Sandia National Laboratories
PO Box 969, MS 9159
Livermore, CA 94550
csafta@sandia.gov

April 30, 2015

Geoscientific Model Development
European Geosciences Union

Dear Editor:

We herewith submit our revised manuscript, "Global Sensitivity Analysis, Probabilistic Calibration, and Predictive Assessment for the Data Assimilation Linked Ecosystem Carbon Model", by Safta, Ricciuto, Sargsyan, Debusschere, Najm, Williams, and Thornton for publication in the Geoscientific Model Development. The original manuscript had been returned to us after a round of reviews, with requests for major modifications. The reviewers had also posed some questions to us.

We have gone over the reviewer comments in detail. We have benefited much from many of the reviewer comments, shown in cursive font in this letter, and have incorporated them in the revised text. Our detailed responses are shown in normal font immediately following each question.

We feel that we have addressed all comments, and have done the major revisions to the paper consistent with the Editor's and reviewers' requests. Some of the major changes are briefly described on the next page.

We respectfully ask the Editor to reconsider the revised paper.

Sincerely,

Cosmin Safta

A brief outline of major changes in the current manuscript compared to the previous version is provided below.

- We reduced the size of the Introduction. We refocused the paper to emphasize the comparison between the steady and transient model setups. This comparison is now the focus of all Sections in the revised manuscript.

- In the GSA section we replaced the discussion of first-order Sobol indices, S_i with total effect indices, S_i^T. We believe that total effect indices, which include the first order effects and the joint and higher order interaction effects, provide a better picture on which parameters matter for specific quantities of interest.

- In the current version of the manuscript we employ informative priors for all model parameters. These priors are described in Section 4.3. In addition to the model calibration study, these priors were also used in Section 3 for Global Sensitivity Analysis.

- The model error is no longer ignored in the current version of the manuscript. We discuss the modeling associated with the statistical model error term in Section 4.2. In Section 4.4.1 we present a convergence study for the parameters controlling the model error.

- Given the new emphasis on steady state/transient model setups, we removed the section Fisher Information matrix (FIM) and subsequent discussion and results based on FIM. Overall the length of the paper, not including figures and tables, increased by one page due to the substantial increase in the discussion in other parts of the paper.

- All figures and most of the associated text, except perhaps the sketches, were updated in the revised manuscripts to account for the changes in priors and model error term. We reduced the number of figures by two, from 19 to 17.

Response to Reviewer #1

1. *Unfortunately, there were a few substantial problems in the paper. First and most important is that model error was ignored during calibration. Second, prior information on model parameters was also ignored, which leads to a misleading uncertainty analysis and potentially biologically implausible parameter estimates (as an aside, the plausibility of model posterior estimates is never discussed or compared to data). Third, lacking an estimate of model error, predictive distributions are done by propagating observation error into the forecasts, which is inappropriate. Finally, overall the paper was too long, though in many places I felt like I was being hit by a barrage of indices that were all Methods and Results without Discussion.*

   We thank this reviewer for his critical assessment of the first version of our manuscript. We have addressed the two major comments, regarding model error and prior information. We updated the formulation to include these comments resulting in major changes throughout most of the manuscript. We provide detailed replies to specific comments by this reviewer, including a reply included in item #20 below to the third issue raised above.

2. *Pages 6895-6897: Too much background. Theres been multiple recent reviews of data assimilation in ecosystem models that can be pointed to for readers that want all the nitty gritty, so instead you should focus on your message/context.*

   The Introduction section in the current version of the manuscript is significantly more compact compared to the previous version. We eliminated unnecessary details, refer to a review paper, and provide a more focused description on the motivation for our study.

3. *Page 6897, line 25: Unlike the last 2 pages, this comment is unsupported. Indeed, there have been a number of recent papers performing detailed sensitivity and uncertainty analyses specifically in the context of leading up to data assimilation. This is worth mentioning because it is important to note that while the details of the workflow the authors put forth is unique, the general workflow they are following is definitely not, and thats a part of the literature that DOES need to be discussed if youre going to claim that what you are going is novel.*

   This comment is now reformulated to state that few, if any, studies focus on the steady vs transient assumptions for the ecosystem models.

4. *Page 6898, line 27: This bit is very important and much more rare (indeed, I know of examples where teams have taken one approach or the other, but Im not aware of a paper that compares the two explicitly), however in the context of this paragraph I had no idea what you were talking about. It was only when I got to the methods that it became clear that you were discussing the alternatives of assuming spin up to steady-state vs. including the IC in the assimilation. This needs to be explained and highlighted more in the intro.*

   We now provide a more detailed explanation, both in the "Introduction" as well as in the following section, "Description of Carbon Cycle Model", on the two approaches for using DALEC, steady-state vs transient model setups.

5. *Page 6898, L10-12: Extraneous. That youre developing this workflow to apply it to more complex models like CLM is relevant, but tweaks to the model can be left to the methods*

   The description of changes to the DALEC model was moved to Section 2.

6. *Pg 6899, L12: Great, but the reader has no idea what UQTk is is it a project, a model, a conference, a piece of software? What does UQTk stand for? Given that the answer is that its a piece of software, push this into the methods, be more explicit about its use, and make the overarching workflow for performing this analysis public (even if the source code of the toolbox isnt). Having just that toolbox is insufficient to allow this analysis to be reproduced.*

   We have updated the description of the suite of tools in UQTk. The source code for this software toolbox is

available for download. We also state that additional scripts/wrappers specific to this paper are available upon request from the first author.

7. *I fundamentally disagree. A sensitivity analysis (even a global one) connects PERTURBATIONS in the model inputs to perturbations in the model output. What you describe is an uncertainty analysis. The fundamental difference is that to perform an UA rather than a SA you need to know the uncertainties in the parameters. But you dont have those uncertainties, you just have arbitrarily assigned uniform ranges. Personally, I dont think Sobol indices are an appropriate technique for SA since they are variance-based their interpretation only really makes sense for UA.*

   We modified the start of this paragraph to eliminate some ambiguity in the intent of this section. We agree with the reviewer that Sobol indices, being variance based, are not equivalent to indices from a perturbation-type sensitivity analysis over the supports of the given probability density functions of model parameters. At the same time, we submit that, in our experience, we have seen perturbative SA done only in local SA, while all GSA studies we have seen actually rely on variance-based probabilistic analysis. In fact, applying perturbative analysis with large perturbations associated with the support of parameter PDFs can be problematic, e.g. in cases where the PDF support is exceedingly large or infinite, as it would give sole attention to the bounds of the support. On the other hand, probabilistic variance-based GSA is naturally weighted by the input measure and well behaved for any input PDF. Given the predominance of the probabilistic framework for GSA, we respectfully propose to retain this framework here as is.

8. *Pg 6902, L15-20: The logic here is completely backwards. You state you are only given prior information on parameter bounds, but it is you that is only giving yourself that information. If you wanted to give yourself more information you easily could since Mats made the model and has been working with it for a decade. More to the point, for almost every parameter in DALEC there is more prior information available in the literature.*

   The prior ranges are chosen following previous studies (e.g. Fox *et. al.* [1]) and more broadly reflect temperate deciduous forests, since many carbon cycle models characterize sites by plant functional type (PFT) rather than focusing on specific species present at a site. We now employ informative priors for all model parameters using nominal PFT-level values and some basic assumptions about their uncertainty. Informative priors for the initial carbon pool amounts now use actual site-specific measurements. The text has been updated to reflect these changes.

9. *Pg 6903, L6: This info really needs to make it into the legends for these figures I looked at the figures first and was scratching my head since neither grey nor white was in the legend or caption*

   We have improved the colormap description in the figure captions.

10. *Pg 6905, L16: The assumption that NEE data are independent is not plausible and will result in considerably overconfident posterior distributions*

    The largest source of error in NEE measurements is associated with turbulent transport, which is expected to be independent (Hollinger and Richardson [2]). However, systematic errors associated with the eddy covariance method are likely, and while they are corrected when possible, some are likely to remain. Such errors will now be represented in the statistical model error term that we have added in this version. This should address the overconfidence problem but systematic data errors may contribute to biases in the posterior distributions of model parameters.

11. *Pg 6905, L18: The decision to neglect model error is completely inappropriate, will lead to incorrect posterior distributions, and renders your predictive intervals meaningless. This assumption implies that you believe that your process model is perfect and the only reason for deviations from observations is due to observation error, which is untenable. The inclusion of model error as a fit parameter is neither conceptually difficult or computationally costly updating the model error doesnt require model runs and the prior could easily be chosen to be conjugate to your Normal likelihood, allowing the update to be done*

using Gibbs Sampling

In the current version of this study we model the discrepancy between the model and the "truth" as a multivariate normal distribution with constant bias and square-exponential covariance matrix. The hyperparameters introduced by this model error term are estimated together with the original model parameters. We also discuss the convergence of these hyperparameters with respect to the bandwidth of the covariance matrix.

12. *L6906, L14: This is awesome. However, you dont state what your priors are. As I note above in the discussion of informed priors, you should actually be able to construct fairly informative priors for most of your C pools since the biometric data for Harvard Forest is pretty good and all public through the LTER.*

    We now use informative priors for all DALEC parameters, including the intial Carbon pools' amounts in the transient model setup. These parameters are provided in Table 4 in the current version of our manuscript.

13. *P6906, L20: Could you use more meaningful acronyms?? Also, these acronyms are not defined in the tables and figures that include them*

    We have changed the labels for the steady state/transient model setups to D_{ST} and D_{TR}. We are now providing explicit definitions in the figure/table captions.

14. *P6907 L1: D is replaced by a random vector of NEE observations This bit doesnt make sense to me, why would you be randomizing your NEE observations such that their was no pairing between observations and model predictions? Based on my reading theres no requirement for this randomization in Fisher Information approach.*

    and

15. *P6907 L12-13: Given that Sigma is just a constant with respect to the different model parameters (thetas), this standardization by...*

    The above comments pertain to the section on Fisher Information Matrix. This section was removed from the current version of the manuscript.

16. *Pg 6909, L8: high-lighted*

    The paragraph containing this typo was re-written in the current version of the manuscript.

17. *P6910, L9-12: This is a really neat trick.*

18. *P6911, L5-6: This is also a new and useful contribution (at the least, its an approach that I havent seen before in the ecosystem modeling literature)*

    We thank the reviewer for the above comments.

19. *P6912, L26-28: Neat. I've never seen a piecewise posterior before, but this is a great example of how input accuracy and if statements can impact posterior inference.*

    We thank the reviewer for this comment. In the current version of the manuscript the piecewise constant posterior is now piecewise (quasi) linear due to the inclusion of informative priors.

20. *P6915, L12-14: I dont think this analysis makes sense. First, observation error should not be propagated into a forecast. Model error, which wasnt estimated, DOES need to be propagated into the predictive distributions. Given that, Id strongly recommend that the authors restrict the current analysis to looking just at the models credible interval (parameter uncertainty) not predictive intervals (unless they end up quantifying model error as well)*

    Respectfully, we cannot agree with the reviewer on this point. The reviewer is under the misapprehension that the posterior predictive is for predictive forecasts with the model. In fact it is not. A predictive

forecast would be made with only the model-error term included, and is what we call the pushed forward posterior. The point of the posterior predictive is not so much prediction of QoIs with the calibrated model, but rather prediction of the noisy data itself as a diagnostic check on the quality of the statistical inference. One cannot do a posterior predictive check relevant to the measured noisy data without including the data noise term in the posterior predictive. In the current version of the manuscript we include the newly inferred model error together with the data noise in the posterior predictive check.

21. *P6916, L1: computed with KDE is unclear*

    This paragraph was re-written and this expression does not appear in the current version of the manuscript.

22. *P 6916, L14: CRPS is also new to me. I think this is cool, but in the end theres not much interpretation / discussion of the results. Needs to be a more clear set of take-home messages in this section*

    and

    P6917, L14: Likewise, CRPSS is new to me, and while it is interesting, like with CRPS it isnt explained or interpreted enough. If the only take home message is that there was 40% improvement, then you should make this section much shorter in order to get to that point more quickly

    We have now expanded the discussion of CRPS and CRPSS, in particular when comparing the steady vs transient model setups.

23. *Figure 2: what is NPP2 and why does leaf biomass not affect LAI?*

    One of the green arrows was missing in Fig. 2. A change in LAI leads to a change in *vc1* which is the leaf biomass. *NPP2* is a variable internal to DALEC representing the NPP available for allocation to stem and root pools after allocation to foliage. It is computed by subtracting the net change in LAI from the Net Primary Production (NPP).

24. *Figure 3-6: please put the months in order! Start with January, end with December rather than starting with November.*

    We have updated these figures per reviewer's request.

25. *Figure 4: Why is DALEC so sensitive to leaf fall in December, which is 2 months after leaf fall occurs?*

    The GPP average during the winter months is negligible, basically just noise of the same order of magnitude to machine epsilon. We have updated the plotting scripts to skip plotting sensitivity data when this information is not physical. Thank you for pointing this out.

References

[1] Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, Agricultural and Forest Meteorology, 149, 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.

[2] Hollinger, D. Y., and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree Physiology, 25, 873–885, 2005.

Response to Reviewer #2

We thank this reviewer for his critical assessment of our manuscript. Please find below our itemized replies to the issues raised by this reviewer.

1. *The use of uninformed priors greatly reduces the utility of the study as a number of studies have demonstrated ...*

   In the current version of the manuscript, we employ informative priors for all parameters.

2. *Overall the paper is too long, with too many figures especially given the lack of any meaningful discussion about much of the results.*

   We have eliminated the section presenting the Fisher information matrix and several paragraphs throughout the paper related to that discussion. We also eliminated several figures. On the other hand, in the new version we emphasize the comparison between steady state and transient model setups. Additionally, the use of informative priors lead to more discussion in several sections in the current version of the manuscript. Overall, the current version is shorter by about three pages and with two figures less than the previous version.

3. *Page 6895-6896: Much of the background section could be replaced by reference to a number of recent reviews of data assimilation techniques used with ecosystem models and UQ workflow tools. Instead more emphasize should be placed on the context of this particular study.*

   The Introduction section in the current version of the manuscript is significantly more compact compared to the previous version. We eliminated unecessary details, refer to a review paper, and provide a more focused description on the motivation for our study.

4. *Page 6897, line 27: The impacts of steady state/non-steady state assumption on SA and UQ are pretty interesting and deserve more discussion at this point. In this case this is spin-up v. initial conditions estimation (although you dont explain that here), but could mean other things.*

   We expanded the discussion in the Global Sensitivity Analysis section of the similarities and differences between steady state/transient assumption for DALEC.

5. *Page 6898, line 9: Dont say modified version without immediately describing the changes which is done in the following section*

   We have moved the discussion on the modifications made to DALEC to Section 2.

6. *Page 6899, line 12: What is UQTk v3.0? OM you follow the link and its software. This looks great, but how was this actually used? How could what youve done be reproduced? I think it would be of great interest to have considerable more details of this tool, and how ecosystem modelers can use it.*

   In the new version of the manuscript we provide a brief description on the set of software tools employed in this study.

7. *Page 6901, line 9: What you describe is an uncertainty analysis, not a sensititivity analysis (which is what you then do)*

   We modified the start of this paragraph to eliminate some ambiguity in the intent of this section. We agree with the reviewer that Sobol indices, being variance based, are not equivalent to results based on perturbation-type sensitivity analysis over the supports of the given probability density functions of model parameters. At the same time, we submit that, in our experience, we have seen perturbative SA done only in local SA, while all GSA studies we have seen actually rely on variance-based probabilistic analysis. In fact, applying perturbative analysis with large perturbations associated with the support of parameter PDFs can be problematic, e.g. in cases where the PDF support is exceedingly large or infinite, as it would give sole attention to the bounds of the support. On the other hand, probabilistic variance-based GSA is

naturally weighted by the input measure and well behaved for any input PDF. Given the predominance of the probabilistic framework for GSA, we respectfully propose to retain this framework here as is.

8. *Page 6902, line 19: See my overall comment there is considerable prior information that could be/should be used to inform priors in this study to give a meaningful and informative analysis.*

   We now use informative priors in the current version of the manuscript.

9. *Page 6903, line 27: Right, monthly timescales are not appropriate for analysis of controls over these large carbon pools. Why was a monthly timescale selected in the first place?*

   We originally selected a monthly timescale in order to study the NEE and GPP fluxes. While monthly timescales are too short for the Carbon pools, we believe that current results offer a nice confirmation for the longer timescales characteristic to the Carbon pools.

10. *Page 6906, line 6: How was steady-state defined in this case?*

    We added in the last paragraph of Section 2 a description on the criteria employed to decide when the model achieves a "numerical" steady state.

11. *Page 6906, line 6: 30-50 cycles Why is there is variability? Youre cycling the same climate files, in the same order? And have a definition of steady state, why isnt it a single, repeatable value?*

    While indeed we use the same climate file for all DALEC simulations, the model parameters values also play a role in the model behavior. In particular, the number of cycles required for spinup depends on the turnover time of the slow carbon pools (i.e. longer stem or SOM turnover requires more cycles). For some samples it take longer for DALEC to achieve a steady state compared to others. In general, we observed that for most runs, it takes between 30 and 50 cycles to achieve a steady state.

12. *Page 6906, line 14: Do you mean 1991 or should it be 1992 to correspond with the start of the met data?*

    We meant 1992. Fixed the typo.

13. *Page 6909, line 13: Why is this done sequentially like this?*

    This paragraph was removed from this version of the manuscript. We describe the sequential procedure in Section 4.4. The main reason is to facilitate an efficient MCMC sampling. In high-dimensional settings the posterior landscape can be very flat and require accurate starting covariances and sample points. This is achieved by starting in a lower-dimensional setting with parameters that are likely to be informed by the data, thus ensuring a good sample coverage for the intermediate posterior densities.

14. *Page 6913, line 8: paramaters typo*

    Fixed typo

15. *Page 6916, line 1: What is KDE here?*

    This context was removed from the current version of the manuscript

16. *Page 6918, line 3: This requires further elaboration/explanation.*

    We now employ informative priors, including for the Carbon pools for the transient model setup. The original statement referred to in this comment was removed

17. *Page 6919, line 27: Actually, it seems like these robust statistical methodologies that are sampling based cant be used with model like CLM thats why youre developing emulators!*

    This paragraph was removed from the current version of this manuscript.

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/07/29 7.12 Copernicus papers of the LaTeX class copernicus.cls.
Date: 30 April 2015

Global sensitivity analysis, probabilistic calibration, and predictive assessment for the Data Assimilation Linked Ecosystem Carbon model

C. Safta[1], D. M. Ricciuto[2], K. Sargsyan[1], B. Debusschere[1], H. N. Najm[1], M. Williams[3], and P. E. Thornton[2]

[1]Sandia National Labs, Livermore, CA 94551, USA
[2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[3]School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, EH9 EJN, UK

Correspondence to: C. Safta (csafta@sandia.gov)

Abstract

In this paper we propose a probabilistic framework for an uncertainty quantification study of a carbon cycle model. A and focus on the comparison between steady state and transient simulation setups. A Global Sensitivity Analysis (GSA) study indicates the parameters and parameter couplings that are important at different times of the year for Quantities of Interest obtained with the Data Assimilation Linked Ecosystem Carbon (DALEC) model. We then employ a Bayesian approach Bayesian approach and a statistical model error term to calibrate the parameters of DALEC using net ecosystem exchange observations at the Harvard Forest site. The calibration exercise is guided by GSA and by Fisher information matrix results that quantify the amount of information carried by the experimental data about specific model parameters. The calibration results results are employed in the second part of the paper to assess the predictive skill of the model via posterior predictive checks. These checks show a better performance for the non-steady state model during the growing season compared to the one employing steady state assumptions. Overall, this study leads to a 40% improvement in the predictive skill of DALEC and highlights the importance of considering correlations in the model parameters as informed by the data.

1 Introduction

Climate studies strongly depend on the modeling of the Carbon cycle. Carbon cycle models, in turn, strongly depend on the capability of current land models to simulate the terrestrial ecosystem and to capture C Carbon exchanges between land and atmosphere. There have been a significant number of studies looking to leverage the increasing amount of experimental observations and calibrate parameters in several terrestrial ecosystem models. These studies have faced a number of challenges related to handling data and measurement errors from multiple sources, formalizing model error, dealing with parameter observability and data sparsity, to name a few. In this paper we propose a probabilistic framework

to estimate parameters for a ~~-~~process-based ecosystem model. Representative studies, both probabilistic and non-probabilistic, are reviewed below.

Over the past two decades several studies employed data assimilation techniques to calibrate Carbon cycle models. Here we briefly discuss the works that motivated the current study. ~~Kaminski et al. (2002)~~ Kaminski et al. (2002, 2012) used an adjoint approach to infer model parameters for a ~~Simple Diagnostic Biosphere Model~~terrestrial biosphere model based on observational data streams. The variational data assimilation problem was formulated based on Bayes ~~formula~~ theorem with both the likelihood and the prior presumed Gaussian. ~~This results in a quadratic cost function that employs an L_2 regularization of the model parameters. This formulation led to optimal values for model parameters. The width of the approximate Gaussian distributions around these optimal values was sensitive to the covariance matrices assumed in the cost function. More recently, Kaminski et al. (2012) employed a similar framework to calibrate the process parameters of a terrestrial biosphere model against two observational data streams. The model~~ It was found that models employing optimized parameters ~~shows~~ show clear improvements when checked against independent observations compared to non-optimized parameters. ~~A similar approach was applied by Rayner et al. (2005) to study the space–time distribution of terrestrial carbon fluxes generated by a terrestrial carbon cycle data assimilation system. Tjiputra et al. (2007) employed an adjoint approach to estimate optimal values for 10 ecosystem control variables in an ocean general circulation model coupled with a carbon cycle model. The optimization problem is based on a quadratic misfit between the simulated surface chlorophyll and observations. Kuppel et al. (2012) used measurements of net CO_2 fluxes (NEE) and latent heat fluxes (LE) to constrain the parameters of a biogeochemical vegetation model. The optimization employed an L-BFGS algorithm for a quadratic cost function similar to the study by Kaminski et al. (2002) . They found that the simulation results are improved by using data from multiple sites, compared to single-site parameter optimization.~~ Similar approaches were employed by Rayner et al. (2005) , Tjiputra et al. (2007) , Kuppel et al. (2012) to estimate parameters of ecosystem models.

Some of the above studies start from a –Bayesian framework when setting the cost function for a –least-square fitting procedure. ~~These studies are based on a Gaussian assumption for the discrepancy between model outputs and observations, and they also employ Gaussian priors to help regularize the problem. However, the~~ The resulting probability densities for model parameters are approximated as multivariate Gaussian ~~distributions~~ densities near the Maximum a –Posteriori (MAP) estimate of the parameter values. This assumption is valid only in the vicinity of MAP values, unless the model is linear in all parameters. ~~In this paper we propose to employ a Bayesian framework to estimate parameters in the Data Assimilation Linked Ecosystem Carbon (DALEC) model (Williams et al., 2005) , without relying on Gaussian assumptions for posterior distributions.~~ Several studies in the past decade, some of which mentioned below, employed sampling techniques to explore non-Gaussian posterior distributions for parameters in ecosystem models.

Knorr and Kattge (2005)~~employed a Bayesian framework to calibrate the parameters of a Terrestrial Ecosystem Model (TEM). A Metropolis–Hastings~~ , Braswell et al. (2005) , Xu et al. (2006) employed Bayesian frameworks to estimate parameters of terrestrial ecosystem models. These studies employed Metropolis-Hastings Markov Chain Monte Carlo (MCMC) ~~approach was used~~ techniques to sample the posterior ~~distribution~~ density of model parameters ~~given a Gaussian likelihood~~ constructed based on eddy covariance measurements of ~~carbon and water fluxes . It was found that about 5 parameters were constrained by the available data and that uniform prior ranges had a strong impact on the posterior distributions. Braswell et al. (2005) performed a synthetic analysis of Net Ecosystem Exchange (NEE) of CO_2 at Harvard Forest using a simplified photosynthesis and evapo-transpiration model. In a Bayesian framework, they employed independent Gaussian daily discrepancies between model predictions and observations. The posterior distributions for modeled parameters, sampled with MCMC, were compared for several synthetic data sets to determine how much information the NEE observations carry about each parameter. A Bayesian framework was also employed by Xu et al. (2006) to study the posterior distributions of C transfer coefficients and pool sizes in a TEM, based on several data sets from the Duke Forest Free-Air CO_2 site.~~ Carbon fluxes as well as based

on synthetic datasets. Tang and Zhuang (2009) employed both Global Sensitivity Analysis (GSA) and a Bayesian framework to improve parameterization of a Terrestrial Ecosystem Model. This study employed Latin Hypercube Sampling from the prior distributions density of model parameters, and sample importance resampling , a sampling-importance resampling method to construct posterior distributions densities for model parameters, and to identify key parameters for the ecosystem model and their effect on seasonal C dynamics. Ricciuto et al. (2008) employed an MCMC approach to sample the posterior densities of key parameters for combined global-scale terrestrial and ocean carbon cycle models. The study found that temporal correlation has a significant impact on the calibrated parameters and subsequently on model predictions. A recent review by Zobitz et al. (2011) provides a primer on data assimilation studies with MCMC.

Several studies compared probabilistic and non-probabilistic several parameter estimation methods for terrestrial biogeochemical models. Several participants to Participants in the OptIC project (Trudinger et al. (2007)) presented results employing optimization, variational, and probabilistic methods. The main conclusion of the study was that modeling choices, i.e. the type of cost function for optimization methods, or the choice of densities for probabilistic methods, had a greater impact on the results than the choice of solution methods. sampling methods. Similarly, the REFLEX project (Fox et al., 2009) (Fox et al. (2009)) selected the DALEC v1 model (Williams et al., 2005) (Williams et al. (2005)) to assess the performance of several parameter estimation algorithms, using both synthetic and observed NEE and LAI data. This study found that it is difficult to analyze the performance of parameter estimation methods in the presence of noisy and sparse data, and that all methodologies should employ uncertainty models that are consistent with observations. Net Ecosystem Exchange (NEE) and Leaf Area Index (LAI) data. More recently, Ziehn et al. (2012) compared variational and probabilistic sampling techniques to estimate parameters for BETHY, a process-based model of the terrestrial biosphere. It was found that the Gaussian approximation is reasonable for most parameters. This study also indicates that probabilistic approaches can be prohibitively expensive for complex ecosystem models.

From this review, we noted a ~~-~~set of critical outstanding research questions in the ~~field of constraining C~~ context of constraining Carbon cycle models. First, few, if any, ~~C cycle models have had a complete parameter sensitivity analysis, particularly with respect to temporal dynamics. Such analyses are vital for organising effective parameter calibration . Second, few, if any, calibration~~ calibration studies have investigated steady state/~~non-steady state~~ transient assumptions. It is also important for the ecological community to understand how information content depends on model assumption, ~~e.g. steady state . Currently, there are no agreed approaches in this community for quantifying information content of data on parameters, or for estimating~~ *e.g.* steady state vs transient. Second, Carbon cycle models require a complete parameter sensitivity analysis, particularly with respect to temporal dynamics. Such analyses are vital for organising effective parameter calibration followed by an estimation of the predictive skill of ecosystem models.

In this paper we propose a ~~-~~Bayesian framework for the estimation of uncertainties in ecosystem land model parameters followed by a ~~-~~forward Uncertainty Quantification (UQ) study to examine the predictive capabilities of the model given the calibrated set of parameters. The Bayesian formulation provides a flexible framework for handling heterogeneous information, and allows for sequential updates of posterior distributions as the prior information is revised.

Figure 1 shows a schematic of this framework, consisting of two intrinsically connected workflows, for *Parameter Estimation* and *Forward UQ*. In this schematic, the ~~same ecosystem Carbon model~~ Data Assimilation Linked Ecosystem Carbon (DALEC) model (Williams et al. (2005)) is used for both the "Measurement Model" ~~$g()$~~ $m()$ and the "Computational Model" ~~$m()$~~. ~~The Carbon model is based on a modified version of the DALEC v1 model (Williams et al., 2005; Fox et al., 2009) . This version of DALEC has been modified to facilitate comparisons with the Community Land Model (Thornton et al., 2007) , and with the Local Terrestrial Ecosystem Carbon Model (Ricciuto et al., 2011) .~~[1] ~~The joint probability density for input parametersis estimated in a Bayesian framework. Bayesian~~

[1] ~~The source code for the modified DALEC version is available upon request from Daniel Ricciuto ()~~

methods provide a flexible framework for handling heterogeneous information, and allow for sequential updates of posterior distributions as the prior information is revised $f()$. We employ two model setups in our analysis. In the first approach, DALEC is run in a spinup mode until the Carbon pools reach a quasi steady state. In the second approach, each ecosystem model run consists of one cycle only. In this approach the Carbon pools are part of the investigation on model parameters, either for the purpose of estimating densities of model inputs or to propagate these densities forward to model outputs. More details on the steady state/transient model setups are provided in Section 2.

To facilitate the estimation of a high-dimensional posterior density for model parameters, we undertake parameter sensitivity tests using a variety of methods. First, parameters are ranked using first rank the importance of specific model parameters on model outputs via Global Sensitivity Analysis. Specifically we employ variance-based decomposition techniques to compute Sobol indices (Sobol (1993); Campolongo et al. (2000)). Posterior densities are estimated first for the most important parameters, while less important parameters are fixed at their nominal values. This constraint is subsequently relaxed to arrive at a joint posterior distribution over the entire parameter space. Second, since the GSA does not consider the error model when ranking parameters, we complement the GSA results with an analysis of the Fisher Information Matrix (FIM) (**?**) . The FIM results quantify the amount of information the experimental observations carry about the set of DALEC parameters for a particular setting for the discrepancy between model predictions and data. This study also allows an investigation of the information content of data based on steady state vs. non-steady state assumptions.

Finally, we undertake a Bayesian posterior predictive check (Lynch and Western (2004)) to assess the adequacy of the calibrated Carbon model to predict the experimental observations. The predictive skill of this model is further assessed via Continuous Rank Predictive Score (Gneiting and Raftery, 2007) computations. (Gneiting and Raftery (2007)) computations. The analysis steps mentioned here are undertaken with the help of the Uncertainty Quantification Toolkit (UQTk).[1] UQTk is a collection of software libraries and

[1] http://www.sandia.gov/UQToolkit

tools for the quantification of uncertainty in numerical model predictions. Additional scripts specific to this study are available upon request from the first author.

This paper is organized as follows. Section 2 provides a description of the processes comprising DALEC and of their associated parameters. Section 3 presents the GSA results, including first total order effects, in Sect.Section 3.1, and joint effects, in Sect.Section 3.2. FIM results and posterior distributions Posterior densities for model parameters are explored in Sect.Section 4 and the predictive capabilities are estimated in Sect.Section 5. We end with conclusions in Sect.Section 6. The methods employed in this paper are part of UQTk v3.0.[2]

2 Description of the Carbon Cycle Model

The schematic in Fig. 2 shows a 1 day time step consisting of a sequence of process-based submodels shown with green boxes. These submodels are connected via fluxes and interact with five major Carbon (C) pools. The fluxes calculated on any given day impact C Carbon pools and processes in subsequent days. The blue arrows in this figure indicate C Carbon pools or model variables that are input parameters to specific sub-models, while green arrows indicate the C Carbon pools or model variables affected by a particular sub process.

This The version of DALEC used in this study is modified from based on a modified version of the DALEC v1 used in Fox et al. (2009) . Both versions of the model consist model (Williams et al. (2005); Fox et al. (2009)). The model has been modified to facilitate comparisons with the Community Land Model (Thornton et al. (2007)), and with the Local Terrestrial Ecosystem Carbon Model (Ricciuto et al. (2011)).[2] It consists of three vegetation C Carbon pools, for leaf, stem, and root, and two soil C Carbon pools, for soil organic matter

[2]. UQTk v3.0 is currently undergoing formal review. In the meantime, the source code is available upon request from Bert Debusschere ()

[2]The source code for the modified DALEC version is available upon request from Daniel Ricciuto (ricciutodm@ornl.gov)

and litter. ~~The photosynthesis~~ Photosynthesis is driven by the Aggregate Canopy Model (ACM) ~~(Williams et al., 2005)~~ (Williams et al. (2005)), which itself is calibrated ~~to~~ againts the Soil-Plant-Atmosphere (SPA) model ~~(Williams et al., 1996) . The following modifications were made: An update was made~~ (Williams et al. (1996)). ACM was updated to employ a ~~-~~temperature-based deciduous phenology used in Ricciuto et al. (2011), driven by the six parameters shown in Fig. 2. Spring phenology is driven by a ~~-~~linear relationship to growing degree days, while senescence is driven by mean air temperature. To reduce model complexity, the plant labile pool was removed and stem carbon is used to support springtime leaf flush given the spring phenology and the maximum leaf area index parameter. Given the importance of maintenance respiration in other sensitivity analyses (Sargsyan et al. (2014)), this process was added along with parameters controlling the base rate and temperature sensitivity.

In this version of DALEC, ACM shares one parameter, the specific leaf area (*lma*), with the deciduous phenology and employs two additional parameters, leaf C:N ratio (*leafcn*) and Nitrogen use efficiency (*nue*). The autotrophic respiration model computes the growth and maintenance respiration components and is controlled by three parameters: the growth respiration fraction (*rg_frac*), and the base rate at 25 °C (*br_mr*) and temperature sensitivity for maintenance respiration (*q10_mr*), respectively. The allocation sub-model partitions ~~C~~ Carbon to several vegetation ~~C~~ Carbon pools. Leaf allocation is first determined by the phenology submodel, and the remaining available ~~C~~ Carbon is allocated to the root and stem pools depending on the fractional stem allocation parameter (*astem*). The "Litterfall" submodel redistributes the ~~C~~ Carbon content from vegetation pools to soil pools and is based on the turnover times for stem (*tstem*), root (*troot*), and leaves (*tleaf*). The sequence of sub-models concludes with the "Decomposition" which models the heterotrophic respiration component and the decomposition of litter into soil organic matter (SOM). This sub-model is driven by temperature sensitivity for heterotrophic respiration (*q10_hr*), the base turnover times for litter and SOM at 25 °C (*br_lit*,*br_som*), respectively, and by the decomposition rate (*dr*) from litter to SOM.

Model parameters and their nominal values are provided in Table 1. These parameters are grouped according to the sub-model that employs them. Except for leaf mass per unit area (*lma*) which impacts both the deciduous leaf phenology and ACM, all other parameters are employed in single submodels. The numerical ranges and nominal values for these parameters are ~~also~~ provided in the table~~. These ranges, corresponding to the Harvard Forest site (Urbanski et al., 2007) , are set to capture a broad range of reasonable values used in past studies (Fox et al., 2009; White et al., 2000) and will serve as a base for the GSA study presented in the next section~~, and are designed to reflect average values and broad uncertainties associated with the temperate deciduous forest plant functional type that includes Harvard Forest (Fox et al. (2009); White et al. (2000); Ricciuto et al. (2011)). In addition to the model parameters, several processes are driven by the observed air temperature, solar radiation, vapor pressure deficit, and CO_2 concentration at the flux tower site.

As mentioned in the Introduction, for this study we consider two approaches for running DALEC. The first approach employs a steady state assumption, with DALEC run in a spinup mode until it reaches a quasi-steady state. For this study we declare the model to be in a quasi-steady state when the relative L_2 error between successive cycles becomes less than a threshold value of 10^{-6} for select model ouputs. For the range of parameters employed in the runs presented here, the model spinup takes typically 30-50 cycles of the 1992-2006 meteorology (450-750 total years) depending on the parameter values, especially the turnover time of slow carbon pools. In this context, each cycle corresponds to running the model for 15 years with the meteorology inputs of 1992-2006. At the start of the first cycle, the Carbon pools are initialized to zero with the exception of stem carbon, which is set at a value to "seed" leaf growth in the following season. For subsequent cycles, the Carbon pools are initialized with the final state from the previous cycle. The daily quantities of interest output by DALEC in the first cycle after the system reaches a steady state are used for several analyses presented in this paper. This approach follows the protocol for the North American Carbon Program (NACP) interim synthesis simulations, but fails to capture, for example, the large negative NEE observed at Harvard Forest. In the second

approach, the initial values of the Carbon pools in January 1992 are added to the set of model parameters to be estimated. This approach employs transient assumptions and, for any given set of parameter values, DALEC is run one cycle only, for 1992-2006. The resulting model output values are then used to study the model behavior under transient conditions. The model evaluations are cheaper compared to the first approach, however the dimensionality of the parameter space of DALEC is increased by 5, with 3 vegetation Carbon pools and 2 soil Carbon pools, from 18 to 23 parameters. Henceforth, we will refer to these two approaches as D_{ST} and D_{TR}.

3 Global Sensitivity Analysis

GSA formally ~~connects uncertainties~~ studies how the change in model output ~~to the underlying uncertainties present~~ can be apportioned to changes in the model inputs. ~~We~~ Given our focus on statistical model calibration and uncertainty quantification, we employ variance-decomposition methods where the variance of the model output is decomposed into fractions associated with input factors and their interactions. The primary quantity of interest (QoI) for GSA is NEE, for which we have experimental observations available. We explore GSA for several other ~~QoI's~~ QoIs to understand the role each parameter or set of parameters ~~play on other DALEC outputs~~ plays in determining other quantities of interest in addition to NEE. Specifically we consider the Gross Primary Production (GPP), the Total Vegetation Carbon (TVC), and the Total Soil Carbon (TSC).

The effects of input parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{N_\theta}\}$ and their interactions on a model output $y = m(\boldsymbol{\theta})$, are quantified through Sobol indices (Sobol (1993); Campolongo et al. (2000)). The first order Sobol indices are given by

$$S_i = \frac{\mathrm{Var}_{\theta_i}[\mathbb{E}_{\boldsymbol{\theta}_{\sim i}}(m(\boldsymbol{\theta})|\theta_i)]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]}, \quad i = 1, \ldots, N_\theta \tag{1}$$

where $\boldsymbol{\theta}_{\sim i} = \{\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_{N_\theta}\}$, $\mathbb{E}_{\theta_{\sim i}}[\cdot]$ is the expectation with respect to $\theta_{\sim i}$, and $\mathrm{Var}_{\theta_i}[\cdot]$ is the variance with respect to θ_i. Note that, in this context, sub-script i can denote

one parameter or a group of parameters. Such a group, corresponding to the Phenology model, is presented below.

Similarly, the joint sensitivity indices S_{ij} are

$$S_{ij} = \frac{\mathrm{Var}_{\theta_i,\theta_j}[\mathrm{E}_{\boldsymbol{\theta}_{\sim(i,j)}}(m(\boldsymbol{\theta})|\theta_i,\theta_j)]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]} - S_i - S_j, \quad i,j = 1,\ldots,N_\theta. \tag{2}$$

While interactions between three or more parameters can be defined in a ~~-~~similar fashion, for most physical models these higher-order interactions are ~~typically~~ negligible.

The sensitivity index S_i can be interpreted as the fraction of the variance in the QoI that can be attributed to the i~~th~~-th input parameter only, while S_{ij} is the variance fraction that is due to the joint contribution of the i~~th~~-th and j~~th~~-th input parameters. The ~~Sobol indices~~total sensitivity index combines the first-order sensitvity indices with joint sensitivity and higher-order interactions to yield

$$S_i^T = S_i + \sum_{\substack{j \\ i\neq j}} S_{ij} + \sum_{\substack{j,k \\ i\neq j\neq k\neq i}} S_{ijk} + \ldots = \frac{\mathbb{E}_{\boldsymbol{\theta}_{\sim i}}[\mathrm{Var}_{\theta_i}(m(\boldsymbol{\theta})|\boldsymbol{\theta}_{\sim i})]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]} \tag{3}$$

This index measures the fractional contribution to the total variance due to parameter θ_i and all interactions with all other model parameters.

Starting from the derivation of these indices, based on the decomposition of variance, the sum of all first-order order indices and joint and higher-order interaction indices sums to one

$$1 = \sum_i S_i + \sum_{\substack{i,j \\ i\neq j}} S_{ij} + \ldots \tag{4}$$

Given that all Sobol indices are greater or equal to zero, it follows that $\sum_i S_i \leq 1$. The reverse is true for the total effect indices, $\sum_i S_i^T \geq 1$, due to multiple counting of joint and higher order parameter interactions.

Total effect indices are useful to ascertain which parameter or group of parameters has the most impact on a particular QoI, and also decide which parameters are less important and can potentially be fixed at their nominal value without a significant impact on the model output. Joint sensitivity indices can be used to verify or discover interactions between the computational model components as related to a specific model output. In this paper we will present results for total effect and joint sensitivity Sobol indices, while skipping first order Sobol indices for brevity.

The Sobol indices (1)~~and (2-(3)~~ can be written in integral forms, but these integrals ~~will not be~~ are not analytically tractable when the input parameter space is high-dimensional. In order to evaluate these indices numerically we employ a ~~-~~Monte-Carlo approach enhanced by techniques described by Saltelli (2002) and modified by Kucherenko et al. (2012) to account for parameter dependencies. This method employs sampling of the input parameters from their prior distributions and an efficient re-use of model evaluations to reduce the computational cost of estimation of the above conditional variances.

We employ ~~maximum entropy (MaxEnt) arguments to choose prior distributions for the model parameters, since we are only given prior information on parameter bounds (Table 1). The MaxEnt principle states that the maximum entropy distribution is the least informative distribution (**?**) . Among distributions with finite support, the uniform distribution has the largest entropy, hence, given available prior information, we choose uniform prior distributions~~ informative priors, described in Sec 4.3, for all model parameters~~, with bounds provided in Table 1~~. The prior distributions for ~~these parameters are~~ all parameters are assumed independent, except for the ~~spring phenology parameters *gdd_min* and *gdd_max*~~Spring phenology parameters *gdd_min* and *gdd_max*, which are bound by the inequality constraint $gdd\_min < gdd\_max$. Consequently, for these two parameters we will compute a compound sensitivity index, namely S_i^T for $i = (gdd\_min, gdd\_max)$ which is the total effect index based on joint prior distribution of this set of parameters, including all interactions between either *gdd_min* or *gdd_max*, or both, and the rest of the DALEC parameters.

For each of the QoIs mentioned above, we compute monthly averages corresponding to the entire simulation, ~~i.e.~~ *i.e.* the January average is computed using the January daily QoI values for all available years. ~~Global averages for all QoIs are also analyzed for comparison purposes.~~ The simulations are driven by daily minimum and maximum temperatures, global radiation, and CO_2 concentration for years ~~1992–2006~~1992 – 2006, at the Harvard Forest site (Urbanski et al. (2007)).

3.1 ~~First-order effects~~Total Effect Indices

Figures 3~~–~~6 show matrices of ~~first-order Sobol indices~~total effect indices, S_i^T, for the four QoIs mentioned above. ~~The colormap changes from red for large Sobol index values to blue for Sobol indices $\approx 1\%$. The grayscale corresponds to Sobol index values from 1% down to 0.1%, while blank cells indicate values smaller the 0.1%.~~ Each row in these matrices shows the ~~Sobol~~ indices corresponding to a ~~particular~~ particular monthly average QoI. ~~The sum of these values on each row indicates the sum of variance contributions due to individual parameters to the total variance of that particular average QoI. For example, in Fig. 3, the first order Sobol indices for the September average NEE sum up to 0.73. The remaining 0.27 fraction of the total variance for this month is due to pairwise interactions between parameters or higher order interactions.~~

Different parameters have larger impacts at certain times of the year. For NEE corresponding to D_{ST}, in Fig. 3a, phenology parameters *tsmin* and *leaffall*, which control the senescence of leaves in the Fall, have a ~~-~~significant impact on NEE during this period only. Specifically, *tsmin*, which is the critical temperature at which *leaffall* begins, mainly affects NEE in October. For D_{TR}, in Fig. 3b, the base rate of maintenance respiration *br_mr*, which represents a Carbon cost plants must continuously spend during their lifetime, becomes the dominant parameter for NEE. In the transient configuration, the autotrophic respiration sub-model controls most of NEE variance. The total effect index for several parameters, i.e. *astem*, *tstem*, *troot*, and *tleaf* are not shown in this figure, since they have a negligible contribution to NEE variance.

Similar behavior is seen for parameters that control GPP~~. Parameter *gdd_min*~~, in Fig. 4. Parameter *gdd_min*, which is ~~the~~ part of the pair *gdd*=(*gdd_min*,*gdd_max*) in this figure, is the number of growing degree days at which leaf budbreak occurs~~,~~. This parameter has the most impact in March and April. The strong dependence of these fluxes on phenology parameters highlights the importance of an accurate phenology model, as has been shown in other modeling studies, ~~e. g. (Richardson et al., 2012) .~~ *e.g.* (Richardson et al. (2012)). On the other hand, the Nitrogen use efficiency ~~*nue*~~*nue*, which controls the amount of GPP per unit leaf Nitrogen, is important throughout most of the growing season (~~June–September~~June-September). This is broadly consistent with other sensitivity studies that have shown strong sensitivity to leaf nitrogen, ~~e.g. Sargsyan et al. (2014) .~~ *e.g.* Sargsyan et al. (2014) . Unlike for NEE, the GPP fluxes exhibit a similar dependence on the parameters controlling the phenology and aggregate canopy modes for both D_{ST} and D_{TR}.

TVC and TSC are carbon pools and tend to vary on a ~~much~~ larger timescale than GPP or NEE, which are fluxes. Therefore, the Sobol indices do not exhibit significant seasonal variability. TVC is a sum of three Carbon pools, *vc1* (for leaf C), *vc2* (for stem C), and *vc3* (for root C). For both D_{ST} and D_{TR}, in Fig. 5, this quantity of interest is most strongly controlled by the base rate of maintenance respiration ~~*br_mr*, which represents a Carbon cost plants must continuously spend during their lifetime. TSC is most strongly controlled by~~ *br_mr*. For D_{TR}, the initial value of *vc2* exhibits a small, but non-negligible, total effect index of about 10% on the total variance of TVC.

TSC corresponding to D_{ST}, in Fig. 6a, is mostly controlled by both *br_mr* and the base rate of decomposition for soil organic matter ~~*br_som*~~*br_som*, which effectively determines the pool residence time. Given the same inputs, a ~~pool~~ with a ~~longer~~ residence time will contain more Carbon. For D_{TR}, in Fig. 6b, the initial value of soil organic matter pool (*sc2*) becomes dominant and exhibits a total effect index of about 50%. For this setup, the impact of *br_mr* and *br_som* on the total variance of TSC is about 40%, down from about 80% for the quasi-steady state setup for D_{ST}.

The total sensitivity index results indicate that, for some quatities of interest like GPP and TVC, the simulation setup, i.e. D_{ST} vs D_{TR}, does not change significantly the effect of model parameters on the model outputs. For these two model outputs the dominant parameters are similar for both setups, given the priors employed for the model parameters, including the Carbon pools for D_{TR}. Unlike for GPP and TVC, the simulation setup changes the relative importance of model parameters on NEE and TSC. This takes place either through a change in the relative importance of phenology and ACM model parameters (for NEE) or by bringing a significant contribution from the Carbon pools (for TSC). In the next section we examine joint effect indices for parameter pairs to determine what fraction of the total effect indices is due to interactions between model parameters.

3.2 Joint ~~effects~~Effects

Figures 7~~–8 -9~~ show relevant joint sensitivity indices corresponding to ~~NEE and GPP which exhibit seasonal variability for the first-order Sobol indices~~the four quantities of interest examined in this study. In these figures, each node shows relevant parameters while the label on each link corresponds to the joint Sobol index S_{ij}, in ~~% units. In this figure the joint~~ % units. She joint sensitivity Sobol index values are rounded to the nearest integer for clarity. ~~The results for these months, selected based on the relevant active processes affecting these two model outputs, show that~~

Both NEE and GPP exhibit seasonal variability for the total effect Sobol indices. For these parameters the joint parameter interactions are ~~also important . For example, during Spring, the interaction between *gdd_min* and other ACM~~ only relevant during Fall, accounting for about 10-15% of the total variance in the corresponding quantity of interest, and play an important role in determining the evolution of the Carbon cycle during the senescence period. Figures 7 and 8, showing these interactions during October, are representative of results throughout Fall. For both NEE and GPP the interaction *tsmin* and *leaffall* is signifincant during Fall, while interactions between other phenology, ACM, and AR parameters are negligible. In general joint sensitivity maps for NEE and GPP are similar between D_{ST} and ~~AR model parameters account for around 10–20of the total variance for~~

both NEE and GPP. Conversely, in the Fall, *tsmin* and *br_mr* have important interactions with several other phenology, AR, and Decomposition model parameters D_{TR}. These interactions account for about 15% of the total variance in both NEE and GPP and play an important role in determining the evolution of the Carbon cycle during the senescence period

Similar to the total effect index results for TVC and TSV, the joint sensitivity indices display little seasonal variability. The results shown in Fig. 9 for these QoIs correspond to September and are representative of all monthly averages (results not shown). Moreover, only D_{ST} results are shown in this figure since the corresponding D_{TR} results are almost identical to D_{ST}. For TVC the data in Fig. 9a indicates that the interaction between AR (through *br_mr*) and ACM (*nue*) and Literfall (*tstem*) sub-models, respectively, contribute about 10% to TVC variance. In fact these joint interactions represent about half of the total effect index of *nue* and *tstem*, shown in Fig. 5. The results in Fig. 9b show that the interactions between model parameters are important for TSC as well. For this quantity of interest, the interaction AR (*br_mr*) and Decomposition (*br_som*) sub-models acounts for about 10-30% of the corresponding total effect index values, shown in Fig. 6.

The GSA results can be used to understand the effect of model parameters on particular quantities of interest and discard, from the analysis, parameters that have a negligible impact. In this study, we will use the GSA results to facilitate the calibration of model parameters, by grouping parameters into sub-sets according to their effect on the relevant quantities of interest. More details are presented in the following section.

4 Parameter ~~calibration~~Calibration

We employ a Bayesian framework to compute posterior probabilities for ~~the~~ model parameters discussed in the previous sections. This framework is well-suited for dealing with uncertainties from different sources, including parametric and model uncertainty ~~and~~ as well as experimental errors ~~(Sivia, 1996) . In the Bayesian approach, the probability density~~

~~for the model parameters is characterized as~~(Sivia (1996)). Bayes rule is given as:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{L_{\mathcal{D}}(\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})} L_{\mathcal{D}}(\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathcal{D}), \tag{5}$$

~~Here~~ where $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\mathcal{D})$ are the prior and posterior probability densities, respectively, for model parameters $\boldsymbol{\theta}$. These densities represent our knowledge of $\boldsymbol{\theta}$ before and after learning from the data $\mathcal{D}$. The likelihood function $L_{\mathcal{D}}(\boldsymbol{\theta}) = p(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood of the data $\mathcal{D}$ for a ~~-~~particular instance of model parameters $\boldsymbol{\theta}$. The denominator in Eq. (5), $p(\mathcal{D})$, is the "evidence", computed by integrating the numerator over the support of $\boldsymbol{\theta}p(\boldsymbol{\theta})$. It plays a ~~-~~role of a ~~-~~normalizing constant in the parameter estimation context, and is not computed here.

4.1 Calibration Data

The data available for the calibration of model parameters consists of the Harvard ~~Forest's daily Net Ecosystem Exchange daily~~ Forest's daily NEE values processed for the North American Carbon Program Site Synthesis study ~~(Barr et al., 2013) . Flux data were measured by the site PI's (Urbanski et al., 2007)~~ (Barr et al. (2013)) based on flux data measured at the site (Urbanski et al. (2007)). Hill et al. (2012) estimated that daily NEE estimates follow a ~~-~~normal distribution. ~~Consistent with Hill et al. (2012) , systematic biases are not included in the present study. We further assume that daily measurement noise/errors, ϵ_d, are independent. Daily measurement errors or standard deviations are provided by the North American Carbon Program (NACP) interim synthesis (Barr et al., 2009) . For this study we neglect the model error, ϵ_g in Fig. 1. Given these assumptions, the likelihood $L_{\mathcal{D}}(\boldsymbol{\theta})$ is written as~~

$$L_{\mathcal{D}}(\boldsymbol{\theta}) = \prod_{k=1}^{N_d} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(y_k - \mathcal{D}_k)^2}{2\sigma_k^2}\right)$$

where $\mathcal{D}_k$ and σ_k are the observed NEE value and its standard deviation for day k, while $y_k = m_k(\boldsymbol{\theta})$ is the corresponding NEE value predicted by the DALEC model. The daily observations cover a period of 15 years starting with year 1992. A snapshot of these observations, including the magnitude of the observation error, is provided in Fig. 10. The standard deviations for the daily NEE values were estimated using a bootstrapping technique using half-hourly NEE data (Papale et al., 2006; Barr et al., 2009) .

For this study we consider two approaches for running the forward model and generating the output needed for the computation of the likelihood in Eq. (10(Papale et al. (2006); Barr et al. (2009)). The first approach employs a steady state assumption, with DALEC run in a spinup mode until a steady state is reached. This takes typically 30–50 cycles of the 1992–2006 meteorology (450–750 total years). In this context, each cycle corresponds to running the model for 15 years with the meteorology inputs of 1992–2006. At the start of the first cycle, the Carbon pools are empty. For subsequent cycles, the C pools are initialized with the final state from the previous cycle. The daily model-predicted NEE values used for parameter estimation are those of the first cycle after the system reaches a steady state. This approach follows the protocol for NACP interim synthesis simulations, but fails to capture the large negative NEE observed at Harvard Forest. In the second approach, the initial values of the C pools in January 1991 are added to the set of model parameters to be estimated. This approach employs unsteady assumptions and, for any given set of parameter values, DALEC is run one cycle only, for 1992–2006. The resulting model output values are employed to compute the likelihood. The model evaluations are cheaper compared to the first approach, however the dimensionality of the parameter space is increased by 5, 3 vegetation C pools and 2 soil Carbon pools, from 18 to 23 parameters. Henceforth, we will refer to these two approaches as D18 and D23. mean standard deviation is about 0.7, with a range of variation between 0.2 and 2.5.

4.2 Fisher information matrix

We first proceed to estimate the amount of information datasets consisting of NEE observations are expected to carry about the DALEC model parameters. This is quantified

via Fisher Information (**?**) , which is defined as the amount of information the observable NEE carries, as a random vector, about the unknown parameter vector θ. Let Z be a random vector of NEE observations. For this work, the probability density for Z is the multivariate normal likelihood defined by Eq. (10). In

4.2 Likelihood Construction

In general, the current context, the specific dataset of NEE observations, $\mathcal{D}$, is replaced by a random vector of NEE observations Z. The Fisher Information Matrix (FIM) is defined as (**?**)

$$\underline{\mathcal{I}}(\boldsymbol{\theta})_{i,j} = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L_Z(\boldsymbol{\theta})|\boldsymbol{\theta}\right] = -\int_{\Omega(Z)} \frac{\partial^2 \log L_Z(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} L_Z(\boldsymbol{\theta})\mathrm{d}Z$$

where $\Omega(Z)$ represents the space of all possible values of Z. Since $L_Z(\boldsymbol{\theta})$ is a multivariate normal, $Z \sim N(\boldsymbol{m}(\boldsymbol{\theta}), \boldsymbol{\Sigma})$, with a constant covariance matrix $\boldsymbol{\Sigma}$, the Fisher Information Matrix (FIM) entries in Eq. (**??**) can be shown to be

$$\underline{\mathcal{I}}(\boldsymbol{\theta})_{i,j} = \frac{\partial \boldsymbol{m}^T}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{m}}{\partial\theta_j}.$$

discrepancy between model predictions and the data can be formalized as

$$z = m(t;\boldsymbol{\theta}) + \epsilon_m + \epsilon_d \tag{6}$$

Here, ~~$\boldsymbol{m}(\boldsymbol{\theta})$~~ t is the time in day units and z is the daily NEE observation described above. Further, ϵ_m is the discrepancy between the model prediction $m(t;\boldsymbol{\theta})$ and the physical truth, while ϵ_d denotes the experimental error. In general it is not straightforward to disambiguate between these two sources of error. For the present study, we presume the experimental

error to be known (Papale et al. (2006); Barr et al. (2009)). Given that measurements are taken at different times, we further assume that daily measurement noise/errors, ϵ_d, are independent, hence

$$\epsilon_d \propto N(0, \Sigma_d), \quad \Sigma_d = \text{diag}[\underbrace{\sigma_{d,1}^2, \sigma_{d,2}^2, \ldots}_{N_d}], \tag{7}$$

where N_d is the number of days. Next we will focus our attention on modeling ϵ_m. We propose a multivariate Gaussian distribution, employing a constant bias $\boldsymbol{\mu} = [\mu, \mu, \ldots, \mu]^T$ and a $N_d \times N_d$ square exponential covariance matrix Σ_m with

$$\Sigma_{m\,i,j} = \sigma_m^2 \exp\left(-(t_i - t_j)^2 / l_c^2\right) \tag{8}$$

Given that t_i is simply a notation for day #i, the covariance matrix entries are given by $\Sigma_{m\,i,j} = \sigma_m^2 \exp\left(-(i-j)^2/l_c^2\right)$, where l_c is a correlation length. This analytical expression for Σ_m is adopted based on the intuition that model errors for succesive days are highly correlated while model errors for days that are far apart are uncorrelated. The magnitude of l_c controls the rate of decrease of daily model error correlations.

Given the above formulations of model and data errors, one can group these two into one multivariate normal error term

$$\epsilon = \epsilon_m + \epsilon_d \propto N(\boldsymbol{\mu}, \Sigma), \quad \Sigma_{i,i} = \sigma_m^2 + \sigma_{d,i}^2, \quad \Sigma_{i,i\pm k} = \sigma_m^2 \exp\left(-k^2/l_c^2\right) \tag{9}$$

and the likelihood $L_{\mathcal{D}}(\boldsymbol{\theta})$ is written as

$$L_{\mathcal{D}}(\boldsymbol{\theta}) = (2\pi)^{-N_d/2} \sqrt{|\Sigma|} \exp\left(-(\boldsymbol{z} - \boldsymbol{m} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{z} - \boldsymbol{m} - \boldsymbol{\mu})\right) \tag{10}$$

Here, $\boldsymbol{z} = [z_1, z_2, \ldots]$ is the vector of daily NEE values output by DALEC, $\boldsymbol{m}(\boldsymbol{\theta}) = (m_1(\boldsymbol{\theta}), \ldots, m_{N_d}(\boldsymbol{\theta}))^T$, with $m_k = \text{NEE}_k$ for day k. The covariance matrix $\boldsymbol{\Sigma}$

~~is diagonal, constructed with daily variance values, σ_k^2, on the diagonal, see also Eq. (10). In order to compare the FIM entries corresponding to different parameters , the parameter values are normalized by their corresponding prior range. The normalized FIM entries are then computed as $\mathcal{I}^*(\boldsymbol{\theta})_{i,j} = \mathcal{I}(\boldsymbol{\theta})_{i,j}\Delta_{\theta_i}\Delta_{\theta_j}$. Here Δ_{θ_i}~~ NEE observations, $\boldsymbol{m} = [m(t_1,\boldsymbol{\theta}), m(t_2,\boldsymbol{\theta}), \ldots]$ is the ~~range corresponding to θ_i, computed based on values given~~ vector of model NEE predictions, and $\boldsymbol{\mu}$ is the model bias vector described above. All these vectors are N_d long. In addition to the model parameters $\boldsymbol{\theta}$, we now have three additional hyperparameters characterizing the model error: the model bias μ, model error standard deviation σ_m, and correlation length l_c. Unlike for DALEC parameters, for which we employ informed priors described in the next section, for these hyperparameters we employ uninformed priors.

In practice, estimating the likelihood $L_{\mathcal{D}}(\boldsymbol{\theta})$ can be costly, and prone to numerical instabilities when considering the full $N_d \times N_d$ covariance matrix Σ. Therefore we will work with band-diagonal covariance matrices, obtained by setting the diagonals of the model error covariance matrix Σ_m to zero beyond a certain bandwidth k_b

$$\Sigma_{i,i\pm k} = 0 \text{ for } k > k_b \tag{11}$$

The effect of covariance matrix bandwidth on the model error terms $\{\mu, \sigma_m, l_c\}$ and DALEC parameters is studied in Section 4.4.1.

4.3 Parameter priors

Following LeBauer et al. (2012) we proceed to construct informed priors for the DALEC model parameters as well as for the initial Carbon pool amounts employed in D_{TR}. Considering the nominal values and bounds presented in Table 1, we separate model parameters into two categories. In the first category we place parameters with a range that spans approximately one order of magnitude or less. For these parameters we employ truncated normal densities as priors, with the mode set at the nominal values and standard deviations set to one-eight of the range of variation for each parameter.

The FIM values are expectations over the data, computed for specific values for model parameters $\boldsymbol{\theta}$. Thus, an uncertain $\boldsymbol{\theta}$ leads to uncertain FIM entries. We use Monte Carlo sampling to generate random samples from the MaxEnt-derived priors on $\boldsymbol{\theta}$. This yields an ensemble of FIM values, from which we can construct histograms for each FIM component. Since the model output dependence on $\boldsymbol{\theta}$ is not given analytically, we compute the partial derivatives in Eq. (**??**) by numerical differentiation[3]

$$\frac{\partial m_k}{\partial \theta_i} \approx \frac{\delta m_k}{\delta \theta_i} = \frac{m_k(\ldots, \theta_{i-1}, \theta_i + \delta\theta_i, \theta_{i+1}, \ldots) - m_k(\boldsymbol{\theta})}{\delta \theta_i}.$$

The perturbation $\delta\theta_i$ is set to $\sqrt{\epsilon_m}\theta_i$, where ϵ_m is the upper bound of relative error due to rounding in floating-point arithmetic, and is typically of the order of 2.2×10^{-16} for double precision (64 bit) computations.

Normalized histograms for select diagonal entries of the FIM are shown in Figs. **??** and **??**, using the model setup for D18. In these figures, FIM_{θ_i} stands for $\mathcal{I}(\boldsymbol{\theta})^*_{i,i}$, defined above. Convergence tests, results not shown, indicate that about 10^4 Monte Carlo samples are sufficient to generate converged histograms. The results were grouped, by visual inspection, according to the magnitude of $\log(\text{FIM}_{\theta_i})$. Figure **??** shows model parameters with larger FIM diagonal entries while Fig. **??** shows parameters corresponding to generally smaller values. This indicates that NEE observations are informative about the In the second category we place parameters for which the range of variation spans more than two orders of magnitude. For these parameters we set truncated log-normal density priors. Similarly to the first set of parameters shown in Fig. **??** (see the figure caption for the list of parameter names), and that, consequently, the probability distributions for these parameters are likely to be significantly updated through the Bayesian parameter estimation discussed in this section. It is interesting to note that these parameters were found to be important, , the parameters of these densities are set such that the mode occurs at the nominal

[3] Alternatively, a set of ordinary differential equations for $\partial m_k/\partial\theta_i$ can be derived and solved numerically.

value and the standard deviation is set to one-eight of the range of variation for each corresponding parameter in this category. For both normal and log-normal densities, we truncate the priors based on the ~~GSA results in the previous section, for NEE. Specifically, *gdd_min*~~ranges presented in Table 1.

For all parameters, except the pair (*gdd_min*,~~*q10_mr*, *br_mr*, and *rg_frac* exhibit relevant first order effects, shown in Fig. 3, while *gdd_max*~~ *gdd_max*) we consider independent prior distributions. For the growing degree days parameters, given the inequality constraint $gdd\_min < gdd\_max$, we employ a truncated joint normal density set up as a product of one-dimensional normal densities for both *gdd_min* and ~~*leaffall* are relevant mostly through interactions with other parameters, in Fig. 7.~~

~~Conversely,~~ *gdd_max*. This joint density is appropriately scaled so that it integrates to 1 over non-rectangular space (due to the inequality constraint) for these two parameters. Similarly, the ~~FIM values shown inFig. **??** are much smaller, hence the calibration exercise is not expected to update the probability distributions for these parameters significantly. Among these parameters, *br_som* was found to be the most important for the TSC. However, based on the FIM results, the NEE observations do not carry information about this parameter and its prior density will likely not be updated by using the NEE data only. This likely results from the long residence time of the SOM pool – the 15 year NEE record is not long enough to constrain it sufficiently. These conjectures will be verified using the calibration results presented in the next section.~~ truncated normal and log-normal densities for the other model parameters are appropriately scaled to account for the finite parameter ranges.

~~Based on the FIM results presented in this section and on the GSA results in Sect. 3, we separate the DALEC parameters into three groups. In the first group we include *gdd_min*~~For D_{TR}, the initial Carbon pool amounts (representing values on January 1st, ~~*gdd_max*, *q10_mr*, *br_mr*, *rg_frac*, and *leaffall*. These parameters were highligted both by the GSA and FIM results as being important for NEE. In the second group we include *tsmin*, *q10_hr*, *br_lit*, and *lma*. These parameters were selected since either the GSA or the FIM results suggested they are relevant to NEE. Finally, we place the remaining parameters~~

in the third group. In the next section the posterior distributions for model parameters are constructed sequentially starting with the most important group of parameters, then gradually adding parameters, one group at a time1992) are also estimated in addition to the DALEC parameters and the hyperparameters defining the model error. For the carbon pool initial values we also employ truncated normal and log-normal densities. These prior distributions are informed by site observations (Table 2). The initial leaf carbon (vc1) is set to zero with a small standard deviation because of the starting date of the simulation, which is in mid-winter well after leaf fall. Initial litter and soil organic mean (sc1, sc2) values and standard deviations are taken from Gaudinski et al. (2000) , while stem carbon is estimated from Urbanski et al. (2007) . Specifically, we employ truncated normal densities for all Carbon pools except litter carbon (sc1). For sc1, the mean and the range differ by two orders of magnitude, hence we employ a truncated log-normal density for this pool.

4.4 Posterior distributions via MCMC

A —Markov Chain Monte Carlo (MCMC) algorithm is used to sample from the posterior probability density $p(\boldsymbol{\theta}|\mathcal{D})$ in Eq. (5). MCMC is a —class of techniques that allows sampling from a —probability density by constructing a —Markov Chain that has the target density as its stationary distribution (Gamerman, 1997; Gilks et al., 1996) (Gamerman (1997); Gilks et al. (1996)). In particular, we employ an adaptive Metropolis algorithm (Haario et al., 2001) (Haario et al. (2001)), which uses the covariance of the previously visited chain states to find better proposal distributions, allowing it to explore the posterior distribution in an efficient manner. Haario et al. (2001) shows show that, for Gaussian distributions, the adaptive sampling algorithm is similar in performance to the Metropolis algorithm. For non-Gaussian posterior densities, the adaptive procedure is superior to non-adaptive procedures, however the adaptive procedure is challenged by the dimensionality of the parameter space.

To facilitate the convergence of the adaptive MCMC algorithm we proceed gradually, starting with the first a group of parameters mentioned in the previous sectionidentified as

important for NEE through GSA in Section 3. The schematic in Fig. 11 shows one iteration in the sequence of MCMC simulations. ~~For the first iteration,~~ We also add the model error hyperparameters, in addition to select DALEC parameters, to start the first iteration

$$\theta^{(1)} = \{gdd\_min, gdd\_max, ~~q10\_mr~~tsmin, ~~br\_mr~~leaffall, ~~rg\_frac~~nue, ~~leaffall~~q10\_mr, br\_mr\}$$

with initial values $~~\theta_{ini}^{(1)}~~ \theta_{ini}^{(1)}$ set to the nominal conditions provided in Table 1 for DALEC parameters, and $\mu = 0$, $\sigma_m = l_c = 1$ for model error hyperparameters, respectively. The rest of parameters are held constant at their nominal values. The initial covariance matrix, $~~\mathbf{C}_{ini}^{(1)}~~ C_{ini}^{(1)}$, allows the MCMC algorithm to explore a ~~-~~number of possible states before adapting the sample covariance based on the sample history. For this study we found that a ~~-~~diagonal covariance matrix with entries set to a ~~-~~fraction of about $1/16$ of the variances for the corresponding prior ~~distributions provided a~~ density provided a good start for the MCMC algorithm.

The MCMC states obtained during the first iteration are used to compute the covariance matrix corresponding to the first set of parameters $~~\mathbf{C}^{(1)}~~ C^{(1)}$ which is then used to construct the initial covariance matrix for the second iteration, $~~\mathbf{C}_{ini}^{(2)}~~ C_{ini}^{(2)}$. This process is shown schematically in Fig. 11. The initial parameter values for the ~~2nd~~ 2-nd iteration consist of the Maximum A ~~-~~Posteriori (MAP) for $\theta^{(1)}$ augmented with the nominal values for

$$\theta^{(2)\backslash(1)} = \{~~tsmin~~lma, ~~q10\_hr~~rg\_frac, ~~br\_lit~~q10\_hr, ~~lma~~br\_lit\}$$

The iterative process is completed after the third iteration, with $\theta^{(3)\backslash(2)}$ containing the rest of the DALEC parameters. This iterative algorithm breaks the original high-dimensional problem into a sequence of steps of increasing dimensionality, with each intermediate step starting with a better proposal covariance compared to an approach for which this covariance is empirically chosen.

We employ the ~~Raftery–Lewis diagnostic (Raftery and Lewis, 1992)~~ Raftery-Lewis diagnostic (Raftery and Lewis (1992)) to determine when the MCMC samples converge

to stationary posterior distributions. For ~~D18~~D_{ST}, approximately 4×10^6 samples are necessary to predict the ~~5, 50, and 95 %~~ 5%, 50%, and 95% quantiles of all parameters to within ~~±1 % accuracy with 95 %~~ ±1% accuracy with 95% probability. For ~~D23~~D_{TR}, the Raftery-Lewis diagnostic test shows that 6×10^6 are necessary for converged posterior ~~distributions~~densities. Given 5×10^6 MCMC samples, the Effective Sample Size ~~(Kass et al., 1998) (~~(Kass et al. (1998)) (ESS) for ~~D18 varies between 10 000 and 15 000~~ D_{ST} varies between 10,000 and 15,000 samples depending on each parameter, while for ~~D23~~D_{TR}, ESS is between ~~8000 and 12 000~~8,000 and 12,000. This shows that there is significant autocorrelation between chain samples, which is somewhat typical for MCMC samplers in high-dimensional spaces. To ensure converged posterior ~~distributions~~densities, and since the computational model is cheap, results presented below are based on ~~10^7~~ 7.5×10^6 MCMC samples for both ~~D18 and D23.~~ D_{ST} and D_{TR}. When processing the MCMC samples, we skip the first 10^6 samples, and then "thin" the rest of the samples by picking every 10th sample.

4.4.1 Effect of covariance bandwidth on posterior distributions

We performed several MCMC runs to examine the effect of covariance bandwidth on the estimates of model pararameters and hyperparameters. The bandwidth is parameterized by k_b, in Eq. (11), which denotes the number of non-zero diagonals on either side of the main diagonal.

Figure 12(a)-(c) shows the estimated MAP values for the hyperparameters μ, σ_m, and l_c, respectively corresponding to the model error. In addition to D_{ST} and D_{TR}, we also show results for "D_{TR}^{up}". This run is similar to D_{TR}, except uniform priors with the same range were employed for all Carbon pools. The error bars shown in this figure correspond to two standard deviations estimated from the MCMC samples.

It seems that the model bias μ, in Fig. 12(a), is not significantly affected by the band-diagonal trim of the covariance martrix. For all runs considered here μ is consistently negative signaling that, on average, DALEC overpredicts the NEE data. The other two model parameters, σ_m and l_c, in Figs. 12(b,c), are sensitive to the bandwidth setup,

until they reach statistically converged values around $k_b = 10$. The model error standard deviation σ_m mean values for both D_{ST} and D_{TR} are slightly below 0.4, compared to a mean value of 0.7 for the NEE measurement error (in Section 4.1).

The 2D joint marginal density for σ_m and l_c, shown in Fig. 12(d) for $k_b = 12$, indicate a relatively strong dependence and a negative correlation between these two hyperparameters. Results for larger covariance bandwidths (not shown) confirm that densities of both σ_m and l_c exhibit converged moments for $k_b > 10$.

It is interesting to note the value for the converged mean correlation length l_c. It seems that this hyperparameter does not depend on a particular model setup, at least for the site and time range considered in this study. Further tests, with uniform priors for all parameters lead to similar mean values for l_c. A value of l_c = 4, indicating that the model error discrepancy exhibits a time scale of about 8 days, seems to be an intrinsic property of the model. This most likely suggests that model errors follow the variability of NEE over synoptic timescales associated with the periodic passage of weather systems and precipitation events (Mahecha et al. (2010)). Further tests, with alternate model error terms, are necessary to verify this observation.

4.4.2 Comparison between D_{ST} and D_{TR}

We first proceed to analyse the model calibration results for ~~D18~~D_{ST}, when DALEC is run to a ~~-~~quasi-steady state for each parameter sample. In order to measure the degree of dependence in the ~~joint posterior distribution~~ posterior distributions for the 18 model parameters we examine the "distance correlation" values ~~(Székely et al., 2007)~~ (Székely et al. (2007)) estimated based on the MCMC samples. The distance correlation is a ~~-~~measure of *dependence* between two random variables, being zero when they are independent. Given random variables X and Y with finite first moments, the distance correlation $\mathcal{R}(X,Y) \in [0,1]$ is defined as

$$\mathcal{R}(X,Y) = \frac{\vartheta^2(X,Y)}{\sqrt{\vartheta^2(X)\vartheta^2(Y)}} \tag{12}$$

where $\vartheta^2(X,Y)$ is the "distance covariance" between X and Y and $\vartheta^2(X)$ is the "distance variance", $\vartheta^2(X) = \vartheta^2(X,X)$. The distance covariance $\vartheta^2(X,Y)$ is defined as

$$\vartheta^2(X,Y) = E(||X - X'_{-}||||Y - Y''_{-}||)) + E(||X - X'_{-}||)E(||Y - Y''_{-}||) - 2E(||X - X''_{-}||||Y - Y'''_{\sim}||)) \quad (13)$$

where ~~(X',Y'), (X'',Y'')~~ (X',Y'), (X'',Y'') are independent and identically distributed random variables, ~~drawn from~~ with the same joint density as (X,Y). Székely et al. (2007) provide numerical algorithms to compute $\mathcal{R}(X,Y)$ given samples of random variables X and Y. The results are shown in Table 3. In this table, parameters are grouped in blocks according to the sub-model they participate in. The entries in the diagonal blocks show dependencies between parameters in the same sub-model while the entries in off-diagonal blocks indicate dependencies between parameters from different sub-models.

The most important statistical dependencies are between ~~*nue* and *lma*~~ *nue* and *lma* that control the gross photosynthesis (ACM) and between ~~*rg_frac* and *nue*~~ *rg_frac* and *nue* that control net photosynthesis. Relevant dependencies are also observed between ~~*q10_mr*, a~~ *q10_mr*, a parameter of the autotrophic respiration process, and ~~*q10_hr* which participates in the heterotrophic respiration process~~and the gross photosynthesis parameters. In order to further understand the dependencies between model parameters we compute ~~1-D and 2-D~~ 1D and 2D joint marginal densities, via Kernel Density ~~Estimates~~ Estimation (KDE) ~~(Scott, 1992; Silverman, 1986)~~ (Scott (1992); Silverman (1986)), for the model parameters that exhibit ~~distance correlation factors greater than 0.3~~at least one distance correlation factor that is greater than 0.4. These results are shown in Fig. 13. The statistical dependencies identified above through $\mathcal{R}$ are also evident in ~~2-D~~ 2D joint marginal densities for the same parameters.

Figure 14 shows ~~1-D~~ 1D marginal densities for the rest of the parameters. These parameters show little dependence on other parameters and so the ~~1-D~~ 1D marginal distribution is sufficient to characterize their density. Some parameters~~are well constrained towards the center of the prior range, for instance *br_mr* and *br_lit*, these parameters control the~~

basal autotrophic and litter respiration rates, which occur on timescales for which the NEE data have high information content.For *tleaf* and *br_som*, the NEE observations are not informative, see Sect. **??**, and their posterior densitiesremain nearly uniform, the same as their prior densities. For *tleaf*, e.g *astem*, *tleaf* , *br_mr*, show little update from prior to posterior densities. For *br_som*, its turnover rate is slow enough such that the NEE data contain little useful information. For *tleaf*, the lack of information is due to the fact that the effects of leaf turnover on net fluxes are much more strongly controlled by their timing, as determined by the phenology parameters, than by the background turnover rate. For *br_som*, its turnover rate is slow enough such that the NEE data contain little useful information. The posterior densities for other parameters, e.g. *laimax*e.g. *laimax*, are tilted toward one end of their prior range. This might indicate that the model error term is not sufficient to describe the discrepancy between the model and the data, and the calibration process attempts to compensate for structural discrepancies between observations and model predictions by pushing some parameters toward either the minimum or the maximum value of their prior range.

While the posterior distribution for *br_mr* is well updated in the calibration, the $\hat{\mathcal{R}}$ values between this parameter and all other parameters are smaller than the threshold used to select parameters that show mutual dependencies. The posterior distribution for *tsmin* is piecewise uniformThe posterior density for *tsmin* exhibits an interesting piecewise quasi-linear profile. This is due to the fact that minimum daily temperatures, in degrees Celsius, are provided with one decimal digit accuracy and this parameter is a threshold for leaf drop, i.e. its participation in the computational model is through an "if" statement. Hence all samples between successive one-digit accurate thresholds are equally probable likely during the MCMC sampling process., and the product between piecewise uniform likelihood and the normal prior results in the posterior density profile observed in Fig. 14.

Nextwe , we analyze the calibration results for D_{TR}. For this model setup, the initial values for the Carbon pools at the beginning of year 1992 are part of the set of model parameters and each DALEC simulation consists of only one cycle, for the time span 1992-2006. The distance correlation matrix for D_{TR} parameters that are common to D_{ST} has entries that are

by-and-large similar to the ones shown in Table 3 indicating that the dependence between model parameters is not altered by the model setup. This observation is confirmed by visual inspection of the 1D and 2D joint marginal densities based on D_{TR} results for the same parameters as the ones shown in Fig. 13 (results not shown).

Finally, Figure 15 shows marginal densities for two Carbon pools that were updated in the calibration exercise D_{TR}. *vc3* corrresponds to the stem Carbon while *sc1* and *sc2* correspond to the litter Carbon and soil organic matter, respectively. Both *vc3* and *sc2* exhibit some dependence on the temperature sensitivities for maintenance respiration and heterotrophic respiration, *q10_mr* and *q10_hr*, respectively. These dependencies are consistent with the flow of information depicted in Fig. 1.

Next we examine the departure of each parameter's posterior density from its ~~uniform prior as a~~ prior density as a result of the Bayesian update via Eq. (5). We quantify these changes via the ~~Kullback–Leibler~~ Kullback-Leibler (KL) divergence between prior and marginal posterior densities,

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx, \tag{14}$$

where p is the posterior ~~density~~ density and q is the prior density. KL divergence results ~~for certain paramaters~~ are presented in Fig. 16. In this figure, parameters are sorted in ascending order based on ~~their D_{KL} values. Also shown in the figure is the inverse of the scaled standard deviation based on the MCMC sample values for each model parameter. This is obtained from the standard deviation of the MCMC samples for each parameter, σ_i, by scaling with the standard deviation based on the corresponding prior density, $\sigma_i^* = \sigma_{p,i}/\sigma_{q,i}$. The calibration exercise had negligible effect on the probability density for the first two parameters, *tleaf*~~ the D_{KL} values for D_{ST}. Parameters that exhibit $D_{KL} < 0.5$ for both D_{ST} and ~~*br_som*~~D_{TR} are excluded from this figure for clarity. Moreover, the C pools shown in this figure are only present for D_{TR}, hence there is no D_{ST} result for these parameters. ~~$D_{KL}(p||q)$ values for these two parameterswere less than 10^{-2} and were rounded to zero.~~

The corresponding σ^* values are close to 1, indicating little change from their prior uniform densities. This result confirms that NEE data contain little information on the turnover rate of SOM, or on the rapidity of leaf drop (rather than the timing of leaf drop, see below). At the other end of the spectrum, *br_mr*, *gdd_min*, and *q10_mr*, exhibit the largest $D_{KL}(p||q)$, indicating a larger departure of their posterior densities from their prior densityThe right half of this figure contains parameters that were identified as important for NEE in Section 3. These parameters are well constrained by the NEE data, reflecting the useful information in the flux dataon , for example on the timing of phenological events (*gdd_mingdd_min*) and the dynamics of autotrophic respiration *(br_mr, q10_mr*). The large $D_{KL}(p||q)$ values for these parameters are accompanied by small σ^* values, or large values for $1/\sigma^*$ as shown in Fig. 16, indicators that the marginal posterior density is significantly narrower than their prior density. Back-of-the-envelope regression tests empirically suggest a power law dependence between $D_{KL}(p||q)$ and σ^*.

To further assess the importance of statistical dependencies between model parameters, we revisit the GSA exercise to determine the relative importance of model parameters, based on the posterior densities of model parameters instead of the prior ones. Figure **??** shows select first order Sobol indices given posterior distributions based on D18. In this figure, several parameters are grouped together if the distance correlation values, shown in Table 3, imply that their mutual dependence is significant. These groups, named G1, G2, and G3, respectively, consist of the following parameters: G1: lma, nue, rg_frac G2: q10_mr, q10_hr G3: gdd. In general D_{KL} results are similar for D_{ST} and D_{TR}, perhaps with the exception of *br*

msin, gddom. For mD_{ST}, the NEE data contain little information on the turnover rate of SOM. For D_{TR}, the inclusion of Carbon pools, in particular the SOM pool (*sc2*ax The results shown with red bars in Fig. **??** are based on joint posterior distributions for G1, G2 and G3, and marginal distributions for the rest of parameters. For the results shown with blue bars, the joint distributions for G1 through G3 are products of marginal distributions of each parameter in the group, hence neglecting any statistical dependence. Visual inspection of the relative importance of parameters or groups of parameters to the total variance of the

~~average monthly NEE values shows that neglecting joint dependecies between parameters can significantly alter the results. This is true both for parameters that show significant dependence, e.g. see group G1, and for parameters that show little dependence on other parameters, e.g. *br_mr.*~~

~~Next, we analyze the calibration results for D23. For this model setup, the initial values for the C poolsat the beginning of year 1991 are part of the set of model parameters and each DALEC simulation consists of only one cycle, for the time span 1992–2006. Table **??** shows $\mathcal{R}$ values for D23 results. Only the parameters common between D23 and D18 are listed in this table. The $\mathcal{R}$ values observed in this table for D23 are similar to the ones observed above for D18 with the exception of pair (*q10_mr*, *q10_hr*). While, for D18, these two parameters exhibit significant dependence, for D23 they are nearly independent. Figure **??** shows 1-D and 2-D),~~ impacts the Bayesian update of this parameters due to the dependencies observed in the joint marginal densities~~for parameters with distance correlation factors greater than 0.3 based on D23 results. In general, these marginal densities are similar to the ones based on D18.~~

~~Finally,~~ , shown in Fig. 15~~shows marginal densities for two Carbon pools that were updated in the calibration exercise D23. *vpool2* corrresponds to the stem C while *spool2* corresponds to the soil organic matter. While *dr* shows little dependence on other model parameters, it has a dominant role in the conversion of the litter C pool into soil organic matter, and the distance correlation between this parameter and *spool2* is about 0.5~~.

5 Predictive ~~assessment~~Assessment

In this section we explore the predictive skill given the posterior distributions for the model parameters for ~~D18 and D23~~D_{ST} and D_{TR}. First, we employ the Bayesian posterior predictive distribution ~~(Lynch and Western, 2004)~~ (Lynch and Western (2004)) to assess the adequacy of the calibrated DALEC model, and the Gaussian data noise model, for prediction of the NEE observations. Specifically, the posterior distribution for the ~~*predicted*~~ *predicted* NEE data, $p(y|\mathcal{D})$, is computed by marginalization of the likelihood over the posterior distri-

bution of model parameters and hyperparameters, here θ:

$$p\left(\underline{y}\boldsymbol{y}|\mathcal{D}\right) = \int_{\theta} p(\underline{y}\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})\underline{\mathrm{d}}\underline{d}\theta. \tag{15}$$

For the present work, ~~$y|\boldsymbol{\theta} \sim N(\boldsymbol{m}(\boldsymbol{\theta}), \boldsymbol{\Sigma})$, where $\boldsymbol{y} = \{y_k | k = 1 \dots N_d\}$ is a~~ $(\boldsymbol{y} - \boldsymbol{m})|\boldsymbol{\theta} \sim \epsilon_m + \epsilon_d$, where $y = [y_1, y_2, \dots]$ is a N_d-dimensional vector with NEE predictions over a ~~-~~range of N_d days, ~~and Σ is a diagonal covariance matrix with variances of daily NEE observations on the diagonal. The 1-D-marginal posterior predictive distributions~~ ϵ_m is the model error term and ϵ_d is the data noise term.

The 1D-marginal posterior predictive density for daily NEE values for a ~~-~~two-year snapshot around 1995 are shown in Fig. 17. These ~~distributions~~ densities were computed by sampling ~~,~~ the posterior distribution of model parameters $\boldsymbol{\theta}$, i.e. by using the MCMC samples ~~behind the posterior densities presented in the previous section. For computational efficiency, $p(y|\boldsymbol{\theta})$ was computed with KDE for each θ sample . This density was then interpolated on a uniform grid centered around the daily observed NEE value~~that are already available. We employ about 4000 MCMC samples, for each sample we draw 20 samples from the multivariate normal distribution $\epsilon_m + \epsilon_d$, and then add these samples to the model evaluations. These results are saved into daily bins, from which we extract several quantiles corresponding to the 1D-marginal posterior predictive density. It is worth to note that the variance of the posterior predictive distributions can also be estimated analytically as the sum of the measurement error variance and the pushed-forward variance, i.e. the variance of the output quantity of interest with respect to posterior variability.

The top frame in Fig. 17 corresponds to ~~D18~~ $\mathrm{D_{ST}}$ and the bottom frame to ~~D23~~$\mathrm{D_{TR}}$. Generally, the predicted data spread covers well the observed NEE values ~~except for a time rangearound May-June-July when the observations, with red line, are frequently outside the 5–95~~for the entire time range. Ocasional spikes can be seen outside the $5-95\%$ predictive band, shown in blue. ~~This discrepancy occurs mostly for years 1993–1996. For other years, the predicted data covers the May-June-July observational data well.~~

In order to quantitatively compare the predictive capability of the calibrated models for ~~D18 and D23~~D_{ST} and D_{TR}, we adopt a ~~-~~probabilistic score based on the predictive cumulative distribution function (CDF). The Continuous Rank Predictive Score (CRPS) ~~(Gneiting and Raftery, 2007)~~ (Gneiting and Raftery (2007)) measures the difference between the CDF of the provided data and that of the forecast/predicted data, ~~i.e.~~ *i.e.* data generated based on the posterior predictive distribution. Thus,

$$\text{CRPS}\mathit{CRPS}(\mathcal{F},\mathcal{D}) = \frac{1}{N_d}\sum_{k=1}^{N_d}\int_{-\infty}^{\infty}\left(\mathcal{F}_k(y_k|\mathcal{D}) - \mathcal{H}_{\mathcal{D}_k}(y_k)\right)^2 \mathrm{d}y\mathrm{d}y_k \tag{16}$$

Here, $\mathcal{F}_k(y_k|\mathcal{D})$ is the ~~1-D~~ 1D marginal posterior predictive CDF for day/component k computed using ~~1-D~~ 1D marginal posterior predictive distributions

$$\mathcal{F}_k(y_k|\mathcal{D}) = \int_{-\infty}^{y_k} p_k\left(y''_k|\mathcal{D}\right)\mathrm{d}y''_k \tag{17}$$

where

$$p_k\left(y_k|\mathcal{D}\right) = \int p\left(y_1, y_2\ldots, y_{N_d}\boldsymbol{y}|\mathcal{D}\right)\mathrm{d}y_1\cdots\mathrm{d}y_{k-1}\mathrm{d}y_{k+1}\cdots\mathrm{d}y_{N_d}\boldsymbol{y}_{\sim k}. \tag{18}$$

is the ~~1-D~~ 1D marginal posterior predictive ~~distribution~~ density corresponding to day k, based on $p(\boldsymbol{y}|\mathcal{D})$ computed via Eq. (15). Herre, $\mathrm{d}\boldsymbol{y}_{\sim k} = \mathrm{d}y_1\cdots\mathrm{d}y_{k-1}\mathrm{d}y_{k+1}\cdots\mathrm{d}y_{N_d}$. The CDF of the provided data is approximated as a ~~-~~Heaviside function centered at the observation value $\mathcal{D}_k$ ~~(Hersbach, 2000)~~ (Hersbach (2000)), $\mathcal{H}_{\mathcal{D}_k}(y_k) = \mathbb{1}_{y_k \geq \mathcal{D}_k}$.

~~Table 4 displays CRPS values based on posterior distributions obtained by averaging over several time ranges. The first row shows the values corresponding to a 60day time frame, from mid-June to mid-August, while the second row corresponds to the remainder of the year. The last row shows the aggregated values, considering the entire year. The averages in Eq. (16) are taken over all years considered in this study. The~~ We employ the

posterior predictive check data presented above to compute CRPS values for both D_{ST} and D_{TR}. For D_{ST} we obtain a value **0.67** while for D_{TR} the CRPS value is **0.60**. The lower values for ~~D23 compared to D18 indicatea better predictive skill~~ D_{TR} compared to D_{ST} indicate, on average, tighter marginal predictive CDF's that are better centered around the NEE data for the setup when DALEC is run for one cycle and the ~~C~~ Carbon pools are treated as parameters. ~~The results reveal that the largest improvement, about 6 %, occurs for the June–August time frame, while for the rest of the year the improvement is about 2 % only~~This indicates a better predictive skill for D_{TR} compared to D_{ST}.

In order to measure the effect of calibration on the predictive capability of DALEC we employ the Continuous Rank Predictive Skill Score (CRPSS) ~~(Wilks, 2011)~~ (Wilks (2011))

$$\text{CRPSS} CRPSS = \frac{\text{CRPS}_{\text{psp}} - \text{CRPS}_{\text{prp}}}{\text{CRPS}_{\text{prf}} - \text{CRPS}_{\text{prp}}} \frac{CRPS_{psp} - CRPS_{prp}}{CRPS_{prf} - CRPS_{prp}} \tag{19}$$

where ~~CRPS_{psp}~~ $CRPS_{psp}$ is the CRPS computed above based on the posterior predictive distribution, ~~CRPS_{prp}~~ $CRPS_{prp}$ is based on the prior predictive distribution, and ~~CRPS_{prf}~~ $CRPS_{prf}$ is the CRPS based on "perfect" predictions. For the current study, the "perfect" predictions ~~have a multivariate normal distribution~~ correspond to the hypothetical case with no model error and posterior densities for model parameters centered on the ~~observations and diagonal covariance matrix Σ defined above~~NEE observations. The prior predictive distribution is defined analogous to the posterior predictive distribution in Eq. (15), with the posterior density $p(\theta|\mathcal{D})$ being replaced by $p(\theta)$, the prior density for model parameters θ.

A ~~–~~CRPSS value of 0 implies no improvement of the predictive skill for the calibrated model parameters compared to the predictions based on the prior information, while a ~~–~~value of 1 can be achieved when the posterior distribution reduces to a ~~–~~point and the model prediction is the same as the corresponding experimental data. ~~For the current study, $\text{CRPS}_{\text{prp}} = 2.38$~~ The CRPS values corresponding to the prior ($CRPS_{prp}$), posterior $CRPS_{prp}$, and the ideal case $CRPS_{prf}$ are presented in Table 4 for both D_{ST} and ~~$\text{CRPS}_{\text{prf}} = 0.53$ for D18. This leads to CRPSS $= 0.4$, indicating a 40 % improvement in the predictive skill of DALEC as a result of calibration. We only show here the CRPSS values for~~

~~D18 since for D23, the C pools employed improper priors for which the $CRPS_{prp}$ is not well defined~~D_{TR}. Based on the values in this table the CRPSS for D_{TR} shows a much stronger improvent in predictive capabilities for this model setup compared to D_{ST}.

6 Conclusions

We presented uncertainty quantification results for a ~~-~~process-based ecosystem Carbon model. We assembled several probabilistic methodologies in a ~~-~~framework that tackles the connected problems of parameter estimation and forward propagation of input uncertainties. Depending on the simulation setup, the model employs either steady state or ~~non-steady~~ transient assumptions, respectively, and it is driven by meteorological data corresponding to years ~~1992–2006~~ 1992-2006 at the Harvard Forest site. Daily Net Ecosystem Exchange (NEE) observations were available to calibrate the model parameters and test the performance of the model.

We first discussed global sensitivity analysis (GSA) results for the complete set of input parameters. Based on their contribution to the variance, we find that different parameters have larger impacts for NEE at certain times of the year when the processes they control become important. One example is the ~~*tsmin*~~ *tsmin* parameter, which is the critical temperature at which leaf fall begins, and mainly affects NEE in October. We ~~also~~ found that parameter interactions can also be relevant to the variability of NEE or Gross Primary Production (GPP). Unlike NEE and GPP which are fluxes, the Carbon pools, either vegetation ~~or soil~~(TVC) or soil (TSC), tend to vary more slowly and their month-to-month variability depends on a ~~-~~small subset of parameters.

~~We also employed Fisher Information Matrix (FIM) computations to estimate the relative information the NEE data contains on the model parameters. To our knowledge this type of study is employed for the first time in the context of a Carbon model. We ranked model parameters according to the relative magnitude of the diagonal entries in the FIM and generally found that most "informed" parameters are also ranked as important based on the GSA results~~We also found that the simulation setup affects the relative importance of

parameters for NEE and TSC while GPP and TVC are less sensitive to the change between steady and transient assumptions.

We then proceeded to calibrate the model parameters in a ~~Bayesian framework~~ Bayesian framework using informative priors for all parameters. In this context we examined both steady and ~~unsteady~~ transient assumptions for the Carbon model simulations. ~~The daily discrepancies between measured~~ In the latter approach the initial values for the Carbon pools are part of the calibration process. The discrepancy between actual and predicted NEE values ~~were modeled as independent and identically distributed Gaussians with prescribed daily variance according to the recorded instrument error. All model parameters were assumed to have uninformative priors with bounds set according to expert opinion~~was modeled as a multivariate normal distribution with constant mean and a square exponential covariance matrix. A convergence study was performed to determine the effect of covariance matrix bandwidth on the parameters of the discrepancy term. It was found that the converged correlation length does not depend on the simulation setup and that the model discrepancy for NEE data exhibits a time scale of about one week.

The posterior distribution of model parameters was sampled sequentially by first considering the most relevant parameters and then progressively adding less important parameters, according to ~~GSA and FIM results~~GSA-based ranking. The posterior samples, obtained with a ~~-~~Markov Chain Monte Carlo algorithm, exhibit significant dependencies between some of the model parameters. ~~Further, a GSA analysis based on marginal posterior distributions shows the importance of considering parameter dependencies when establishing the importance of each parameter or set of parameters for given quantities of interest~~Comparison of posterior densities for parameters that are common to the two model setups indicate similar calibration results.

The predictive capabilities of the model, employing the parameters' posterior distribution, were assessed qualitatively through posterior predictive checks and quantitatively through Continuous Rank Predictive Score (CRPS) computations. Based on the CRPS values, the ~~unsteady~~ transient model setup, for which ~~C~~ Carbon pools are set as simulation parameters, performed ~~slightly~~ better, in particular ~~during the growing seasons, compared to~~

model setup assuming steady state conditionswhen compared to results based on prior predictive distributions. Given similar calibration results for the parameters common to the two configurations, we attribute the improvement in the predictive capabilities to the calibrated Carbon pools in the transient model setup.

The analysis presented in this paper considered a single data series at one site only. However, the Bayesian framework employed in the parameter calibrations is well-suited to deal with both heterogenous data and modelsheterogeneous data and multiple model setups. We are currently exploring avenues to extend this work to multi-site studies together with employing multiple data streams to better constrain the model parameters.

The framework presented here encompasses robust statistical methodologies that can be employed in the development and analysis of more detailed models like the Community Land Model (CLM). Since some of these methodologies are sampling-based, their application is restricted to computationally inexpensive models. To this end we are currently working on developing efficient surrogate models that can be used in place of expensive models like CLM. With a surrogate model approach in place, one can proceed to study individual CLM sub-models as well as the CLM model as a whole and potentially improve its predictive capabilities.

Acknowledgements. C. Safta, D. Ricciuto, K. Sargsyan, H. N. Najm, B. Debusschere, and P. Thornton were supported by the US Department of Energy, Office of Science, under the project "Climate Science for a Sustainable Energy Future", funded by the This research was conducted by the Accelerated Climate Modeling for Energy (ACME) project, supported by the Office of Biological and Environmental Research (BER) program . M. Williams in the DOE Office of Science. CS acknowledges partial support from the Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. DOE, Office of Science, Advanced Scientific Computing Research. MW was supported by NERC National Centre for Earth Observation. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94-AL85000. Oak Ridge National Laboratory is managed by UT-BATTELLE for DOE under contract DE-AC05-00OR22725.

References

Barr, A., Hollinger, D., and Richardson, A. D.: ~~CO_2~~ CO2 Flux Measurement Uncertainty Estimates for NACP, AGU Fall Meeting, 2009.

Barr, A., Ricciuto, D. M., Schaefer, K., Richardson, A., Agarwal, D., Thornton, P. E., Davis, K., Jackson, B., Cook, R. B., Hollinger, D. Y., van Ingen, C., Amiro, B., ans M. A. Arain, A.~~A. M.~~ A., Baldocchi, D., Black, T. A., Bolstad, P., Curtis, P., Desai, A., Dragoni, D., Flanagan, L., Gu, L., Katul, G., Law, B. E., Lafleur, P., Margolis, H., Matamala, R., Meyers, T., McCaughey, H., Monson, R., Munger, J. W., Oechel, W., Oren, R., Roulet, N., Torn, M., and Verma, S.: NACP Site: Tower Meteorology, Flux Observations with Uncertainty, and Ancillary Data, ~~available at:(last access: 10 October 2014),~~ Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, doi:10.3334/ORNLDAAC/1178, 2013.

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

Campolongo, F., Saltelli, A., Sørensen, T., and Tarantola, S.: Hitchhiker's ~~guide to sensitivity analysis~~Guide to Sensitivity Analysis, in: Sensitivity Analysis, edited by ~~:~~ Saltelli, A., Chan, K., and Scott, E., Wiley, Chicester, 2000.
~~Fisher, R.: Statistical Methods and Scientific Inference, Macmillan Pub Co, 1973.~~

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D. M., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Wijk, M. ~~-~~T. V.: The REFLEX project: ~~comparing~~ Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, ~~Agr. Forest . Meteorol.~~Agricultural and Forest Meteorology, 149, 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.

Gamerman, D.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Chapman & Hall, London, 1997.

Gaudinski, J. B., Trumbore, S. E., Davidson, E. A., and Zheng, S.: Soil carbon cycling in a temperate forest: radiocarbon-based estimates of residence times, sequestration rates and partitioning of fluxes, Biogeochemistry, 51, 33–69, doi:10.1023/A:1006301010014, 2000.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1996.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc.Journal of the American Statistical Association, 102, 359–378, doi:10.1198/016214506000001437, 2007.

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems., Weather Forecast. Wea. Forecasting, 15, 559–570, 2000.

Hill, T. C., Ryan, E., and Williams, M.: The use of CO_2 CO2 flux time series for parameter and carbon stock estimation in carbon cycle research, Glob. Change Biol.Global Change Biology, 18, 179–193, doi:10.1111/j.1365-2486.2011.02511.x, 2012.

Jaynes, E. T.: Prior probabilities, IEEE Transactions on Systems Science and Cybernetics, 4, 227–241, , 1968.

Kaminski, T., Knorr, W., Rayner, P. J., and Heimann, M.: Assimilating atmospheric data into a terrestrial biosphere model: a A case study of the seasonal cycle, Global . Biogeochem. Cy.Biogeochemical Cycles, 16, 14-1–14-1614–1–14–16, doi:10.1029/2001GB001463, 2002.

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Consistent assimilation of MERIS FAPAR and atmospheric COCO$_2$ into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences, 9, 3173–3184, doi:10.5194/bg-9-3173-2012, 2012.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M.: Markov Chain Monte Carlo in practice: a roundtable discussion, Am. Stat.Practice: A Roundtable Discussion, The American Statistician, 52, 93–100, doi:10.1080/00031305.1998.10480547, 1998.

Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, Glob. Change Biol.Global Change Biology, 11, 1333–1351, doi:10.1111/j.1365-2486.2005.00977.x, 2005.

Kucherenko, S., Tarantola, S., and Annoni, P.: Estimation of global sensitivity indices for models with dependent variables, Comput. Phys. Commun.Computer Physics Communications, 183, 937–946, doi:10.1016/j.cpc.2011.12.020, 2012.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, Biogeosciences, 9, 3757–3776, doi:10.5194/bg-9-3757-2012, 2012.

Lehmann, E.and Casella, G.: Theory of Point Estimation, Springer Texts in Statistics, Springer, 2003.

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., and Dietze, M. C.: Facilitating feedbacks between field measurements and ecosystem models, Ecological Monographs, 83, 133–154, doi:10.1890/12-0137.1, 2012.

Lynch, S. M. and Western, B.: Bayesian posterior predictive checks for complex models, ~~Sociol. Method. Res.~~Sociological Methods and Research, 32, 301–335, doi:10.1177/0049124103257303, 2004.

~~Papale,~~

Mahecha, M. D., Reichstein, ~~M.~~M., Jung, M., Seneviratne, S. I., Zaehle, S., Beer, C., Braakhekke, M. C., Carvalhais, N., Lange, H., Le Maire, G., and Moors, E.: Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales, Journal of Geophysical Research: Biogeosciences, 115, ~~Aubinet,~~ doi:10.1029/2009JG001016, 2010.

Papale, D., Reichstein, M., ~~Canfora,~~ Aubinet, M., Canfora, E., Bernhofer, ~~C.~~C., Kutsch, ~~W.~~W., Longdoz, ~~B.~~B., Rambal, ~~S.~~S., Valentini, ~~R.~~R., Vesala, ~~T.~~T., and Yakir, ~~D.~~D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, doi:10.5194/bg-3-571-2006, 2006.

Raftery, A. E. and Lewis, S.: How ~~many iterations~~ Many Iterations in the Gibbs ~~sampler~~Sampler?, in: Bayesian Statistics 4, pp. 763–773, Oxford University Press, ~~763–773,~~ 1992.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 19, ~~GB2026,~~ doi:10.1029/2004GB002254, 2005.

Ricciuto, D. M., Davis, K. J., and Keller, K.: A ~~Bayesian~~ Bayesian calibration of a simple carbon cycle model: ~~the~~ The role of observations in estimating and reducing uncertainty, Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 22, doi:10.1029/2006GB002908, 2008.

Ricciuto, D. M., King, A. W., Dragoni, D., and Post, W. M.: Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: ~~effects~~ Effects of constraining variables and data record length, ~~J. Geophys. Res.-Biogeo.~~Journal of Geophysical Research: Biogeosciences, 116, ~~G01033,~~ doi:10.1029/2010JG001400, 2011.

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., and Xue, Y.: Terrestrial biosphere models need better representation of

vegetation phenology: results from the North American Carbon Program Site Synthesis, Global Change Biology, ~~Glob. Change Biol.,~~ 18, 566–584, doi:10.1111/j.1365-2486.2011.02562.x, 2012.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, ~~Comput. Phys. Commun.~~Computer Physics Communications, 145, 280–297, doi:10.1016/S0010-4655(02)00280-1, 2002.

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B., Ricciuto, D. M., and Thornton, P. E.: Dimensionality reduction for complex models via Bayesian compressive sensing, International Journal of Uncertainty Quantification, 4, 63–93, doi:10.1615/Int.J.UncertaintyQuantification.2013006821, 2014.

Scott, D. W.: Multivariate Density Estimation. Theory, Practice and Visualization, Wiley, New York, 1992.

Silverman, B. W.: Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

Sivia, D. S.: Data Analysis: A Bayesian Tutorial, Oxford Science, 1996.

Sobol, I. M.: Sensitivity ~~estimates for nonlinear mathematical models~~Estimates for Nonlinear Mathematical Models, Math. Modeling and Comput. Exper., 1, 407–414, 1993.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K.: Measuring and ~~testing dependence by correlation of distances, Ann. Stat.~~Testing Dependence by Correlation of Distances, Annals of Statistics, 35, 2769–2794, doi:10.1214/009053607000000505, 2007.

Tang, J. and Zhuang, Q.: A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based ~~terrestrial ecosystem model, J. Geophys. Res.-Atmos.~~Terrestrial Ecosystem Model, Journal of Geophysical Research: Atmospheres, 114, ~~D15303,~~ doi:10.1029/2009JD011724, 2009.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to ~~CO_2~~CO2 fertilization and climate variability, Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 21, ~~GB4018,~~ doi:10.1029/2006GB002868, 2007.

Tjiputra, J. F., Polzin, D., and Winguth, A. M. E.: Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: ~~sensitivity~~ Sensitivity analysis and ecosystem parameter optimization, Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 21, ~~GB1001,~~ doi:10.1029/2006GB002745, 2007.

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S.: OptIC project: ~~an~~ An intercomparison of optimization techniques for parameter estimation in

terrestrial biogeochemical models, ~~J. Geophys. Res. Biogeo.~~Journal of Geophysical Research: Biogeosciences, 112, ~~G02027,~~ doi:10.1029/2006JG000367, 2007.

Urbanski, S., Barford, C., Wofsy, S. C., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling ~~CO_2~~CO2 exchange on timescales from hourly to decadal at Harvard Forest, ~~J. Geophys. Res. Biogeo.~~Journal of Geophysical Research: Biogeosciences, 112, 1–25, doi:10.1029/2006JG000293, 2007.

White, M. A., Thornton, P. E., Running, S. W., and Nemani, R. R.: Parameterization and ~~sensitivity analysis of the BIOME–BGC terrestrial ecosystem model: net primary production controls, Earth Interac.~~Sensitivity Analysis of the BIOME–BGC Terrestrial Ecosystem Model: Net Primary Production Controls, Earth Interactions, 4, 1–85, doi:10.1175/1087-3562(2000)004<0003:PASAOT>2.0.CO;2, 2000.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 2011.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a ~~Quercus–Acer~~ Quercus–Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, Plant~~Cell Environ.~~, Cell & Environment, 19, 911–927, doi:10.1111/j.1365-3040.1996.tb00456.x, 1996.

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

Xu, T., White, L., Hui, D., and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: ~~analysis~~ Analysis of uncertainty in parameter estimation and model prediction, Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 20, ~~GB2007,~~ doi:10.1029/2005GB002468, 2006.

Ziehn, T., Scholze, M., and Knorr, W.: On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models, Global ~~Biogeochem. Cy.~~Biogeochemical Cycles, 26, ~~GB3025,~~ doi:10.1029/2011GB004185, 2012.

Zobitz, J. M., Desai, A. R., Moore, D. J. P., and Chadwick, M. A.: A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC), Oecologia, 167, 599–611, doi:10.1007/s00442-011-2107-9, 2011.

Table 1. Description of model parameters.

	Param.	Nom. val.	Range	Description	Units
Phen.	gdd_min	100	$10 \cdots 250$	threshold for leafout	[°C day]
Phen.	gdd_max	200	$50 \cdots 500$	threshold for max. LAI	[°C day]
Decid.	tsmin	5	$0 \cdots 10$	Temperature for leaffall	[°C]
Decid.	laimax	4	$2 \cdots 7$	Seasonal max. LAI	[m^2 leaf / m^2]
Decid.	leaffall	0.1	$0.03 \cdots 0.95$	rate of leaffall	[day^{-1}]
ACM	lma	80	$20 \cdots 150$	specific leaf area	[gC / m^2 leaf]2ex>2ex
ACM	leafcn	25	~~fixed~~ fixed	leaf C:N ratio	[gC/gN]
A.R.	nue	7	$1 \cdots 20$	Nitrogen use efficiency	[ ]2ex>2ex
A.R.	q10_mr	2	$1 \cdots 4$	Maintenance resp. T-sensitivity	[]
A.R.	br_mr	10^{-4}	$10^{-5} \cdots 10^{-2}$	Base rate for maintenance resp.	[gC m^{-2} day^{-1} / gC biomass]
A.R.	rg_frac	0.2	$0.05 \cdots 0.5$	growth respiration fraction	[]
A.	astem	0.7	$0.1 \cdots 0.95$	Allocation to plant stem pool	[]
Litter.	tstem	$\frac{1}{50 \times 365}$	$\frac{1}{250 \times 365} \cdots \frac{1}{10 \times 365}$	stem turnover time	[day^{-1}]
Litter.	troot	$\frac{1}{5 \times 365}$	$\frac{1}{25 \times 365} \cdots \frac{1}{365}$	root turnover time	[day^{-1}]
Litter.	tleaf	10^{-2}	$10^{-3} \cdots 10^{-1}$	leaf turnover time	[day^{-1}] 2ex>2ex
Decomp.	q10_hr	2	$1 \cdots 4$	Heterotrophic resp. T-sensitivity	[]
Decomp.	br_lit	$\frac{1}{2 \times 365}$	$\frac{1}{5 \times 365} \cdots \frac{10}{5 \times 365}$	base turnover for litter	[gC m^{-2} day^{-1} / gC litter]
Decomp.	br_som	$\frac{1}{30 \times 365}$	$\frac{1}{100 \times 365} \cdots \frac{1}{10 \times 365}$	base turnover for SOM	[gC m^{-2} day^{-1} / gC SOM]
Decomp.	dr	10^{-3}	$10^{-4} \cdots 10^{-2}$	decomposition rate	[day^{-1}]

Table 2. ~~Distance correlation factors~~ Prior setup for ~~D18. The diagonal blocks are marked according to~~ the ~~process the parameters contribute to, see also Fig. 2 and Table 1. The entries~~ initial Carbon pool amounts employed in ~~the diagonal block show dependencies between parameters from the same process, while the entries in the off-diagonal block show dependencies between parameters from different processes~~D_{TR}.

Name	ID	Mean	St. Dev	Constraints
leaf C	*vc1*	0	20	$0 < vc1$
stem C	*vc2*	9000	1800	$0 < vc2$
root C	*vc3*	1500	300	$0 < vc3$
litter C	*sc1*	10	25	$0 < sc1 < 1000$
som C	*sc2*	8800	1760	$0 < sc2$

Table 3. Distance correlation factors for ~~D23~~D_{ST}. The diagonal blocks are marked according to the process the parameters contribute to, see also Fig. 2 and Table 1. The entries in the diagonal block show dependencies between parameters from the same process, while the entries in the off-diagonal block show dependencies between parameters from different processes.

	gdd_min	gdd_max	tsmin	laimax	leaffall	lma	nue	q10_mr	br_mr	rg_frac	astem	tstem	troot	tleaf	q10_hr	br_lit	br_som	dr
gdd_min	1																	
gdd_max	0.5	1																
tsmin	0	0	1															
laimax	0	0.1	0	1														
leaffall	0.1	0	0.2	0	1													
lma	0	0.2	0	0.1	0.1	1												
nue	0.1	0.3	0	0.	0.2	0.9	1											
q10_mr	0	0.2	0	0	0.1	0.6	0.6	1										
br_mr	0	0.2	0.1	0.1	0.4	0.1	0	0.2	1									
rg_frac	0.3	0.1	0	0	0.2	0.4	0.6	0.2	0.1	1								
astem	0	0	0	0	0	0	0	0	0	0	1							
tstem	0	0	0	0	0.2	0	0	0.1	0.1	0.1	0	1						
troot	0	0	0	0	0	0	0	0	0	0	0	0.1	1					
tleaf	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
q10_hr	0.1	0.1	0.1	0	0.2	0.2	0.1	0.2	0.2	0.2	0	0.2	0	0	1			
br_lit	0.1	0.1	0.1	0	0.4	0	0.1	0.1	0.5	0.1	0	0	0	0	0.3	1		
br_som	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
dr	0	0	0	0	0.1	0.1	0.1	0.1	0.2	0.1	0	0	0	0	0.1	0.2	0.1	1

Table 4. CRPS and CRPSS values for ~~D18~~ D_{ST} and ~~D23~~D_{TR}. The CRPSS value for D_{TR} shows a much larger improvement in predictive capabilities for this model setup compared to D_{ST}.

~~Period~~ Setup	~~CRPS-D18~~ $CRPS_{prf}$	~~CRPS-D23~~ $CRPS_{prp}$	~~change~~ $CRPS_{psp}$	CRPSS
~~Jun–Aug~~ D_{ST}	~~1.26~~ 0.16	~~1.18~~ 0.90	~~6 Rest of year~~ 0.67	~~1.27 1.25 2~~ 0.31
~~Overall~~ D_{TR}	~~1.27~~ 0.16	~~1.24~~ 1.45	~~2.5~~ 0.60	0.65

Table 5. Nomenclature.

ACM	Aggregate Canopy Model
CRPS	Continuous Rank Predictive Score
CRPSS	Continuous Rank Predictive Skill Score
DALEC	Data Assimilation Linked Ecosystem Carbon
FIM	Fisher Information Matrix
GPP	Gross Primary Production
GSA	Global Sensitivity Analysis
MCMC	Markov Chain Monte Carlo
NEE	Net Ecosystem Exchange
QoI	Quantity of Interest
TSC	Total Soil Carbon
TVC	Total Vegetation Carbon
$D_{KL}(p\|\|q)$	Kullback–Leibler divergence between probability densities q and p
$L_{\mathcal{D}} = p(\mathcal{D}\|\boldsymbol{\theta})$	Likelihood of the data $\mathcal{D}$ for a particular instance of model parameters $\boldsymbol{\theta}$
$p(\boldsymbol{\theta})$, $p(\boldsymbol{\theta}\|\mathcal{D})$	prior and posterior probability densities, respectively, for model parameters $\boldsymbol{\theta}$
$p(y\|\mathcal{D})$	posterior distribution for the predicted NEE data y
$p_k(y_k\|\mathcal{D})$	marginal posterior distribution for the predicted NEE component y_k
$\mathcal{R}(X,Y)$	Distance correlation between random variables X and Y
S_i	First-order Sobol index for parameter i
S_{ij}	Joint Sobol index for parameters i and j
S_i^T	Total-order Sobol index for parameter i
$\boldsymbol{\theta}$	Vector of parameters for DALEC

~~Schematic of parameter estimation, on yellow background, and forward UQ workflows, on green background. For this work DALEC is used as both "measurement model", g, and as "computational model", m. In the Bayesian framework the parameters estimation depends both on the model error ϵ_g and on the measurement error ϵ_d.~~

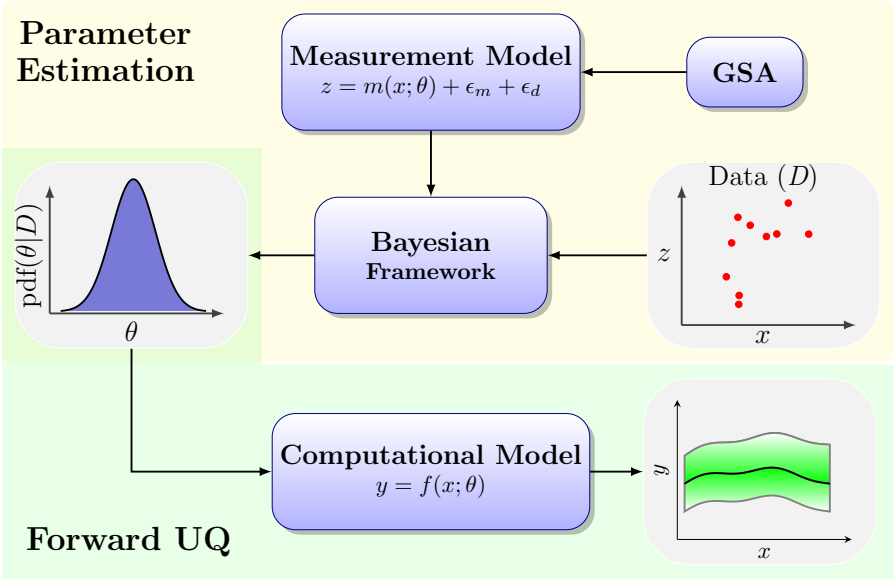


Figure 1. Schematic of parameter estimation, on yellow background, and forward UQ workflows, on green background. For this work DALEC is used as both "measurement model", m, and as "computational model", f. In the Bayesian framework, parameter estimation depends both on the model error ϵ_m and on the measurement error ϵ_d.

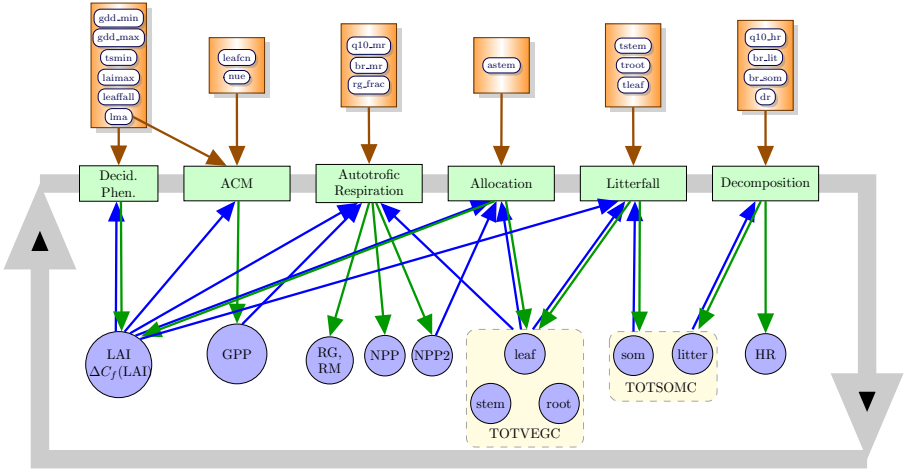


Figure 2. Schematic of processes, shown with green boxes, in DALEC with associated parameters, listed in orange boxes. The blue arrows indicate how internal parameters and QoIs, shown with blue circles, impact DALEC processes, while ~~while~~ the green arrows show the impact of processes on the QoI and other internal parameters.

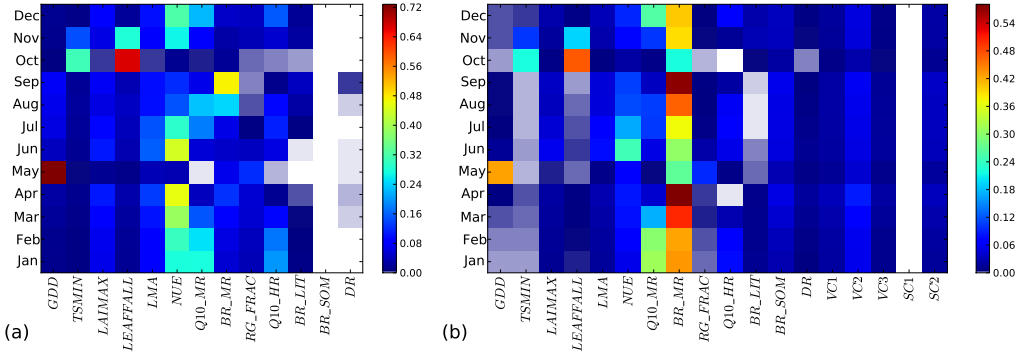


Figure 3. Matrices with ~~first-order~~ total effect Sobol indices, S_i^T, for monthly averages of NEE ~~. Also shown are the main Sobol indices~~ for ~~the global average~~ (~~G~~a) D_{ST} and (b) D_{TR}. The ~~largest value $S_{br\_mr} = 0.49$ occurs~~ colormap changes from red for ~~September average NEE~~large index values to blue for indices $\approx 1\%$. The ~~sum of first-order~~ grayscale corresponds to Sobol ~~indices for each month is shown in parentheses~~index values from 1% down to 0.1%, while blank cells indicate values smaller the 0.1%.

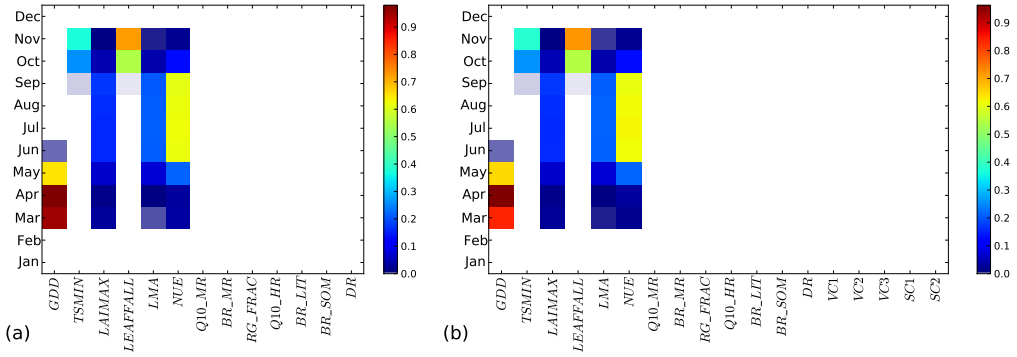


Figure 4. Matrices with ~~main~~ total effect Sobol indices, S_i^T, for monthly averages of GPP ~~. Also shown are the main Sobol indices~~ for ~~the global average~~ (~~G~~a) D_{ST} and (b) D_{TR}. The ~~largest value $S_{\text{leaffall}} = 0.77$ occurs for December average GPP. The sum of first-order Sobol indices for each month~~ colormap setup is ~~shown~~ similar to the one in ~~parentheses~~Fig. 3.

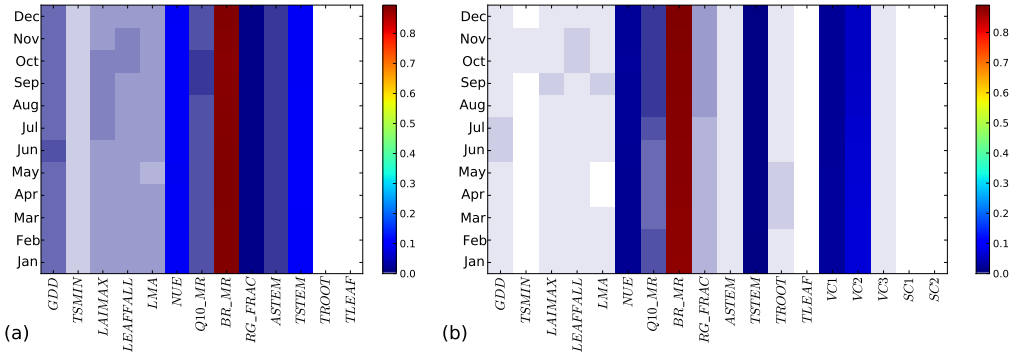


Figure 5. Matrices with ~~main~~ total effect Sobol indices, S_i^T, for monthly averages of TVC ~~. Also shown are the main Sobol indices~~ for ~~the global average~~ (~~G~~a) D_{ST} and (b) D_{TR}. The ~~largest value $S_{br\_mr} = 0.36$ occurs~~ colormap setup is similar to the one in Fig. 3.

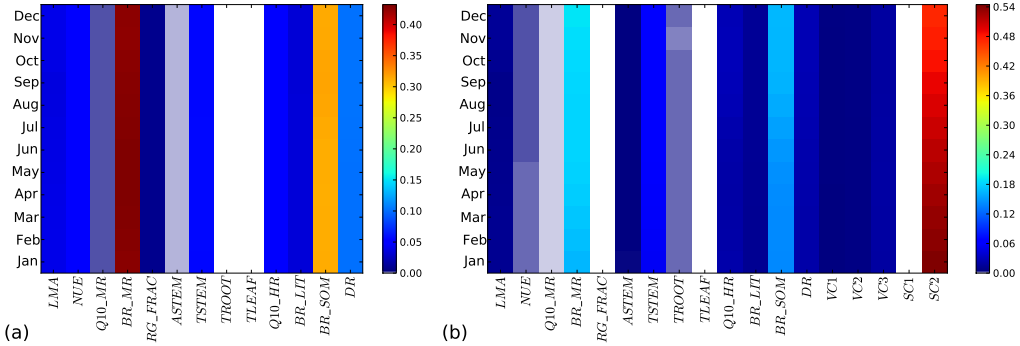


Figure 6. Matrices with total effect Sobol indices, S_i^T, for ~~several TVC~~ monthly averages ~~. The sum of first-order Sobol indices~~ TSC for ~~each month~~ (a) D_{ST} and (b) D_{TR}. The colormap setup is ~~shown~~ similar to results in ~~parentheses~~Fig. 3.

~~Matrices with main Sobol indices for monthly averages of TSC. Also shown are the main Sobol indices for the global average (G). The largest value $S_{\text{br\_som}} = 0.38$ occurs for several TSC monthly averages. The sum of first-order Sobol indices for each month is shown in parentheses.~~

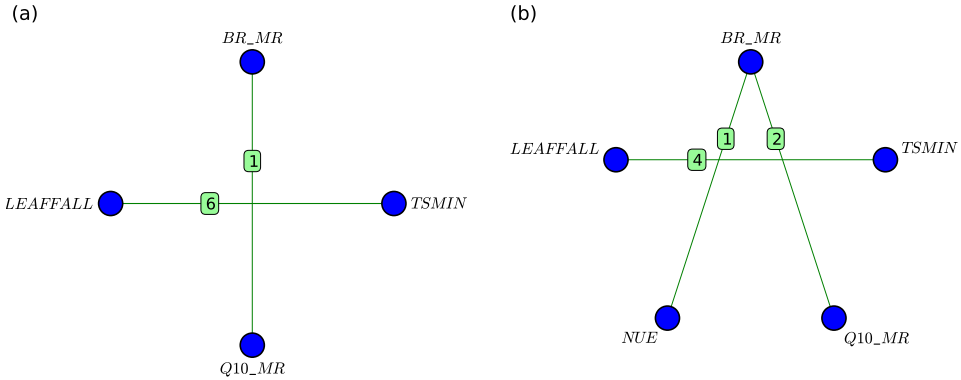


Figure 7. Relevant joint Sobol indices, S_{ij}, corresponding to ~~monthly~~ October NEE averages for ~~**(a)** May~~ (a) D_{ST} and ~~**(b)** October~~(b) D_{TR}. The labels on each line ~~shows~~ show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.

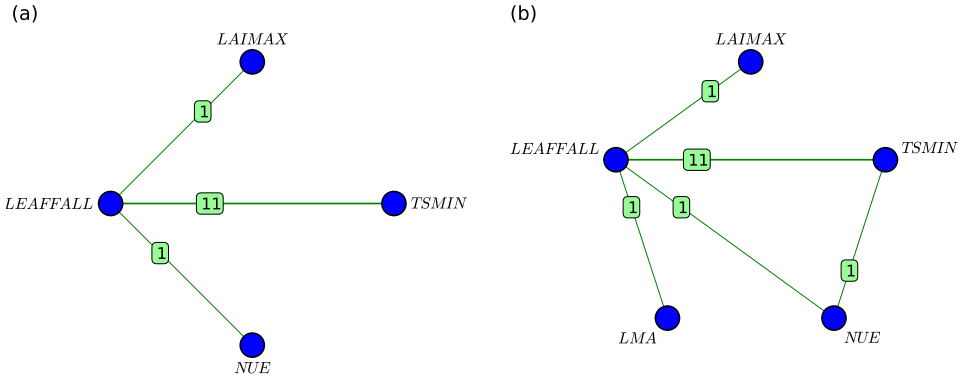


Figure 8. Relevant joint Sobol indices, S_{ij}, corresponding to ~~monthly~~ October GPP averages for ~~**(a)** May~~ (a) D_{ST} and ~~**(b)** November~~(b) D_{TR}. The labels on each line ~~shows~~ show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.

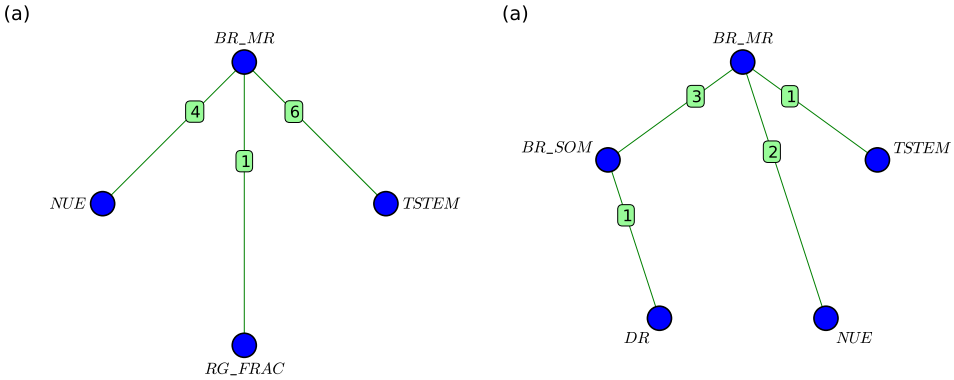


Figure 9. Relevant joint Sobol indices, S_{ij}, corresponding to September averages for (a) TVC and (b) TSC. Both sets of results are based on D_{ST}. The labels on each line show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.

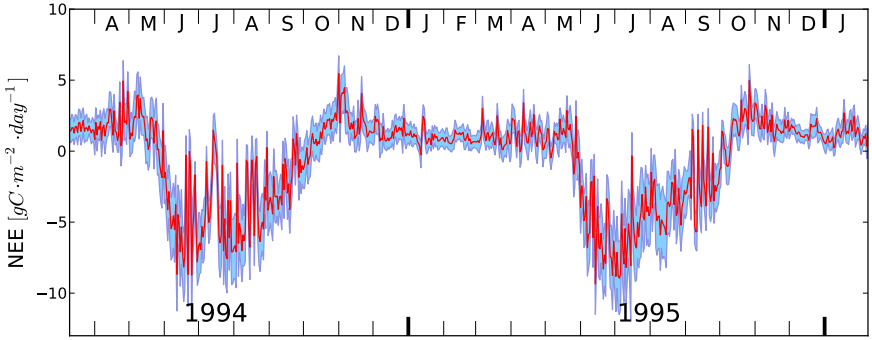
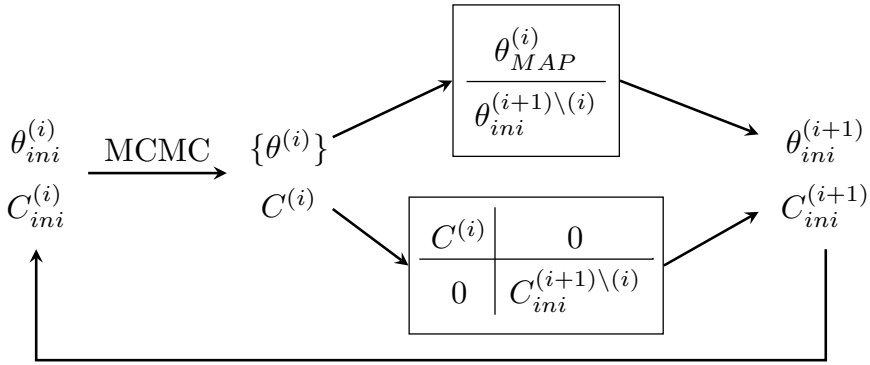


Figure 10. Snapshot of NEE observations (with red line) for the Harvard Forest site. The light blue region, bordered by thick blue lines corresponds to $\pm 2\sigma$ around the daily NEE values.



~~Normalized histograms for diagonal entries of the Fisher Information Matrix corresponding to parameters *astem*, *tstem*, *troot*, *br_lit*, *br_som*, and *dr*. Results are based on NEE and an ensemble of parameter values drawn from the corresponding prior distributions.~~

Schematic of the iterative process for parameter calibration. The MCMC sampling of the joint density for the set of parameters $\theta^{(i)}$ starts at ~~$\theta^{(i)}_{\text{ini}}$~~ $\underset{\sim}{\theta^{(i)}_{ini}}$ using an initial proposal covariance ~~$\mathbf{C}^{(i)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i)}_{ini}}$. For the following iteration, $(i+1)$, the initial condition is constructed using the MAP estimate for $\theta^{(i)}$, augmented with initial conditions, in this case the nominal values, for the rest of parameters, $\theta^{(i+1)\setminus(i)}$. The initial proposal covariance ~~$\mathbf{C}^{(i+1)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i+1)}_{ini}}$ is constructed based on the sample covariance matrix for $\theta^{(i)}$, augmented with an initial proposal covariance for $\theta^{(i+1)\setminus(i)}$, ~~$\mathbf{C}^{(i+1)\setminus(i)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i+1)\setminus(i)}_{ini}}$.

~~Normalized histograms for diagonal entries of the Fisher Information Matrix corresponding to parameters *astem*, *tstem*, *troot*, *br_lit*, *br_som*, and *dr*. Results are based on NEE and an ensemble of parameter values drawn from the corresponding prior distributions.~~

Schematic of the iterative process for parameter calibration. The MCMC sampling of the joint density for the set of parameters $\theta^{(i)}$ starts at ~~$\theta^{(i)}_{\text{ini}}$~~ $\underset{\sim}{\theta^{(i)}_{ini}}$ using an initial proposal covariance ~~$\mathbf{C}^{(i)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i)}_{ini}}$. For the following iteration, $(i+1)$, the initial condition is constructed using the MAP estimate for $\theta^{(i)}$, augmented with initial conditions, in this case the nominal values, for the rest of parameters, $\theta^{(i+1)\setminus(i)}$. The initial proposal covariance ~~$\mathbf{C}^{(i+1)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i+1)}_{ini}}$ is constructed based on the sample covariance matrix for $\theta^{(i)}$, augmented with an initial proposal covariance for $\theta^{(i+1)\setminus(i)}$, ~~$\mathbf{C}^{(i+1)\setminus(i)}_{\text{ini}}$~~ $\underset{\sim}{C^{(i+1)\setminus(i)}_{ini}}$.

Figure 11. ~~Normalized histograms for diagonal entries of the Fisher Information Matrix corresponding to parameters *gdd_min*, *gdd_max*, *leaffall*, *q10_mr*, *br_mr*, and *rg_frac*. Results are based on NEE and an ensemble of parameter values drawn from the corresponding prior distributions.~~

~~Normalized histograms for diagonal entries of the Fisher Information Matrix corresponding to parameters *astem*, *tstem*, *troot*, *br_lit*, *br_som*, and *dr*. Results are based on NEE and an ensemble of parameter values drawn from the corresponding prior distributions.~~

Schematic of the iterative process for parameter calibration. The MCMC sampling of the joint density for the set of parameters $\theta^{(i)}$ starts at $\theta_{\text{ini}}^{(i)}$ $\theta_{ini}^{(i)}$ using an initial proposal covariance $\mathbf{C}_{\text{ini}}^{(i)}$ $C_{ini}^{(i)}$. For the following iteration, $(i+1)$, the initial condition is constructed using the MAP estimate for $\theta^{(i)}$, augmented with initial conditions, in this case the nominal values, for the rest of parameters, $\theta^{(i+1)\setminus(i)}$. The initial proposal covariance $\mathbf{C}_{\text{ini}}^{(i+1)}$ $C_{ini}^{(i+1)}$ is constructed based on the sample covariance matrix for $\theta^{(i)}$, augmented with an initial proposal covariance for $\theta^{(i+1)\setminus(i)}$, $\mathbf{C}_{\text{ini}}^{(i+1)\setminus(i)}$ $C_{ini}^{(i+1)\setminus(i)}$.

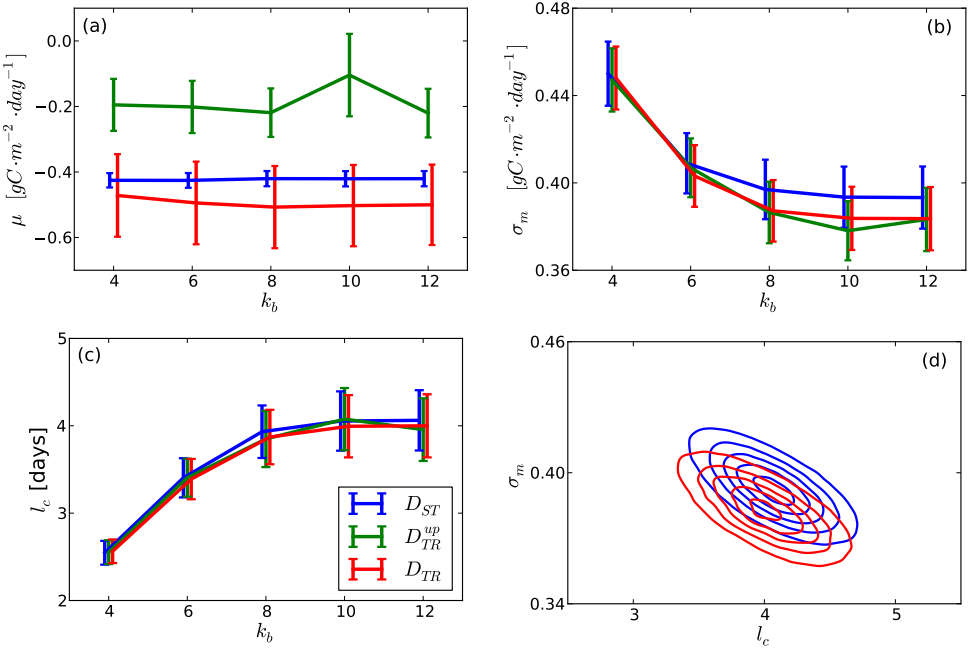


Figure 12. Convergence of model error components with increasing bandwidth of the covariance matrix: (a) μ, (b) σ_m, and (c) l_c. The joint 2D marginal density of (σ_m,l_c) for $k_b = 12$ is shown in (d). In addition to D_{ST} and D_{TR} setups, we also considered "D_{TR}^{up}", a setup equivalent to D_{TR}, but with uniform priors assumed for the vegetation and soil Carbon pools.

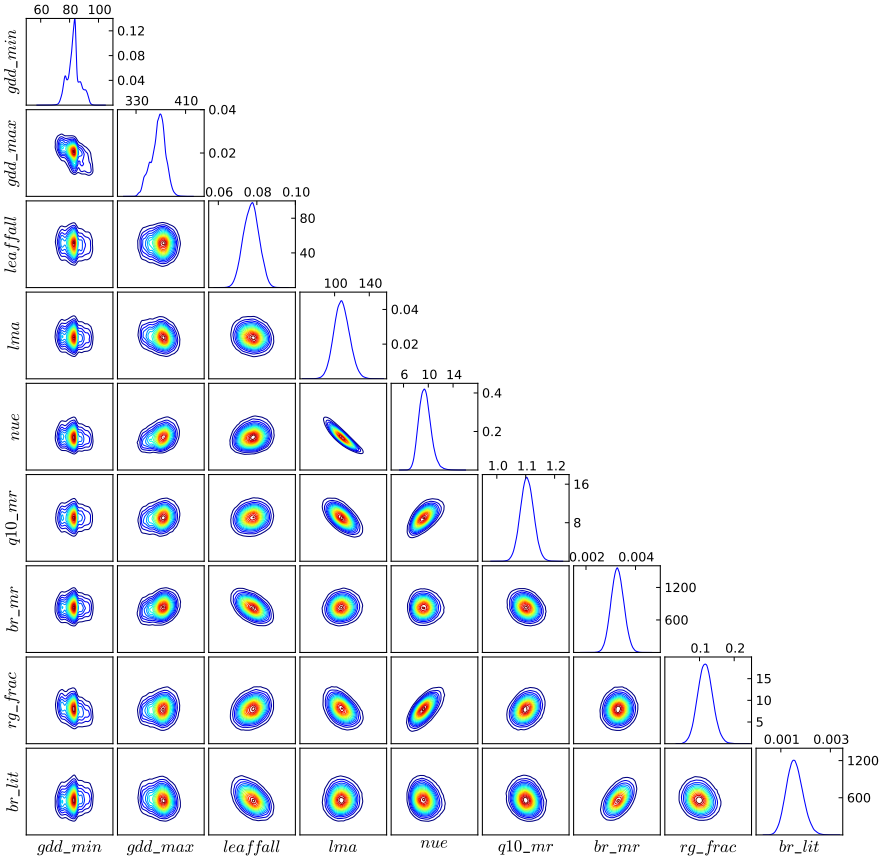
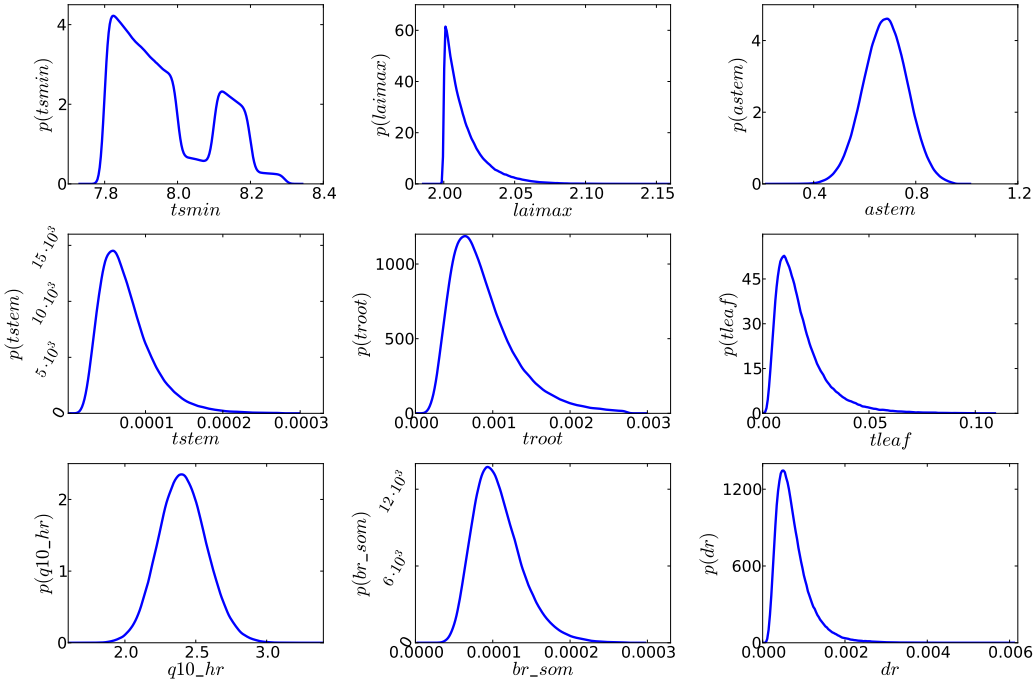


Figure 13. ~~D18-problem~~D_{ST}-problem: ~~1-D~~ 1D marginal and ~~2-D~~ 2D joint marginal PDFs for parameters showing distance correlation factors above ~~0.3~~0.4, see also Table 3. Marginal PDFs are estimated via KDE based on approximately 5×10^5 MCMC samples.



~~Kulback–Leibler divergence, $D_{KL}(p||q)$, between prior q and posterior p densities and scaled inverse standard deviation, $1/\sigma^* = \sigma_q/\sigma_p$ for select parameters.~~

~~First order Sobol indices for June and September average NEE values. The red bars show results~~ Marginal PDFs are estimated via KDE based on ~~joint 3-D (for group G1) and 2-D (for groups G2 and G3) posterior distributions and 1-D marginal posterior distributions for the rest; blue bars show results based on 1-D marginal posterior distributions for all parameters; G1=~~{~~*lma*, *nue*, *rg_frac*~~}~~, ~~G2=~~{~~*q10_mr*, *q10_hr*~~}~~, ~~G3=~~{~~*gdd_min*, *gdd_max*~~}~~approximately 5×10^5 MCMC samples.

~~Kulback–Leibler divergence, $D_{KL}(p||q)$, between prior q and posterior p densities and scaled inverse standard deviation, $1/\sigma^* = \sigma_q/\sigma_p$ for select parameters.~~

~~First order Sobol indices for June and September average NEE values. The red bars show results~~ Marginal PDFs are estimated via KDE based on ~~joint 3-D (for group G1) and 2-D (for groups G2~~

~~and G3) posterior distributions and 1-D marginal posterior distributions for the rest; blue bars show results based on 1-D marginal posterior distributions for all parameters; G1={*lma, nue, rg_frac}, G2={q10_mr, q10_hr}, G3={gdd_min, gdd_max}~~approximately 5×10^5 MCMC samples.

Figure 14. ~~D18 problem~~D_{ST} problem: ~~1-D~~ 1D marginal PDFs for parameters showing distance correlation factors less than ~~0.3~~ 0.4 with other parameters, see also Table 3.
~~Kulback–Leibler divergence, $D_{KL}(p||q)$, between prior q and posterior p densities and scaled inverse standard deviation, $1/\sigma^* = \sigma_q/\sigma_p$ for select parameters.~~
~~First order Sobol indices for June and September average NEE values. The red bars show results~~ Marginal PDFs are estimated via KDE based on ~~joint 3-D (for group G1) and 2-D (for groups G2 and G3) posterior distributions and 1-D marginal posterior distributions for the rest; blue bars show results based on 1-D marginal posterior distributions for all parameters; G1={*lma, nue, rg_frac}, G2={q10_mr, q10_hr}, G3={gdd_min, gdd_max}~~approximately 5×10^5 MCMC samples.

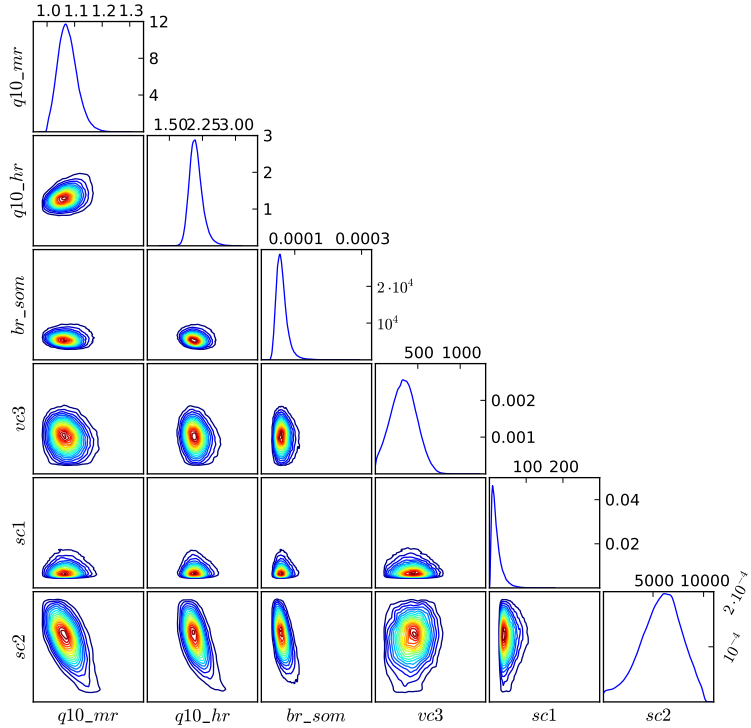


Figure 15. ~~D23-problem~~ D_{TR}-problem: ~~1-D~~ 1D marginal and ~~2-D~~ 2D joint marginal PDFs for select parameters ~~showing distance correlation factors above 0.3, see also Table **??**~~correlated with the Carbon pools. Marginal PDFs are estimated via KDE based on approximately 5×10^5 MCMC samples.

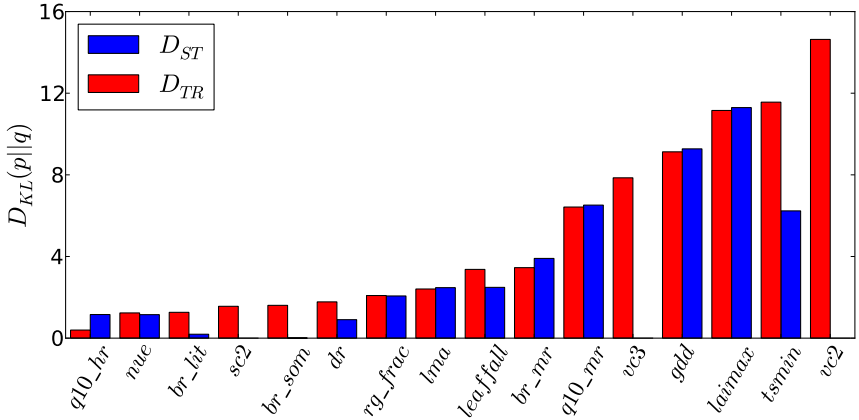


Figure 16. ~~D23-problem: 1-D marginal~~ Kulback-Leibler divergence, $D_{KL}(p||q)$, between prior q and ~~2-D joint marginal PDFs~~ posterior p densities for parameters ~~correlated~~ with ~~the Carbon pools~~ $D_{KL} > 0.5$ for both D_{ST} and D_{TR}.

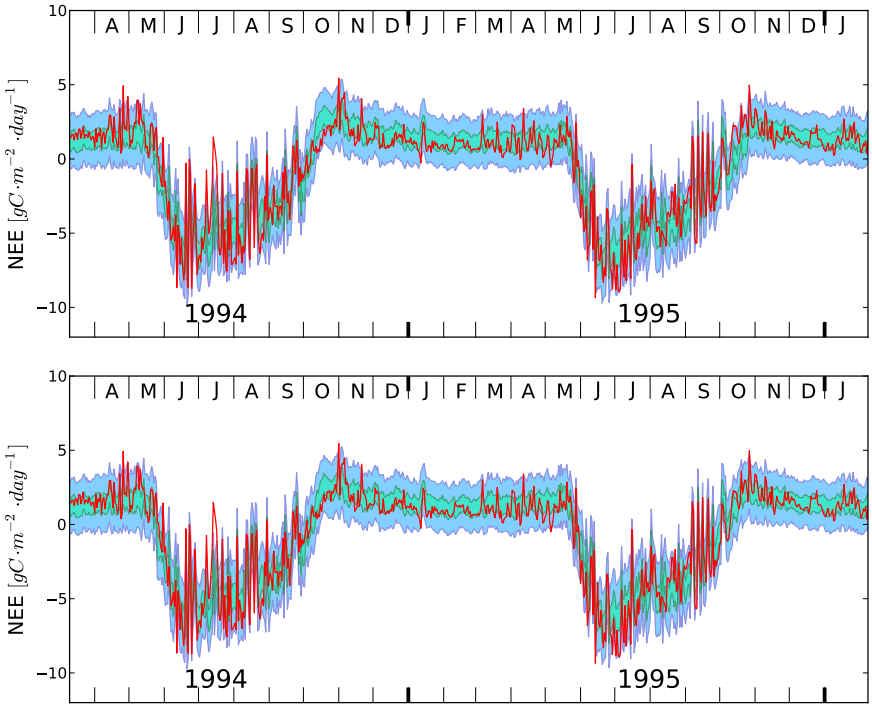


Figure 17. ~~Posterior~~ Marginal posterior predictive ~~distributions~~ densities using the calibration results for ~~D18~~ D_{ST} (top frame) and ~~D23~~ D_{TR} (bottom frame) presented in ~~Sect.~~Section 4. The blue regions correspond to the daily ~~5–95~~5 – 95% quantile range and the green regions to ~~25–75~~25 – 75% quantile range. The red line shows the daily NEE observations.