt, C. (2009). Groundwater availability of the central valley aquifer, california.

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T. (2016). Water availability and land subsidence in the central valley, california, usa. *Hydrogeology Journal*, 24(3):675–684.

Feddes, R. A., Kowalik, P. J., Zaradny, H., et al. (1978). *Simulation of field water use and crop yield*. Centre for Agricultural Publishing and Documentation.

Feynman, R. and Vernon Jr., F. (1963). The theory of a general quantum system interacting with a linear dissipative system. *Annals of Physics*, 24:118–173.

Fisher, J., Sitch, S., Malhi, Y., Fisher, R., Huntingford, C., and Tan, S.-Y. (2010a). Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retranslocation, and fixation. *Global Biogeochemical Cycles*, 24(1).

Fisher, R., McDowell, N., Purves, D., Moorcroft, P., Sitch, S., Cox, P., Huntingford, C., Meir, P., and Ian Woodward, F. (2010b). Assessing uncertainties in a second-generation dynamic vegetation model caused by ecological scale limitations. *New Phytologist*, 187(3):666–681.

Fisher, R., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N., Knox, R., Koven, C., Holm, J., Rogers, B., et al. (2015). Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes. *Geosci. Model Dev. Discuss*, 8(4):3293–3357.

Francos, A., Elorza, F. J., Bouraoui, F., Bidoglio, G., and Galbiati, L. (2003). Sensitivity analysis of distributed environmental simulation models: understanding the model behaviour in hydrological studies at the catchment scale. *Reliability Engineering & System Safety*, 79(2):205–218.

Fraternali, P. (1999). Tools and approaches for developing data-intensive web applications: a survey. *ACM Computing Surveys (CSUR)*, 31(3):227–263.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al. (2006). Climate-carbon cycle feedback analysis: Results from the c4mip model intercomparison. *Journal of Climate*, 19(14):3337–3353.

Gardenass, A., Hopmans, J., Hanson, B., and Simunek, J. (2005). Two-dimensional modeling of nitrate leaching for various fertigation scenarios under micro-irrigation. *Agricultural water management*, 74(3):219–242.

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., et al. (2015). Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *Journal of Operational Oceanography*, 8(sup1):s168–s187.

Gelman, A., Roberts, G. O., Gilks, W. R., et al. (1996). Efficient metropolis jumping rules. *Bayesian statistics*, 5(599-608):42.

Gelman, A. and Rubin, D. B. (1992). A single series from the gibbs sampler provides a false sense of security. *Bayesian statistics*, 4:625–631.

Geromel, J. C. (1999). Optimal linear filtering under parameter uncertainty. *IEEE Transactions on Signal Processing*, 47(1):168–175.

Ghanem, R. G. and Spanos, P. D. (1991). Spectral stochastic finite-element formulation for reliability analysis. *Journal of Engineering Mechanics*, 117(10):2351–2372.

Granzow, G. D. (2014). A tutorial on adjoint methods and their use for data assimilation in glaciology. *Journal of Glaciology*, 60(221):440–446.

Green, S. and Clothier, B. (1988). Water use of kiwifruit vines and apple trees by the heat-pulse technique. *Journal of Experimental Botany*, 39(1):115–123.

Groenendijk, M., Dolman, A., Van der Molen, M., Leuning, R., Arneth, A., Delpierre, N., Gash, J., Lindroth, A., Richardson, A., Verbeeck, H., et al. (2011). Assessing parameter variability in a photosynthesis model within and between plant functional types using global fluxnet eddy covariance data. *Agricultural and forest meteorology*, 151(1):22–38.

Gupta, H. V., Bastidas, L., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. (1999). Parameter estimation of a land surface scheme using multicriteria methods. *Journal of Geophysical Research: Atmospheres*, 104(D16):19491–19503.

Gupta, H. V., Wagener, T., and Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22(18):3802–3813.

Haaker, M. and Verheijen, P. (2004). Local and global sensitivity analysis for a reactor design with parameter uncertainty. *Chemical Engineering Research and Design*, 82(5):591–598.

Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14(3):375–396.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, pages 223–242.

Hanski, I., Turchin, P. K., and Korpimäki, E. (1993). Population oscillations of boreal rodents: regulation by mustelid predators leads to chaos. *Nature*, 364:232–235.

Harbaugh, A. W., Banta, E. R., Hill, M. C., and McDonald, M. G. (2000). Modflow-2000, the u. s. geological survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. *Open-file Report. U. S. Geological Survey*, (92):134.

Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., and et al. (2013). Station for measuring ecosystem-atmosphere relations: Smear. In Harti, P., Heliövaara, K., and Kulmala, L., editors, *Physical and Physiological Forest Ecology*, pages 471–487. Springer Netherlands, Dordrecht.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hays, G. C., Richardson, A. J., and Robinson, C. (2005). Climate change and marine plankton. *Trends in ecology & evolution*, 20(6):337–344.

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3):327–367.

Hickler, T., Smith, B., Sykes, M. T., Davis, M. B., Sugita, S., and Walker, K. (2004). Using a generalized vegetation model to simulate vegetation dynamics in northeastern usa. *Ecology*, 85(2):519–530.

Hölttä, T., Mencuccini, M., and Nikinmaa, E. (2009). Linking phloem function to structure: analysis with a coupled xylem–phloem transport model. *Journal of theoretical biology*, 259(2):325–337.

Houghton, R., Lawrence, K., Hackler, J., and Brown, S. (2001). The spatial distribution of forest biomass in the brazilian amazon: a comparison of estimates. *Global Change Biology*, 7(7):731–746.

Hsieh, L.-S. and Wang, R.-Y. (1999). A semi-distributed parallel-type linear reservoir rainfall-runoff model and its application in taiwan. *Hydrological Processes*, 13(8):1247–1268.

Huntingford, C., Zelazowski, P., Galbraith, D., Mercado, L. M., Sitch, S., Fisher, R., Lomas, M., Walker, A. P., Jones, C. D., Booth, B. B., et al. (2013). Simulated resilience of tropical rainforests to co2-induced climate change. *Nature Geoscience*, 6(4):268–273.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., et al. (2013). The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360.

Issa, R. I. (1986). Solution of the implicitly discretised fluid flow equations by operator-splitting. *Journal of computational physics*, 62(1):40–65.

Jackson, R. B., Jobbágy, E. G., and Nosetto, M. D. (2009). Ecohydrology in a human-dominated landscape. *Ecohydrology*, 2(3):383–389.

Jahanshiri, E. and Shariff, A. R. M. (2014). Developing web-based data analysis tools for precision farming using r and shiny. In *IOP Conference Series: Earth and Environmental Science*, volume 20, page 012014. IOP Publishing.

Janott, M., Gayler, S., Klier, C., and Priesack, E. (2009). A one-dimensional model of water flow in soil-plant systems. In *EGU General Assembly Conference Abstracts*, volume 11, page 4357.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London a: mathematical, physical and engineering sciences*, volume 186, pages 453–461. The Royal Society.

Jones, H. G. (2013). *Plants and microclimate: a quantitative approach to environmental plant physiology*. Cambridge university press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Katz, R. W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, 20(2):167–185.

Keating, E. H., Doherty, J., Vrugt, J. A., and Kang, Q. (2010). Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research*, 46(10).

Keenan, T. F., Carbone, M. S., Reichstein, M., and Richardson, A. D. (2011). The model–data fusion pitfall: assuming certainty in an uncertain world. *Oecologia*, 167(3):587–597.

Kendall, B. E., Briggs, C. J., Murdoch, W. W., Turchin, P., Ellner, S. P., McCauley, E., et al. (1999). Why do populations cycle? a synthesis of statistical and mechanistic modeling approaches. *Ecology*, 80(6):1789–1805.

Kioutsioukis, I., Tarantola, S., Saltelli, A., and Gatelli, D. (2004). Uncertainty and global sensitivity analysis of road transport emission estimates. *Atmospheric Environment*, 38(38):6609–6620.

Kitajima, K. and Poorter, L. (2010). Tissue-level leaf toughness, but not lamina thickness, predicts sapling leaf lifespan and shade tolerance of tropical tree species. *New Phytologist*, 186(3):708–721.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water resources research*, 22(4):499–507.

Kizito, F., Campbell, C., Campbell, G., Cobos, D., Teare, B., Carter, B., and Hopmans, J. (2008). Frequency, electrical conductivity and temperature analysis of a low-cost capacitance soil moisture sensor. *Journal of Hydrology*, 352(3):367–378.

Kleijnen, J. P. (1975). A comment on blanning's 'metamodel for sensitivity analysis: the regression metamodel in simulation'. *Interfaces*, 5(3):21–23.

Kleijnen, J. P. (2010). Design and analysis of computational experiments: overview. In *Experimental Methods for the Analysis of Optimization Algorithms*, pages 51–72. Springer.

Konikow, L. F. and Kendy, E. (2005). Groundwater depletion: A global problem. *Hydrogeology Journal*, 13(1):317–320.

Kramer, P. J. and Boyer, J. S. (1995). *Water relations of plants and soils*. Academic press.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(1).

Kuhnt, S. and Steinberg, D. M. (2010). Design and analysis of computer experiments. *AStA Advances in Statistical Analysis*, 94(4):307–309.

Kumagai, T. (2001). Modeling water transportation and storage in sapwood model development and validation. *Agricultural and Forest Meteorology*, 109(2):105 – 115.

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., Lighty, L., Eastman, J. L., Doty, B., Dirmeyer, P., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental modelling & software*, 21(10):1402–1415.

Kutzbach, J., Bonan, G., Foley, J., and Harrison, S. (1996). Vegetation and soil feedbacks on the response of the african monsoon to orbital forcing in the early to middle holocene.

Lacointe, A. and Minchin, P. E. (2008). Modelling phloem and xylem transport within a complex architecture. *Functional Plant Biology*, 35(10):772–780.

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage markov chain monte carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5):2664–2682.

Lange, O. L., Kappen, L., Schulze, E.-D., et al. (1976). *Water and plant life. Problems and modern approaches.* Springer-Verlag.

Larcher, W. (2003). *Physiological plant ecology: ecophysiology and stress physiology of functional groups*. Springer Science & Business Media.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the community land model. *Journal of Advances in Modeling Earth Systems*, 3(1).

Lawson, L. M., Hofmann, E. E., and Spitz, Y. H. (1996). Time series sampling and data assimilation

in a simple marine ecosystem model. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2):625–651.

Lawson, L. M., Spitz, Y. H., Hofmann, E. E., and Long, R. B. (1995). A data assimilation technique applied to a predator-prey model. *Bulletin of Mathematical Biology*, 57(4):593–617.

Le Maître, O. and Knio, O. M. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media.

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., and Dietze, M. C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, 83(2):133–154.

Leuning, R. (2002). Temperature dependence of two parameters in a photosynthesis model. *Plant, Cell & Environment*, 25(9):1205–1210.

Levy, S. and Steinberg, D. M. (2010). Computer experiments: a review. *AStA Advances in Statistical Analysis*, 94(4):311–324.

Lo, M.-H. and Famiglietti, J. S. (2013). Irrigation in california's central valley strengthens the southwestern us water cycle. *Geophysical Research Letters*, 40(2):301–306.

Lu, Y. and Mohanty, S. (2001). Sensitivity analysis of a complex, proposed geologic waste disposal system using the fourier amplitude sensitivity test method. *Reliability Engineering & System Safety*, 72(3):275–291.

Luo, Y. (2007). Terrestrial carbon–cycle feedback to climate warming. *Annu. Rev. Ecol. Evol. Syst.*, 38:683–712.

Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., and Schimel, D. S. (2011). Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications*, 21(5):1429–1442.

Luo, Y. and Schimel, D. S. (2011). Model improvement via data assimilation toward ecological forecasting 1. *Ecological Applications*, 21(5):1427–1428.

Machac, D., Reichert, P., Rieckermann, J., and Albert, C. (2016). Fast mechanism-based emulator of a slow urban hydrodynamic drainage simulator. *Environmental Modelling & Software*, 78:54–67.

Malhi, Y. and Davidson, E. A. (2009). Biogeochemistry and ecology of terrestrial ecosystems of amazonia. *Amazonia and global change*, pages 293–297.

Malhi, Y. and Grace, J. (2000). Tropical forests and atmospheric carbon dioxide. *Trends in Ecology & Evolution*, 15(8):332–337.

Man, J., Li, W., Zeng, L., and Wu, L. (2016). Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based kalman filter. *Advances in Water Resources*, 92:258–270.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M. (2015). A particle batch smoother approach to snow water equivalent estimation. *Journal of Hydrometeorology*, 16(4):1752–1772.

Martiny, A., Vrugt, J. A., Primeau, F. W., and Lomas, M. W. (2013). Regional variation in the particulate organic carbon to nitrogen ratio in the surface ocean. *Global Biogeochemical Cycles*, 27(3):723–731.

Masson, V., Champeaux, J.-L., Chauvin, F., Meriguet, C., and Lacaze, R. (2003). A global database of land surface parameters at 1-km resolution in meteorological and climate models. *Journal of climate*, 16(9):1261–1282.

Massoud, E.M., e. a. (In Review). Identification of key biological controls in forest dynamics based on a sensitivity analysis to the community land model with ecosystem demography, clm4.5(ed). *Journal of Advances in Modeling Earth Systems*, XX:xx–xx.

Maurer, E. P. and Hidalgo, H. G. (2008). Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods.

May, R. M. (1973). *Stability and complexity in model ecosystems*, volume 6. Princeton University Press.

McDowell, N., Hanson, P., Ibanez, I., Phillips, R. P., and Ryan, M. G. (2016). Physiological responses of forests to drought. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States).

McDowell, N. G., Williams, A., Xu, C., Pockman, W., Dickman, L., Sevanto, S., Pangle, R., Limousin, J., Plaut, J., Mackay, D., et al. (2015). Multi-scale predictions of massive conifer mortality due to chronic temperature rise. *Nature Climate Change*.

McGuire, V. (2009). Water-level changes in the high plains aquifer, predevelopment to 2007, 2005-06, and 2006-07. *Publications of the US Geological Survey*, page 17.

McMahon, S. M., Dietze, M. C., Hersh, M. H., Moran, E. V., and Clark, J. S. (2009). A predictive framework to understand forest responses to global change. *Annals of the New York Academy of Sciences*, 1162(1):221–236.

McNutt, M. (2014). The drought you can't see. *Science*, 345(6204):1543–1543.

Medlyn, B. E., De Kauwe, M. G., and Duursma, R. A. (2016). New developments in the effort to model ecosystems under water stress. *New Phytologist*, 212(1):5–7.

Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., Hickler, T., Jain, A. K., Luo, Y., Parton, W., et al. (2015). Using ecosystem experiments to improve vegetation models. *Nature Climate Change*, 5(6):528–534.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953).

Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Milly, P. and Shmakin, A. (2002). Global modeling of land water and energy balances. part i: The land dynamics (lad) model. *Journal of Hydrometeorology*, 3(3):283–299.

Mirfenderesgi, G., Bohrer, G., Matheny, A. M., Fatichi, S., Frasson, M., Prata, R., and Schäfer, K. V. (2016). Tree level hydrodynamic approach for resolving aboveground water storage and stomatal conductance and modeling the effects of tree hydraulic strategy. *Journal of Geophysical Research: Biogeosciences*, 121(7):1792–1813.

Mo, X., Chen, J. M., Ju, W., and Black, T. A. (2008). Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble kalman filter. *Ecological Modelling*, 217(1):157–173.

Moorcroft, P. R., Hurtt, G., and Pacala, S. W. (2001). A method for scaling vegetation dynamics: the ecosystem demography model (ed). *Ecological monographs*, 71(4):557–586.

Morton, K. W. and Mayers, D. F. (2005). *Numerical solution of partial differential equations: an introduction*. Cambridge university press.

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res*, 12(3):513–522.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part i?a discussion of principles. *Journal of hydrology*, 10(3):282–290.

Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J., et al. (2014). The role of data assimilation in predictive ecology. *Ecosphere*, 5(5):1–16.

Nobile, F., Tempone, R., and Webster, C. G. (2008a). An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442.

Nobile, F., Tempone, R., and Webster, C. G. (2008b). A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345.

Noilhan, J. and Planton, S. (1989). A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117(3):536–549.

North, G. B. and Nobel, P. S. (1997). Drought-induced changes in soil contact and hydraulic conductivity for roots of opuntia ficus-indica with and without rhizosheaths. *Plant and Soil*, 191(2):249–258.

Noy-Meir, I. (1973). Desert ecosystems: environment and producers. *Annual review of ecology and systematics*, pages 25–51.

O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1):201–203.

Overpeck, J. T., Rind, D., and Goldberg, R. (1990). Climate-induced changes in forest disturbance and vegetation.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.

Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., and Burlando, P. (2013). Sensitivity analysis of a

process-based ecosystem model: Pinpointing parameterization and structural issues. *Journal of Geophysical Research: Biogeosciences*, 118(2):505–528.

Peng, C., Guiot, J., Wu, H., Jiang, H., and Luo, Y. (2011). Integrating models with data in ecology and palaeoecology: advances towards a model–data fusion approach. *Ecology Letters*, 14(5):522–536.

Petchey, O. L., Pontarp, M., Massie, T. M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18(7):597–611.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79:214–232.

Powell, T. L., Galbraith, D. R., Christoffersen, B. O., Harper, A., Imbuzeiro, H., Rowland, L., Almeida, S., Brando, P. M., Costa, A. C. L., Costa, M. H., et al. (2013). Confronting model predictions of carbon fluxes with measurements of amazon forests subjected to experimental drought. *New Phytologist*, 200(2):350–365.

PRESENTERS, T. C. (2012). Californiawater plan, update 2013.

Prieto, I., Armas, C., and Pugnaire, F. I. (2012). Water release through plant roots: new insights into its consequences at the plant and ecosystem level. *New Phytologist*, 193(4):830–841.

Qian, T., Dai, A., Trenberth, K. E., and Oleson, K. W. (2006). Simulation of global land surface conditions from 1948 to 2004. part i: Forcing data and evaluations. *Journal of Hydrometeorology*, 7(5):953–975.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B. E., Disney, M., and Bowyer, P. (2008).

Assimilating canopy reflectance data into an ecosystem model with an ensemble kalman filter. *Remote Sensing of Environment*, 112(4):1347–1364.

Räisänen, J., Hansson, U., Ullerstig, A., Döscher, R., Graham, L., Jones, C., Meier, H., Samuelsson, P., and Willén, U. (2004). European climate in the late twenty-first century: regional simulations with two driving global models and two forcing scenarios. *Climate dynamics*, 22(1):13–31.

Ratto, M., Castelletti, A., and Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, 34:1–4.

Razavi, S., Tolson, B. A., and Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48(7):1–32.

Regis, R. G. and Shoemaker, C. A. (2007a). Improved strategies for radial basis function methods for global optimization. *Journal of Global Optimization*, 37(1):113–135.

Regis, R. G. and Shoemaker, C. A. (2007b). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS Journal on Computing*, 19(4):497–509.

Reichman, O. J., Jones, M. B., and Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018):703–705.

Reynolds, W., Elrick, D., Youngs, E., Amoozegar, A., Booltink, H., and Bouma, J. (2002). 3.4 saturated and field-saturated water flow parameters. *Methods of soil analysis, Part*, 4:797–801.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, 1(5):318–333.

Richey, A. S., Thomas, B. F., Lo, M.-H., Famiglietti, J. S., Swenson, S., and Rodell, M. (2015a). Uncertainty in global groundwater storage estimates in a total groundwater stress framework. *Water resources research*, 51(7):5198–5216.

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., and Rodell, M. (2015b). Quantifying renewable groundwater stress with grace. *Water resources research*, 51(7):5217–5238.

Rings, J., Kamai, T., Kandelous, M., Hartsough, P., Simunek, J., Vrugt, J., and Hopmans, J. (2013). Bayesian inference of tree water relations using a soil-tree-atmosphere continuum model. *Procedia Environmental Sciences*, 19:26–36.

Rodell, M., Velicogna, I., and Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in india. *Nature*, 460(7258):999–1002.

Rodriguez-Iturbe, I. (2000). Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamics. *Water Resources Research*, 36(1):3–9.

Rodriguez-Iturbe, I., Porporato, A., Laio, F., and Ridolfi, L. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: I. scope and general outline. *Advances in Water Resources*, 24(7):695–705.

Rogers, A. (2014). The use and misuse of v c, max in earth system models. *Photosynthesis research*, 119(1-2):15–29.

Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., Caemmerer, S., Dietze, M. C., Kattge, J., Leakey, A. D., Mercado, L. M., Niinemets, Ü., et al. (2017). A roadmap for improving the representation of photosynthesis in earth system models. *New Phytologist*, 213(1):22–42.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Gonçalves, L. G. G., and Zeng, X. (2013). Towards a comprehensive approach to parameter estimation in land surface parameterization schemes. *Hydrological Processes*, 27(14):2075–2097.

Rowland, L., da Costa, A. C. L., Galbraith, D. R., Oliveira, R., Binks, O. J., Oliveira, A., Pullen,

A., Doughty, C., Metcalfe, D., Vasconcelos, S., et al. (2015). Death from drought in tropical forests is triggered by hydraulics not carbon starvation. *Nature*, 528(7580):119–122.

Running, S., Baldocchi, D., Turner, D., Gower, S., Bakwin, P., and Hibbard, K. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and eos satellite data. *Remote Sensing of Environment*, 70(1):108–127.

Saad, G. and Ghanem, R. (2009). Characterization of reservoir simulation models using a polynomial chaos-based ensemble kalman filter. *Water Resources Research*, 45(4):n/a–n/a. W04417.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, pages 409–423.

Sadegh, M. and Vrugt, J. A. (2014). Approximate bayesian computation using markov chain monte carlo simulation: Dream (abc). *Water Resources Research*, 50(8):6767–6787.

Sadegh, M., Vrugt, J. A., Xu, C., and Volpi, E. (2015). The stationarity paradigm revisited: Hypothesis testing using diagnostics, summary metrics, and dream (abc). *Water Resources Research*, 51(11):9207–9231.

Sala, O. and Lauenroth, W. (1982). Small rainfall events: an ecological role in semiarid regions. *Oecologia*, 53(3):301–304.

Salati, E. (1987). The forest and the hydrological cycle.

Saltelli, A., Chan, K., Scott, E. M., et al. (2000). *Sensitivity analysis*, volume 1. Wiley New York.

Saltelli, A. and Tarantola, S. (2002). On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, 97(459):702–709.

Saltelli, A., Tarantola, S., and Chan, K.-S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56.

Santner, T. J., Williams, B. J., and Notz, W. I. (2013). *The design and analysis of computer experiments*. Springer Science & Business Media.

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., and Thornton, P. (2014). Dimensionality reduction for complex models via bayesian compressive sensing. *International Journal for Uncertainty Quantification*, 4(1).

Sato, H., Itoh, A., and Kohyama, T. (2007). Seib–dgvm: A new dynamic global vegetation model using a spatially explicit individual-based approach. *Ecological Modelling*, 200(3):279–307.

Savtchenko, A. K., Huffman, G., and Vollmer, B. (2015). Assessment of precipitation anomalies in california using trmm and merra data. *Journal of Geophysical Research: Atmospheres*, 120(16):8206–8215.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B. (2012a). Groundwater depletion and sustainability of irrigation in the us high plains and central valley. *Proceedings of the national academy of sciences*, 109(24):9320–9325.

Scanlon, B. R., Jolly, I., Sophocleous, M., and Zhang, L. (2007). Global impacts of conversions from natural to agricultural ecosystems on water resources: Quantity versus quality. *Water Resources Research*, 43(3).

Scanlon, B. R., Longuevergne, L., and Long, D. (2012b). Ground referencing grace satellite estimates of groundwater storage changes in the california central valley, usa. *Water Resources Research*, 48(4).

Schaeffer, D. J., Herricks, E. E., and Kerster, H. W. (1988). Ecosystem health: I. measuring ecosystem health. *Environmental Management*, 12(4):445–455.

Schaibly, J. H. and Shuler, K. E. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. ii applications. *The Journal of Chemical Physics*, 59(8):3879–3888.

Scharnagl, B., Vrugt, J., Vereecken, H., and Herbst, M. (2011). Inverse modelling of in situ soil water dynamics: Investigating the effect of different prior distributions of the soil hydraulic parameters. *Hydrology and Earth System Sciences*, 15(10).

Scheiter, S., Langan, L., and Higgins, S. I. (2013). Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist*, 198(3):957–969.

Schlesinger, W. H., Reynolds, J., Cunningham, G. L., Huenneke, L., Jarrell, W., Virginia, R., and Whitford, W. (1990). Biological feedbacks in global desertification. *Science*, 247(4946):1043–1048.

Scholes, R. J. and Walker, B. H. (2004). *An African savanna: synthesis of the Nylsvley study*. Cambridge University Press.

Schoups, G., Hopmans, J. W., Young, C. A., Vrugt, J. A., Wallender, W. W., Tanji, K. K., and Panday, S. (2005). Sustainability of irrigated agriculture in the san joaquin valley, california. *Proceedings of the National Academy of Sciences*, 102(43):15352–15356.

Schoups, G. and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water Resources Research*, 46(10):1–17.

Schulte, P. J. and Brooks, J. R. (2003). Branch junctions and the flow of water through xylem in douglas-fir and ponderosa pine stems. *Journal of Experimental Botany*, 54(387):1597–1605.

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., Barr, A., Black, T. A., Chen, G., Chen, J. M., et al. (2010). A model-data intercomparison of co2 exchange

across north america: Results from the north american carbon program site synthesis. *Journal of Geophysical Research: Biogeosciences*, 115(G3).

Sen, O. L., Bastidas, L. A., Shuttleworth, W. J., Yang, Z.-L., Gupta, H. V., and Sorooshian, S. (2001). Impact of field-calibrated vegetation parameters on gcm climate simulations. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1199–1223.

Serbin, S. P., Singh, A., Desai, A. R., Dubois, S. G., Jablonski, A. D., Kingdon, C. C., Kruger, E. L., and Townsend, P. A. (2015). Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. *Remote Sensing of Environment*, 167:78–87.

Shmida, A., Evenari, M., and Noy-Meir, I. (1986). Hot desert ecosystems: an integrated view. *Ecosystems of the World*, 12:379–387.

Shuttleworth, W. J. (1988). Macrohydrology?the new challenge for process hydrology. *Journal of Hydrology*, 100(1):31–56.

Siebert, S., Burke, J., Faures, J.-M., Frenken, K., Hoogeveen, J., Döll, P., and Portmann, F. T. (2010). Groundwater use for irrigation–a global inventory. *Hydrology and Earth System Sciences*, 14(10):1863–1880.

Simmons, A. and Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128(580):647–677.

Simuunek, J., van Genuchten, M. T., and Sejna, M. (2008). Development and applications of the hydrus and stanmod software packages and related codes. *Vadose Zone Journal*, 7(2):587–600.

Singh, A., Serbin, S. P., McNeil, B. E., Kingdon, C. C., and Townsend, P. A. (2015). Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications*, 25(8):2180–2197.

Siqueira, M., Katul, G., and Porporato, A. (2008). Onset of water stress, hysteresis in plant conductance, and hydraulic lift: scaling soil water dynamics from millimeters to meters. *Water Resources Research*, 44(1):1–14.

Sitch, S., Huntingford, C., Gedney, N., Levy, P., Lomas, M., Piao, S., Betts, R., Ciais, P., Cox, P., Friedlingstein, P., et al. (2008). Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (dgvms). *Global Change Biology*, 14(9):2015–2039.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M. T., et al. (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.

Sitz, A., Schwarz, U., Kurths, J., and Voss, H. (2002). Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Physical Review E*, 66(1):016210.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Dokl. Akad. Nauk SSSR*, volume 4, page 123.

Sneed, M., Brandt, J., and Solt, M. (2013). *Land subsidence along the Delta-Mendota Canal in the northern part of the San Joaquin Valley, California, 2003-10*. US Department of the Interior, US Geological Survey.

Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1):271–280.

Sorooshian, S. and Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, 16(2):430–442.

Steppe, K., De Pauw, D. J., Lemeur, R., and Vanrolleghem, P. A. (2006). A mathematical model linking tree sap flow dynamics to daily stem diameter fluctuations and radial stem growth. *Tree physiology*, 26(3):257–273.

Sterck, F., Markesteijn, L., Toledo, M., Schieving, F., and Poorter, L. (2014). Sapling performance along resource gradients drives tree species distributions within and across tropical forests. *Ecology*, 95(9):2514–2525.

Strogatz, S. H. et al. (1995). Nonlinear dynamics-ordering chaos with disorder. *Nature*, 378(6556):444–444.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979.

Ter Braak, C. and Van Tongeren, O. (1995). *Data analysis in community and landscape ecology*. Cambridge University Press.

Ter Braak, C. J. (2006). A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249.

Thonicke, K., Venevsky, S., Sitch, S., and Cramer, W. (2001). The role of fire disturbance for global vegetation dynamics: coupling fire into a dynamic global vegetation model. *Global Ecology and Biogeography*, 10(6):661–677.

Triantafyllou, G., Hoteit, I., Luo, X., Tsiaras, K., and Petihakis, G. (2013). Assessing a robust

ensemble-based kalman filter for efficient ecosystem data assimilation of the cretan sea. *Journal of Marine Systems*, 125:90–100.

Tyree, M. T. (1988). A dynamic model for water flow in a single tree: evidence that models must account for hydraulic architecture. *Tree Physiology*, 4(3):195–217.

(Unep), U. N. E. P. (2013). *Global Environment Outlook 2000*, volume 1. Routledge.

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5):892–898.

van Gigch, J. P. (1991). Diagnosis and metamodeling of system failures. In *System Design Modeling and Metamodeling*, pages 297–310. Springer.

Vandermeer, J. (1982). On the resolution of chaos in population models. *Theoretical Population Biology*, 22(1):17–27.

Vandermeer, J. (2004). Coupled oscillations in food webs: balancing competition and mutualism in simple ecological models. *The American Naturalist*, 163(6):857–867.

Vermeer, M. and Rahmstorf, S. (2009). Global sea level linked to global temperature. *Proceedings of the National Academy of Sciences*, 106(51):21527–21532.

Vrugt, J., Hopmans, J., and Simunek, J. (2001a). Calibration of a two-dimensional root water uptake model. *Soil Science Society of America Journal*, 65(4):1027–1037.

Vrugt, J., Wijk, M. v., Hopmans, J. W., and Šimunek, J. (2001b). One-, two-, and three-dimensional root water uptake functions for transient modeling. *Water Resources Research*, 37(10):2457–2470.

Vrugt, J. A. (2016). Markov chain monte carlo simulation using the dream software package: Theory, concepts, and matlab implementation. *Environmental Modelling & Software*, 75:273–316.

Vrugt, J. A., Bouten, W., Dekker, S. C., and Musters, P. A. (2002a). Transpiration dynamics of an austrian pine stand and its forest floor: identifying controlling conditions using artificial neural networks. *Advances in water resources*, 25(3):293–303.

Vrugt, J. A., Bouten, W., Gupta, H. V., and Sorooshian, S. (2002b). Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resources Research*, 38(12).

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41(1). W01017.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8).

Vrugt, J. A. and Robinson, B. A. (2007a). Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences*, 104(3):708–711.

Vrugt, J. A. and Robinson, B. A. (2007b). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and bayesian model averaging. *Water Resources Research*, 43(1).

Vrugt, J. A. and Sadegh, M. (2013). Toward diagnostic model calibration and evaluation: Approximate bayesian computation. *Water Resources Research*, 49(7):4335–4345.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., and Higdon, D. (2009). Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., and Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, 44(12).

Wada, Y., van Beek, L. P., van Kempen, C. M., Reckman, J. W., Vasak, S., and Bierkens, M. F. (2010). Global depletion of groundwater resources. *Geophysical research letters*, 37(20).

Wahba, G. (1990). *Spline models for observational data*. SIAM.

Wang, G., Gertner, G., Liu, X., and Anderson, A. (2001). Uncertainty assessment of soil erodibility factor for revised universal soil loss equation. *Catena*, 46(1):1–14.

Waring, R., Landsberg, J., and Williams, M. (1998). Net primary production of forests: a constant fraction of gross primary production? *Tree physiology*, 18(2):129–134.

Waring, R., Whitehead, D., and Jarvis, P. (1979). The contribution of stored water to transpiration in scots pine. *Plant, Cell & Environment*, 2(4):309–317.

Waring, R. H. and Running, S. W. (2010). *Forest ecosystems: analysis at multiple scales*. Elsevier.

Werth, D. and Avissar, R. (2004). The regional evapotranspiration of the amazon. *Journal of Hydrometeorology*, 5(1):100–109.

Williams, M., Richardson, A., Reichstein, M., Stoy, P., Peylin, P., Verbeeck, H., Carvalhais, N., et al. (2009). Improving land surface models with fluxnet data. *Biogeosciences*, 6(7):1341–1359.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11(1):89–105.

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., and Kavanagh, R. P. (2003). The use of bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17(6):1579–1590.

Wood, E. F., Lettenmaier, D. P., and Zartarian, V. G. (1992). A land-surface hydrology parameterization with subgrid variability for general circulation models. *Journal of Geophysical Research: Atmospheres*, 97(D3):2717–2728.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L., Bierkens, M. F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water. *Water Resources Research*, 47(5).

Wright, S. J., Kitajima, K., Kraft, N. J., Reich, P. B., Wright, I. J., Bunker, D. E., Condit, R., Dalling, J. W., Davies, S. J., Díaz, S., et al. (2010). Functional traits and the growth–mortality trade-off in tropical trees. *Ecology*, 91(12):3664–3674.

Wu, H., Guiot, J., Peng, C., and Guo, Z. (2009). New coupled model used inversely for reconstructing past terrestrial carbon storage from pollen data: validation of model using modern data. *Global Change Biology*, 15(1):82–96.

Xiao, Y. and Friedrichs, M. A. (2014). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, 119(4):2691–2712.

Xiu, D. (2007). Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys*, 2(2):293–309.

Xiu, D. and Hesthaven, J. S. (2005). High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139.

Xiu, D. and Karniadakis, G. E. (2002). The wiener–askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644.

Xu, C., Fisher, R., Wullschleger, S. D., Wilson, C. J., Cai, M., and McDowell, N. G. (2012).

Toward a mechanistic modeling of nitrogen limitation on vegetation dynamics. *PloS One*, 7(5):e37914–e37914.

Xu, C. and Gertner, G. (2007). Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics & Data Analysis*, 51(12):5579–5590.

Xu, C. and Gertner, G. (2011). Understanding and comparisons of different sampling approaches for the fourier amplitudes sensitivity test (fast). *Computational statistics & data analysis*, 55(1):184–198.

Xu, C. and Gertner, G. Z. (2009). Uncertainty analysis of transient population dynamics. *ecological modelling*, 220(3):283–293.

Xu, C., Gertner, G. Z., and Scheller, R. M. (2009). Uncertainties in the response of a forest landscape to global climatic change. *Global Change Biology*, 15(1):116–131.

Xu, C., McDowell, N. G., Sevanto, S., and Fisher, R. A. (2013). Our limited ability to predict vegetation dynamics under water stress. *New Phytologist*, 200(2):298–300.

Ye, M., Meyer, P. D., and Neuman, S. P. (2008). On model selection criteria in multimodel analysis. *Water Resources Research*, 44(3).

Yeh, P. J.-F., Swenson, S., Famiglietti, J., and Rodell, M. (2006). Remote sensing of groundwater storage changes in illinois using the gravity recovery and climate experiment (grace). *Water Resources Research*, 42(12).

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F. (2005). Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochemical Cycles*, 19(3).

Zeng, N., Neelin, J. D., Lau, K.-M., and Tucker, C. J. (1999). Enhancement of interdecadal climate variability in the sahel by vegetation interaction. *Science*, 286(5444):1537–1540.

Zeng, X. (2001). Global vegetation root distribution for land modeling. *Journal of Hydrometeorology*, 2(5):525–530.

Zeng, X., Ye, M., Burkardt, J., Wu, J., Wang, D., and Zhu, X. (2016). Evaluating two sparse grid surrogates and two adaptation criteria for groundwater bayesian uncertainty quantification. *Journal of Hydrology*, 535:120–134.

Zhang, X., Srinivasan, R., and Van Liew, M. (2009). Approximating swat model using artificial neural network and support vector machine1. *JAWRA Journal of the American Water Resources Association*, 45(2):460–474.

Zobitz, J., Desai, A., Moore, D., and Chadwick, M. (2011). A primer for data assimilation with ecological models using markov chain monte carlo (mcmc). *Oecologia*, 167(3):599–611.

Zweifel, R., Steppe, K., and Sterck, F. J. (2007). Stomatal regulation by microclimate and tree water relations: interpreting ecophysiological field data with a hydraulic plant model. *Journal of Experimental Botany*, 58(8):2113–2131.