

CROP PHYSIOLOGY CALIBRATION IN CLM

RESPONSE TO REVIEWERS

NOTES AND MAJOR REMARKS

- All changes made in the manuscript are marked with blue color (revised document enclosed at the end of the response).
- The reviewers' comments in this document are presented in *italics*. Our response follows each comment.
- In order to demonstrate the transferability of the calibrated parameters, we repeated the validation experiment at one more location: Mead, NE, US.
- Addressing the reviewers' comments, we re-ran our calibration procedure with 4 more parameters (i.e., for a total of 14 parameters). The results of this experiment are presented as Exhibit I of this response.

REVIEWER #1

Bilionis et al. present a detailed description of a new calibration procedure for parameters within the CLM 4.5 crop model. They apply the procedure to estimate values for 10 poorly constrained parameters for soybean at a single site in the U.S.A. In a test with artificial data, their calibration method is able to reproduce the parameters used to generate that artificial data, and the calibration improves the performance of the crop model at the site for which it was calibrated. Unfortunately, I find myself insufficiently qualified to assess the quality or the novelty of the mathematics underlying section 4 of the manuscript. I have therefore restricted my review (with the exception of a few typos) to other aspects of the manuscript. As section 4 represents the core of the method, my review should therefore only be considered in addition to others in which this section can be properly assessed. Nonetheless, I find some concerns over the manuscript in its current form, which I detail below, and which I believe should be addressed before publication. Overall I find the standard of language is satisfactory, and the paper fits within the scope of GMD.

MAJOR COMMENTS

1. *As a demonstration of a method to calibrate a site-scale model I suspect that the manuscript has a great deal of merit. But as a method to calibrate a large-scale model it is lacking in details of how the method should be applied to calculate a generally applicable parameterization. It is mentioned in the discussion that there is insufficient data to determine globally relevant parameters. I accept that this is probably the case, but one must question the logic behind developing a calibration procedure for a large-scale model that relies on non-existent data in order for it to be applied at the scale the model is intended to*

operate at. How then can the calibration procedure be used to improve CLM performance? At the very least the authors should make an evaluation of whether the parameters derived for this single site improve the simulation of soybean at other sites in North America. At minimum, data on yield should be available.

This is an excellent suggestion. To that end, we ran a validation experiment with independent data (not used for calibration) from a different site (Mead, NE, USA). The results from the new experiments have been added in the manuscript (see Figure 8). We believe that the results of this independent validation re-enforce the validity of the hypothesis that our scheme generates physically relevant, and thus, transferable parameters.

2. *Regarding the choice of calibration parameters. It seems strange to calibrate to NEE, when NEE also strongly depends on heterotrophic (i.e. soil) respiration, which is largely out of the remit of the parameters being calibrated. By including NEE in the calibration there is a risk of parameterizing the crops to account for deficiencies in the soil scheme.*

We are aware of the implications of using NEE for our calibration procedure. We previously experimented without including NEE for the calibration and found very little impact on the results with the exception of a slightly better agreement with GPP. Our choice to leave NEE in the calibration was related to the lack of crop rotation in the model that is at the site. By calibrating with NEE we can indirectly include a representation for corn soybean rotation without modeling it explicitly. In future studies we plan to address some of the below-ground parameters in this calibration framework and those related to soil properties are a priority.

3. *Presumably the level of re-translocation of carbon from leaves/stems to grain is dependent on nitrogen availability (if not then references should be provided). If so, then the fitting of $fleafcn$ and $fstemcn$ means that the fidelity of the calibration is conditional on the nitrogen availability in the soil. Such effects are referred to in the results and discussion sections, but to me it seems strange to calibrate these parameters at all when observations covering a range of nitrogen availabilities are not available – it once again appears to go against the principle of developing a general parameterization for a large-scale model.*

Previous studies found that the importance of retranslocation as a source of nitrogen for grain in maize is more important than uptake during the grain filling stage of plant development (Gallias et al., 2006; Gallias et al., 2007). In fact, Crawford et al. (1982) found that the amount of retranslocated nitrogen was the same in maize plants that had been fertilized as those that did not have fertilization and that nitrogen fluxes in the grain and cob were unaffected by nitrogen in the rooting medium. The importance of retranslocation during the grain development has been observed for other crop types as well (Simpson et al., 1983); therefore, we feel like nitrogen availability will play a small role. However, retranslocation is a complex process that has been greatly simplified in the model. We have explored the sensitivity of the model to initial soil conditions (including nitrogen) and plan to expand our analysis in future studies.

MINOR COMMENTS

4. *The abstract should provide a description of the results also.*

The following phrase was appended to the abstract: “The model showed significant improvement of crop productivity with the new calibrated parameters. We demonstrate that the calibrated parameters are applicable across alternative years and different sites.”

5. *In the abstract and on pg. 6735, l. 27, it is claimed that the calibration will improve “projections” of plant development and carbon fluxes. This terminology is a bit loose, as it seems to imply long-term (e.g. 100 year) projections under changing climate, atmospheric CO2 mixing ratios and nutrient availability. It is clear that for environmental conditions similar to the calibration conditions, the calibrated model would be expected to perform better. However, it is far from clear why a calibration to current conditions should actually improve projections for conditions that lie outside the range of the calibration conditions. The terminology should either be altered, or better still, this point should be discussed.*

We have changed the wording from “projection” to “simulation”. We agree with the reviewer that we cannot predict the future without calibrating against future data, however, we have no hope of simulating future crop productivity and yield without first getting the current productivity right. Since the model clearly does not do well at current crop representation, by addressing that first, we can assume our future will also be improved. Improving predictions under future climate is a goal that we are working toward through the application of SMC to calibrate parameters using available data. This information is expected to improve projections, but it is difficult to estimate by how much. More data (perhaps sites at different locations) will give us more confidence. Although it is not stated in the introduction we do list your concern as one of the main limitations of this study, but not of the method itself.

6. *p. 6735, l. 19: It is not defined what is coupled or uncoupled.*

The reviewer is right to notice that “coupled or uncoupled modes” are not clearly understood by readers unfamiliar with the internals of the cited multiple very fast simulated annealing (MVFSA) (Yang et al. 2012) method. Since the phrase under question is not essential for our discussion, we chose to remove it.

7. *p. 6737, l. 9: Presumably “fits within the framework of natural vegetation” means that all the processes apart from the allocation scheme are treated exactly the same for crops as for natural vegetation?*

Yes, that is correct. We have changed the wording to clarify that crops are modeled similarly to natural vegetation with the exception of the growing phases. The sentence the reviewer refers to has changed to “Crops are modeled similarly to natural vegetation with the main exception of how allocation is defined via a different growing scheme, which is separated into four phases: planting, emergence, grain fill, and harvest.”

8. p. 6740, l. 19: “Level 4” will be meaningless to most readers. Suggest to remove or explain.

We chose to remove “Level 4”. We reported Level 4 to provide readers with the information needed for reproducibility; however, sufficient detail was provided to readers to determine which AmeriFlux data was used so this additional information is indeed unnecessary.

9. Eq. 3 appears to imply the calculation of a slope for GRAINC (z_3), however this is not mentioned in the accompanying text.

The reviewer is right. The slope of GRAINC (z_3) is not used in the calibration. We have updated Eq. 3 to reflect exactly the variables we use. Also, notice, that we have fixed a typo in the number of calibrated quantities. The previous notation indicated that there were 11 quantities. The right number is 10.

10. p. 6752, l. 15: Over precision of reported numbers.

The reviewer is right. Fixed.

11. p. 6752, l. 20: Timings would be more useful if accompanied by a number of cores and a CPU clock speed.

We do report the number of cores implicitly: “Each particle is assigned to a different computational core.” We have changed this phrase to: “Each particle is assigned to a different computational core, i.e., we use 1280 computational cores.”

12. The text appears to imply that *stemcn* is calibrated, but it is not listed in Table 1. Is *livewdcn* the same as *stemcn*? If so, nomenclature should be consistent throughout.

We apologize for the confusion. The variable *stemcn* was changed to *livewdcn* in model version 4.5 to eliminate redundancy. We have corrected all references of *stemcn* in the manuscript to reflect *livewdcn* to ensure consistency.

13. Figure 4, *grperc* panel: The distribution here appears to move out of the range of allowable parameters. This would appear to suggest that the range should be widened. Alternatively, if such a widening would lead to unphysical values, then this points to a deficiency in the model formulation, which would presumably merit some discussion

grperc is the ratio of growth respiration carbon to new growth carbon. Since observations of growth respiration are difficult to separate from maintenance respiration, there is a great deal of uncertainty in this parameter, and it is often assumed to be proportional to the growth rate. The value of *grperc* is typically used in crop models is small, around 0.2-0.3, although growth respiration can be higher for soybean (Bunce and Ziska, 1996). We don’t believe the range should be expanded larger than 0.5.

TYPOS/GRAMMATICAL ERRORS

14. p. 6734, l. 2: “terrestrial ground” sounds strange. I would suggest “of the land surface”.

Fixed.

15. p. 6734, l. 3: *“applied to”, rather than “used for”.*

Fixed.

16. p. 6734, l. 24: *I would suggest to change to “...version 3.5 to switch to a cold bias in CLIM4.0.”*

Fixed.

17. p. 6735 l. 25: *The logic of the sentence seems backwards. Suggest “This is the case for the CLM-Crop model, as well as for ESMs.”*

Fixed.

18. p. 6742, l. 8: *Suggest “prior to this point” instead of “ahead of it”.*

Fixed.

19. p. 6742, l. 15: *Degrees should be defined. Presumably either “C” or “K”.*

Fixed.

20. p. 6747, l. 7: *“moment” instead of “momment”.*

Fixed.

21. p. 6749, l.9: *“effective”.*

Fixed.

22. p. 6750, l. 16: *“degeneracy”.*

Fixed.

23. p. 6753, l. 24. *“ultimately”.*

Fixed.

BIBLIOGRAPHY

1. Gallias, A., Coque, M., Quilléré, I., Prioul, J.-L., and Hirel, B.: Modelling postsilking nitrogen fluxes in maize(*Zea mays*) using ¹⁵N-labelling field experiments, *New Phytologist*, 172, 696-707, 2006.
2. Gallias, A., Coque, M., Le Gouis, J., Prioul, J. L., Hirel, B., and Quilléré, I.: Estimating the proportion of nitrogen remobilization and of postsilking in nitrogen uptake allocated to maize kernels by nitrogen-15 labeling, *Crop Science*, 47, 685-693, 2007
3. Crawford, T. W., Rendig, V. V., and Broadbent, F. E.: Sources, fluxes, and sinks of nitrogen during early reproductive growth of maize (*Zea mays* L.), *Plant Physiology*, 70, 1654-1660, 1982.
4. Simpson, R. J., Lambers, H., and Dalling, M. J.: Nitrogen redistribution during grain growth in wheat (*Triticum aestivum* L.), *Plant Physiology*, 71, 7-14, 1983.
5. Bunce, J. A., and Ziska, L. H.: Responses of respiration to increases in carbon dioxide concentration and temperature in three soybean cultivars, *Annals of Botany*, 77, 507-514, 1996.
6. Sargsyan, K., Safta, C., Berry, R., Ray, J., Debusschere, B., and Naim, H.: Efficient uncertainty quantification methodologies for high-dimensional climate land models, Sandia National Laboratory, Albuquerque, NM, 2011.

EXHIBIT I

RERUNNING THE CALIBRATION PROCEDURE WITH ADDITIONAL PARAMETERS.

We re-ran the calibration procedure with 4 more parameters (i.e., for a total of 14 parameters). We considered parameters that are, presumably, important for GPP, LAI, and Total Vegetation Carbon. Correlating parameters across these three variables, we find the following important parameters: the percent of nitrogen in the leaf as rubisco, Q10 parameter for heterotrophic respiration, Q10 parameter for maintenance respiration, and base rate for maintenance respiration. Including these additional parameters doesn't change the results significantly, but does modify parameter values slightly. The results are presented in Pages 8-13 of this exhibit. Their format is exactly the same as the paper's.

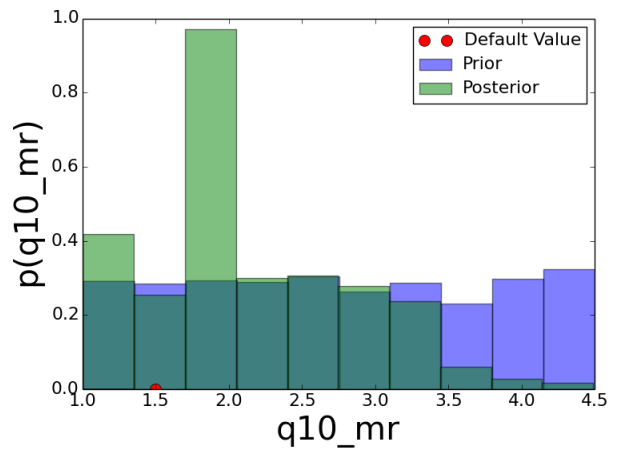
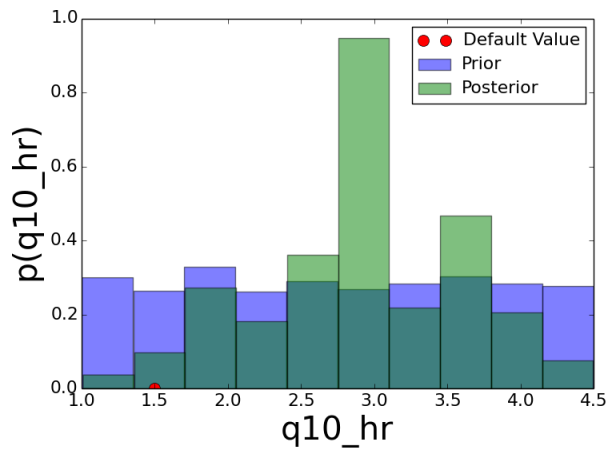
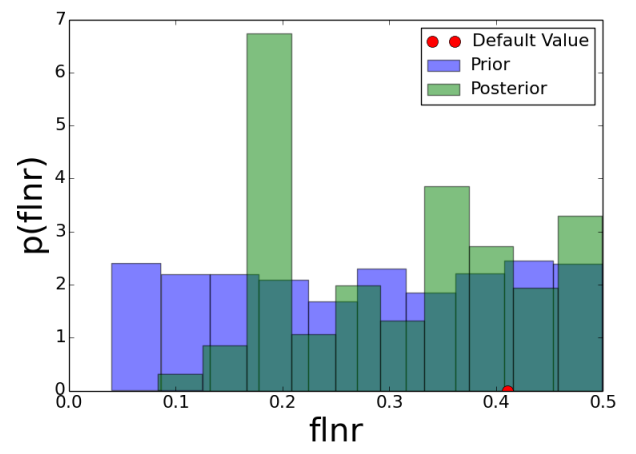
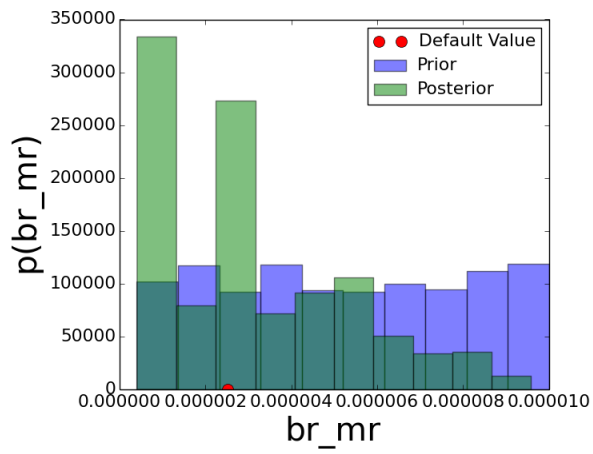
TABLE 1 - PRIOR INFORMATION ON THE PARAMETERS. THE NEW PARAMETERS ARE IN BOLD FONTS.

Name	Description	Constraints	Min	Max	Default
xl	leaf/stem orientation index	none	-0.40	0.60	-0.40
slatop	specific leaf area at top of canopy	none	0.00	0.15	0.07
leafcn	leac C	none	5.00	80.00	25.00
flnr	fraction of leaf N in the Rubisco enzyme (gN Rubisco / gN leaf)	none	0.04	0.50	0.41
frootcn	fine root C	none	40.00	100.00	42.00
livewdcn	live wood (phloem and ray parenchyma C)	none	10.00	100.00	50.00
grperc	growth respiration factor 1	none	0.10	0.50	0.25
grpnw	growth respiration factor 2	none	0.00	1.00	1.00
br_mr	base rate for maintenance respiration	none	4.00e-07	1.00e-05	2.53e-06
q10_mr	Q10 for maintenance respiration	none	1.00	4.50	1.50
q10_hr	Q10 for heterotrophic respiration	none	1.00	4.50	1.50
graincn	organ carbon nitrogen ratio	none	20.00	100.00	50.00
fleafcn	final leaf carbon nitrogen ratio	> leafcn	10.00	100.00	65.00
fstemcn	final stem carbon nitrogen ratio	> livewdcn	40.00	200.00	130.00

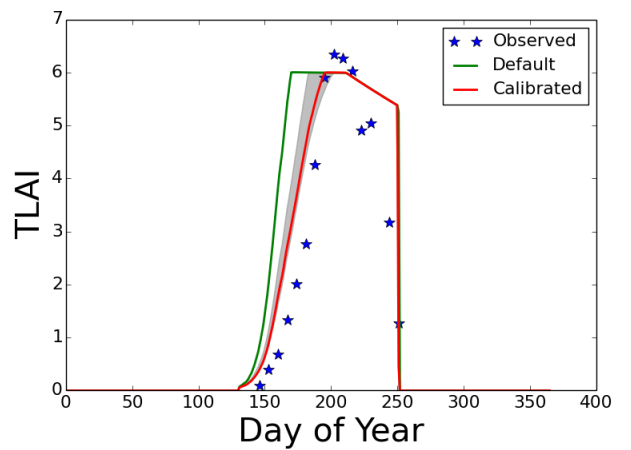
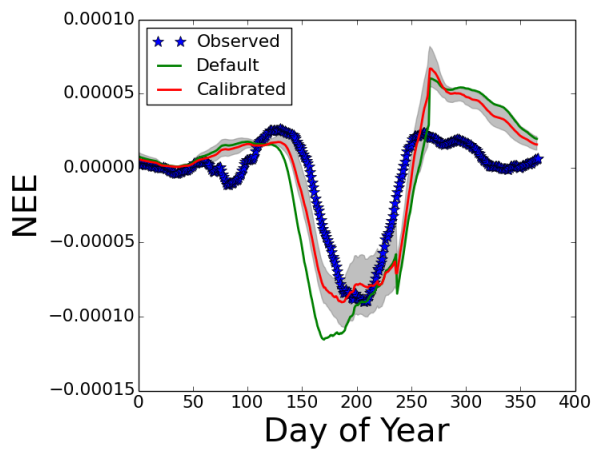
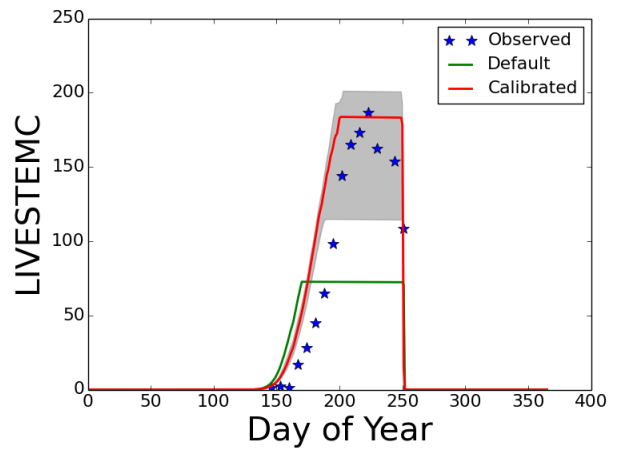
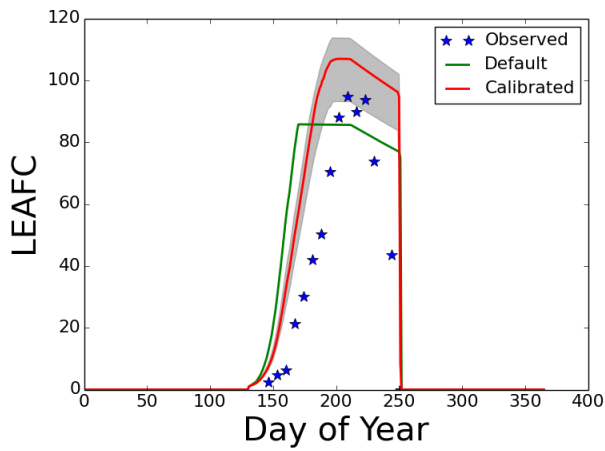
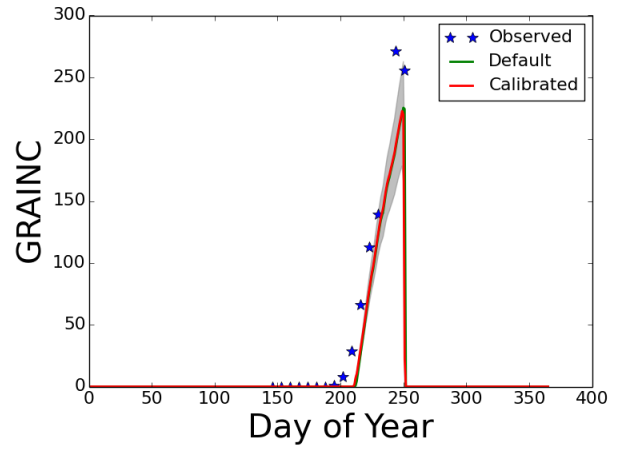
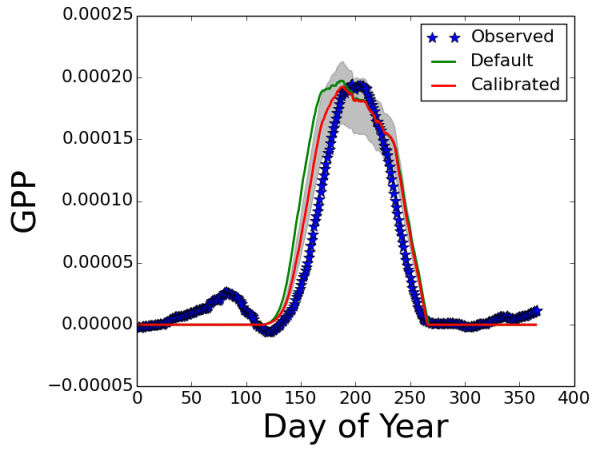
TABLE 2 - CALIBRATION RESULTS AFTER INCLUDING THE NEW PARAMTERS. THE OLD CALIBRATION RESULTS (WHEN AVAILABLE) ARE IN PARENTHESIS. THE OBSERVED DESCREANCIES BETWEEN THE TWO CASES ARE WITHIN THE STATISTICALL ERROR. THE PARAMETERS WERE TIGHTLY BOUND BEFORE (SLATOP AND LEAFCN) REMAIN SO AFTER THE INCLUSION OF THE NEW VARIABLES.

Name	Median	$p = 0.05$	$p = 0.95$
xl	-0.01 (-0.09)	-0.33 (-0.32)	0.35 (0.51)
slatop	0.06 (0.06)	0.05 (0.05)	0.06 (0.07)
leafcn	22.58 (29.51)	11.90 (25.26)	38.83 (35.79)
flnr	0.32	0.17	0.47
frootcn	77.54 (66.90)	48.69 (46.02)	94.89 (95.66)
livewdcn	45.83 (48.16)	20.48 (10.56)	90.69 (85.89)
grperc	0.33 (0.40)	0.23 (0.18)	0.46 (0.50)
grpnow	0.32 (0.54)	0.04 (0.05)	0.84 (0.91)
br_mr	2.42e-06	8.29e-07	7.71e-06
q10_mr	1.86	1.19	3.19
q10_hr	2.88	1.73	4.01
graincn	79.14 (55.69)	33.15 (27.18)	99.36 (89.43)
fleafcn	70.32 (64.30)	30.38 (59.47)	96.18 (180.88)
fstemcn	118.64 (118.02)	47.84 (59.47)	176.14 (180.88)

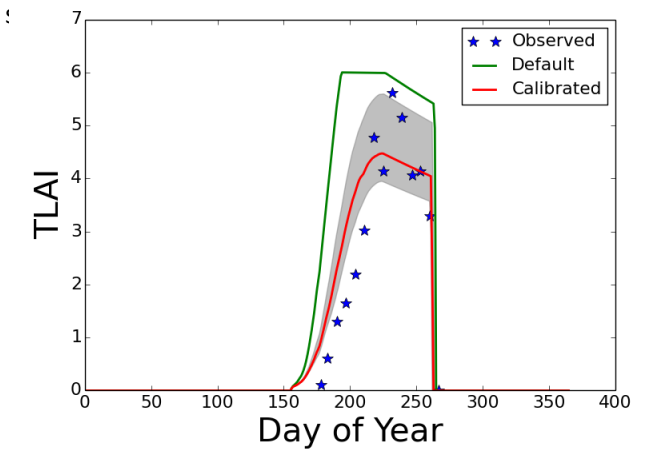
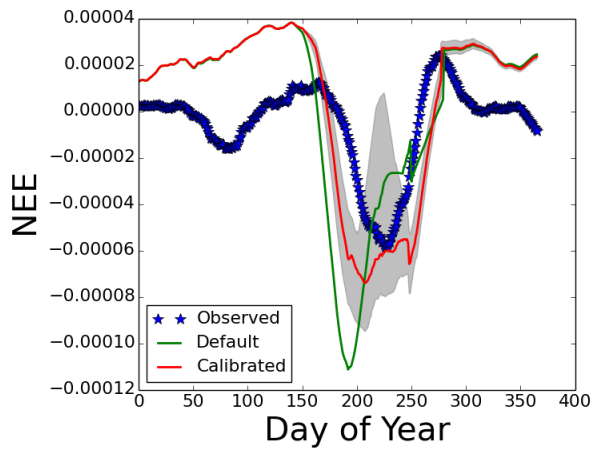
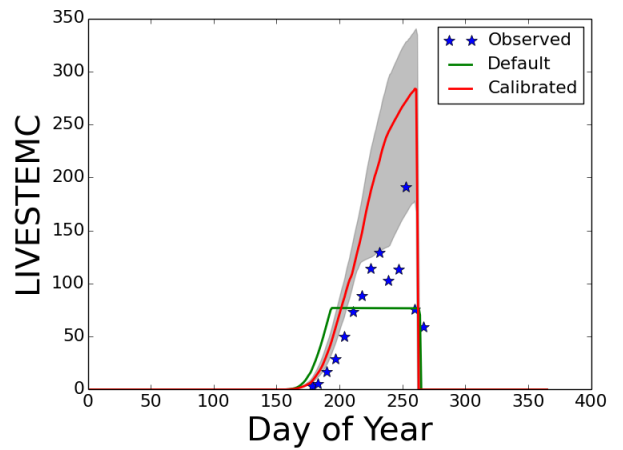
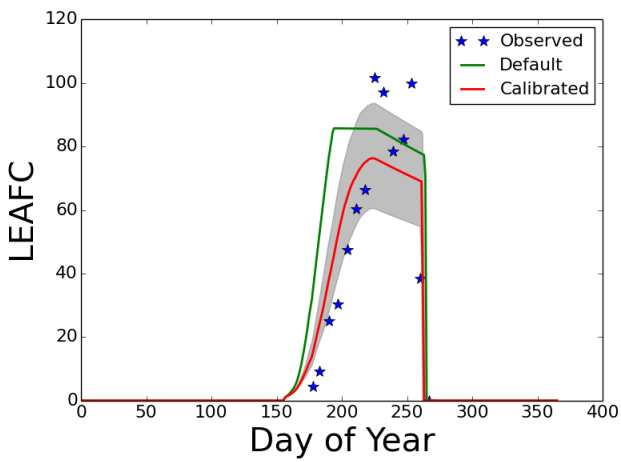
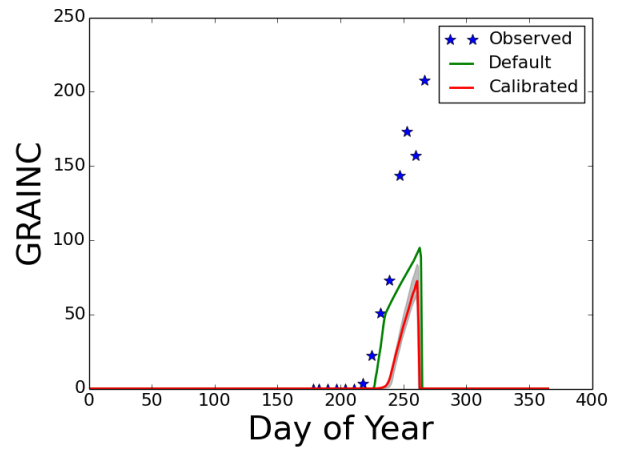
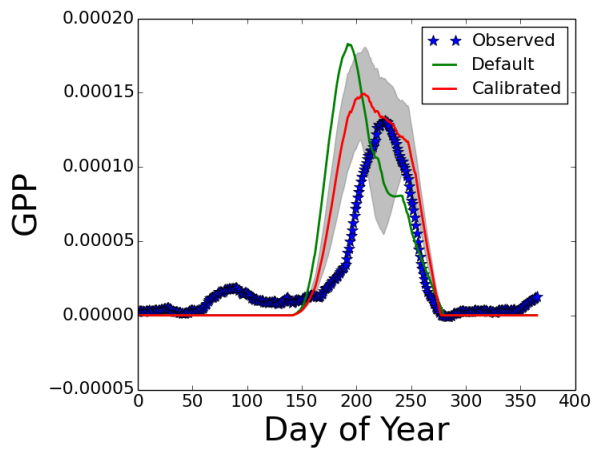
POSTERIOR DISTRIBUTIONS OF NEWLY ADDED PARAMETERS



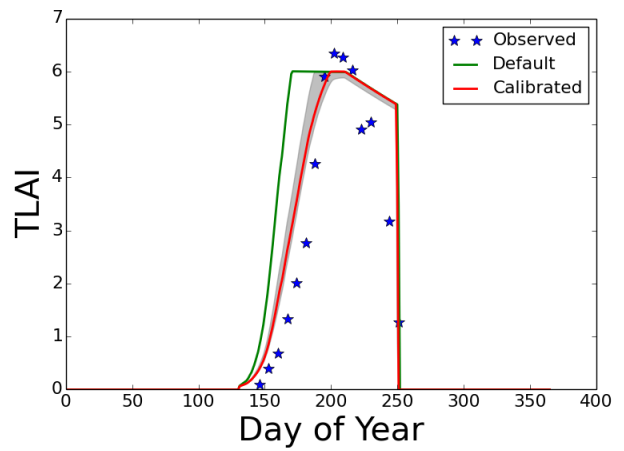
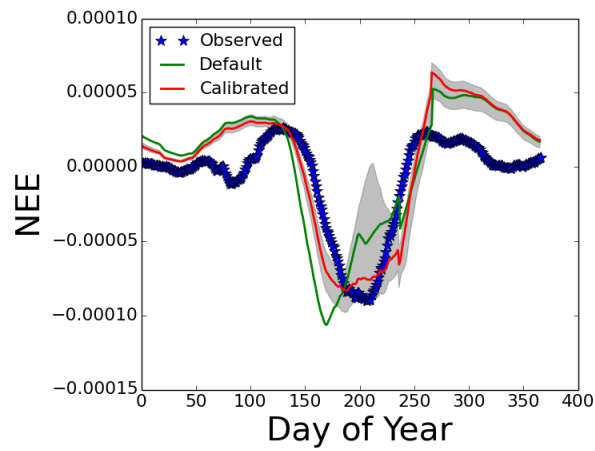
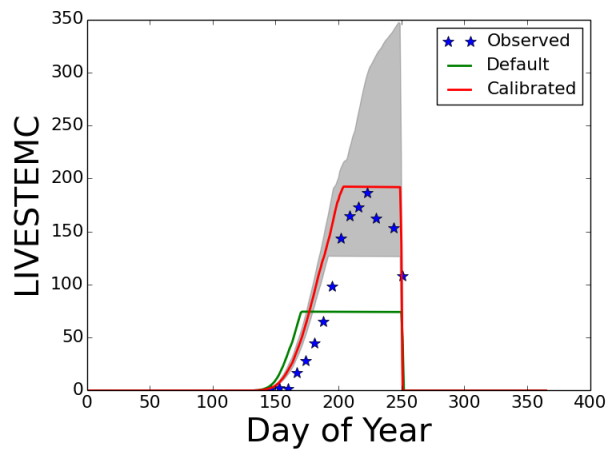
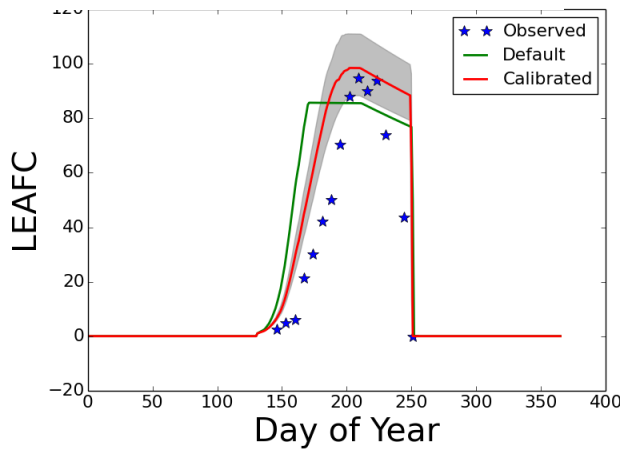
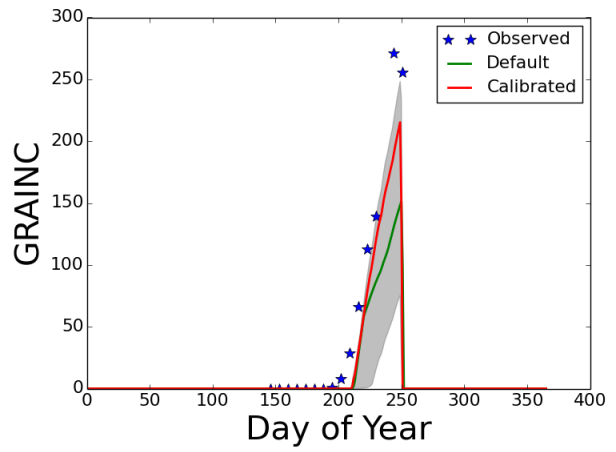
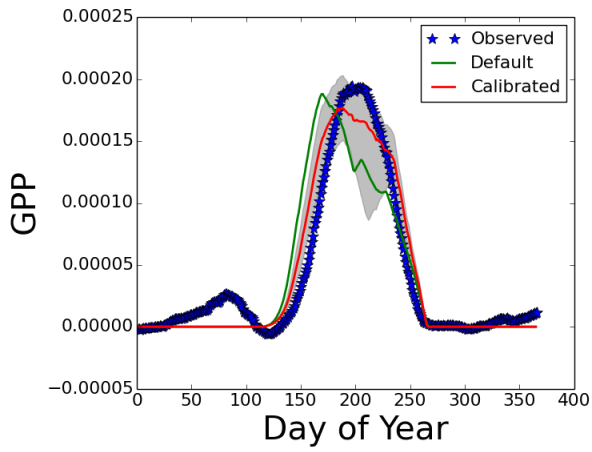
CALIBRATED OUTPUT (BONDVILLE, IL, 2004)



VALIDATION EXPERIMENT (BONDVILLE, IL, 2002)



VALIDATION EXPERIMENT (BONDVILLE, IL, 2004)



Crop Physiology Calibration in CLM

I. Bilionis¹, B. A. Drewniak², and E. M. Constantinescu¹

¹Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA

²Environmental Science Division, Argonne National Laboratory, Argonne, IL, USA

Correspondence to: E. M. Constantinescu
(emconsta@mcs.anl.gov)

Abstract. Farming is using more of the land surface ~~terrestrial-ground~~, as population increases and agriculture is increasingly applied to ~~used-for~~ non-nutritional purposes such as biofuel production. This agricultural expansion exerts an increasing impact on the terrestrial carbon cycle. In order to understand the impact of such processes, the Community Land Model (CLM) has been augmented with a CLM-Crop extension that simulates the development of three crop types: maize, soybean, and spring wheat. The CLM-Crop model is a complex system that relies on a suite of parametric inputs that govern plant growth under a given atmospheric forcing and available resources. CLM-Crop development used measurements of gross primary productivity and net ecosystem exchange from AmeriFlux sites to choose parameter values that optimize crop productivity in the model. In this paper we calibrate these parameters for one crop type, soybean, in order to provide a faithful projection in terms of both plant development and net carbon exchange. Calibration is performed in a Bayesian framework by developing a scalable and adaptive scheme based on sequential Monte Carlo (SMC). ~~The model showed significant improvement of crop productivity with the new calibrated parameters. We demonstrate that the calibrated parameters are applicable across alternative years and different sites.~~

1 Introduction

Development of Earth system models (ESMs) is a challenging process, involving complex models, large input datasets, and significant computational requirements. As models evolve through the introduction of new processes and through improvement of algorithms, the ability of the models to accurately simulate feedbacks between coupled systems improves, although results may not have the desired impact on all areas. For example, Lawrence et al. (2012) es-

imate that changes to the hydrology parameterization may be responsible for the warm bias in high-latitude soils in the Community Land Model (CLM) version 3.5 to ~~switch to a cold bias become-cold-biased~~ in CLM4.0. Although testing of ESMs is extensive, ensuring after new developments are merged that the model can still perform with limited (if any) degradation, on rare occasions model behavior can be negatively affected. The strong nonlinearity of such models also makes parameter fitting a difficult task; and as global models are developed by several different user groups simultaneously, combinations of multiple alterations make identifying the specific cause that leads to a new model output challenging. The CLM has been augmented with a CLM-Crop extension that simulates the development of three crop types: maize, soybean, and spring wheat (Drewniak et al., 2013). The CLM-Crop model is a complex system that relies on a suite of parametric inputs that govern plant growth under a given atmospheric forcing and available resources. CLM-Crop development used measurements of gross primary productivity (GPP) and net ecosystem exchange (NEE) from AmeriFlux sites to choose parameter values that optimize crop productivity in the model.

Global climate models have historically been tuned or calibrated to meet certain requirements, such as balancing the top of the atmosphere radiation budget (Bender, 2008; Hourdin et al., 2012; Mauritsen et al., 2012). Various techniques have been applied to models to adjust parameters, including using data assimilation (Pauwels et al., 2007), applying an ensemble Kalman filter (EnKF) (Hargreaves et al., 2004; Annan et al., 2005; Evensen, 2009), and using a sampling algorithm such as multiple very fast simulated annealing (MVFSA) (Yang et al., 2012). ~~-,which-can-be-used-to-calibrate-one-or-many-parameters-at-a-time-in-coupled-or-uncoupled-modes-~~ Most calibration strategies can be traced to a Bayesian approach that in most cases is simplified (e.g., MVFSA) or augmented with assumptions that make the problem tractable

(e.g., EnKF). The tuning parameters that are not directly observed may be stated as an inverse problem (Tarantola, 2004). Inverse problems are, in general, very challenging especially when the data are sparse, the models are complex, and the state space is large. This is the case for CLM-Crop model, as well as for ESMs ESMs as well as for the CLM-Crop model.

Our goal is to calibrate some of the CLM-Crop parameters in order to improve model projection of plant development and carbon fluxes. To this end, we follow a Bayesian approach (Tarantola, 2004; Kaipio and Somersalo, 2004). We start by summarizing our initial state of knowledge in a prior probability distribution over the parameters we wish to calibrate. After making some observations, our updated state of knowledge is captured by the posterior distribution. Since the posterior is not analytically available, we attempt to approximate it using an ensemble of particles (samples) from it. To construct this particle approximation, we employ ideas from sequential Monte Carlo (SMC) Doucet et al. (2001). Basically, we define a one-parameter family of distributions of increasing complexity that starts at the prior and ends at the posterior. Starting from a particle approximation of the prior, we gradually move it toward the posterior by sequentially applying importance sampling. The scheme is highly parallelizable, since each particle of the approximation can be computed independently. The way we move the particle approximation towards the posterior is adjusted on the fly using the ideas developed by Bilonis and Koutsourelakis (2012) and Bilonis and Zabaras (2014). Each intermediate step of our scheme requires Markov chain Monte Carlo (MCMC) (Metropolis et al., 1953) sampling of the intermediate distributions. One of the novelties of this work, is the automatic construction of MCMC proposals for those intermediate steps using Gaussian mixtures (Blei and Jordan, 2005). The result is an algorithmic framework that can adjust itself to the intricacies of the posterior. As demonstrated by the numerical examples, our scheme can perform model calibration using very few evaluations and, by exploiting parallelism, at a fraction of the time required by plain vanilla MCMC.

We present the results from a twin experiment (self-validation) and calibration results and validation using real observations from two AmeriFlux tower sites in the mid-western United States, for the soybean crop type. The improved model will help researchers understand how climate affects crop production and resulting carbon fluxes, and additionally, how cultivation impacts climate.

2 The CLM-Crop model

CLM-Crop was designed and tested in the CLM3.5 model version (Drewniak et al., 2013) and in CLM4 (Levis et al., 2012). The crop model was created to represent crop vegetation similarly to natural vegetation for three crop types: maize, soybean, and spring wheat. The model simulates GPP and yield driven by climate, in order to evaluate the impact

of climate on cultivation and the impact of agriculture on climate. Crops are modeled within a grid cell sharing natural vegetation; however, they are independent (i.e., they do not share the same soil column). This approach allows management practices, such as fertilizer, to be administered without disturbing the life cycle of natural vegetation. For a full description of the crop model see the study by Oleson et al. (2013); the harvest scheme is described by Drewniak et al. (2013).

Crops are modeled similarly to natural vegetation with the main exception of how allocation is defined via a different growing scheme, which is separated into four phases: planting, emergence, grain fill, and harvest. ~~Although the design of the crop model fits within the framework of natural vegetation, crops have a significantly different growing scheme, separated into four phases: planting, emergence, grain fill, and harvest.~~ Each phase of growth changes how carbon and nitrogen are allocated to the various plant parts: leaves, stems, fine roots, and grain. During planting, carbon and nitrogen are allocated to the leaf, representative of seed. This establishes a leaf area index (LAI) for photosynthesis, which begins during the emergence phase. The emergence phase allocates carbon and nitrogen to leaves, stems, and roots using functions from the Agro-IBIS model (Kucharik and Brye, 2003). During the grain fill stage, decreased carbon is allocated to leaves, stems, and roots in order to fulfill grain requirements. When maturity is reached, harvest occurs: all grain is harvested, while leaves, stems, and roots are turned over into the litter pool. Residue harvest is not active in the model.

The allocation of carbon to each plant part is driven largely by the carbon-nitrogen (CN) ratio parameter assigned to each plant segment. CLM first calculates the potential photosynthesis for each crop type based on the incoming solar radiation and the LAI. The total nitrogen needed to maintain the CN ratio of each plant part is calculated as plant demand. If soil nitrogen is sufficient to meet plant demand, potential photosynthesis is met; however, if soil nitrogen is inadequate, the total amount of carbon that can be assimilated is down-scaled.

During the grain fill stage, a nitrogen retranslocation scheme is used to fulfill nitrogen demands by mobilizing nitrogen in the leaves and stems for use in grain development. This scheme uses alternate CN ratios for the leaf and stem to determine how much nitrogen is transferred from the leaves and stems into a retranslocation storage pool. The total nitrogen transferred at the beginning of the grain fill stage from the leaf and stem is represented by

$$\text{retransn}_{\text{leaf}} = \frac{C_{\text{leaf}}}{\text{leafcn}} - \frac{C_{\text{leaf}}}{\text{fleafcn}}, \quad (1)$$

$$\text{retransn}_{\text{stem}} = \frac{C_{\text{stem}}}{\text{livewdcn}} - \frac{C_{\text{stem}}}{\text{fstemcn}}. \quad (2)$$

C_{leaf} and C_{stem} are the total carbon in the leaf and stem, respectively; leafcn and livewdcn ~~stemcn~~ are the pregrain fill

CN ratios for the leaf and stem; and f_{leafcn} and f_{stemcn} are the post-grain fill CN ratios for the leaf and stem. All of the CN ratios are fixed parameters, which vary with crop type; default values are reported in Table 1.

In addition to the above, CLM-Crop has a fertilizer application and soybean nitrogen fixation, described by [Drewniak et al. \(2013\)](#). Planting date and time to maturity are based on temperature threshold requirements ([Levis et al., 2012](#)). For the calibration procedure, we used the actual planting date reported for the Bondville site for the year 2004. Crops are not irrigated in the model, nor do we consider crop rotation. Although rotation will have an impact on the carbon cycle both above and below-ground, CLM does not support crop rotation at this time.

The version of CLM-Crop detailed by [Drewniak et al. \(2013\)](#) was calibrated against AmeriFlux data for both the Mead, NE, and Bondville, IL, sites' plant carbon measurements, for both maize and soybean, using optimization techniques to fit parameters. When available, parameter values were taken from the literature or other models. Remaining parameters were derived through a series of sensitivity simulations designed to match modeled carbon output with AmeriFlux observations of leaf, stem, and grain carbon at the Bondville, IL, site and total plant carbon at the Mead, NE, (rainfed) site.

When CLM-Crop was ported into the CLM4.5 framework, the parameter values were no longer optimized as a result of various changes in model processes that affected how crops fit into the model framework. In addition, a new below-ground subroutine of carbon and nitrogen cycling is included in CLM4.5 ([Koven et al., 2013](#)), which has a strong influence on crop productivity. Therefore, we needed to retune the model parameters that represented crops with a more sophisticated approach described later in this paper.

2.1 Parameters affecting the crops

Over 100 parameters are defined in CLM4.5 to represent crops. Many of these parameters are similar to those that govern natural vegetation, but some are specific to crops. These parameters define a variety of processes, including photosynthesis, vegetation structure, respiration, soil structure, carbon nitrogen dynamics, litter, mortality, and phenology. To add further complication, parameters are assigned in various parts of the model; some parameters are defined in an external physiology file, some are defined in surface datasets, and others are hardcoded in the various subroutines of CLM4.5.

Performing a full model calibration for all parameters would be a monumental task, so we began our calibration process by narrowing down the parameters that are used only in crop functions or might have a large influence on crop behavior. Of this list, parameter values can be fixed across all vegetation types (or crop types), vary with crop type, or vary spatially and by crop type. We chose to limit the parameters to those that are either constant or vary with crop type.

Crop parameters are taken from the literature (when available) and used to determine a range of values appropriate for each crop type. When parameters are not available, optimization techniques are used to estimate parameter values based on CLM performance. Determining a full range of acceptable values was difficult for several parameters, and in some cases not possible. Of the full list of parameters in need of calibration, we began our approach with the ten parameters listed in Table 1 that may have a large influence on crop productivity and have the greatest uncertainty because the values are based on optimization from a previous model version. Six of the parameters are the carbon-nitrogen (CN) ratios for the various plant parts (leaf, stem, root, and grain). Since the leaf and stem account for nitrogen relocation during grain fill, they are represented by two separate CN ratios, to separate pre- and postgrain fill stages of plant development. They influence how carbon and nitrogen are allocated, thereby affecting growth, nutrient demand, photosynthesis, and so on, and are included as part of the physiology data file. Four additional parameters are included in the calibration process. The leaf/stem orientation is used to calculate the direct and diffuse radiation absorbed by the canopy, the specific leaf area at the top of the canopy is used with the leaf CN ratio to calculate the LAI, and the growth respiration factors determine the timing and quantity of carbon allocated toward respiration of new growth.

2.2 Description of the observational dataset

We used observations from the Bondville, IL, AmeriFlux tower located in the midwestern United States (40.01° N, 88.29° W) using an annual no-till corn-soybean rotation; a full site description is given by [Meyers and Hollinger \(2004\)](#). The site has been collecting measurements since 1996 of wind, temperature, humidity, pressure, radiation, heat flux, soil temperature, CO₂ flux, and soil moisture. Soybeans were planted in 2002 and 2004, and corn was planted in 2001, 2003, and 2005. We used daily averaged eddy covariance measurements of NEE and derived GPP in our model calibration procedure, which are categorized as Level 4 data published on the AmeriFlux site, gap filled by using the Marginal Distribution Sampling procedure outlined by [Reichstein et al. \(2005\)](#). GPP is derived as the difference between ecosystem respiration and NEE, where ecosystem respiration is estimated by using the method of [Reichstein et al. \(2005\)](#). In addition, biomass information (which we convert to carbon assuming half of the dry biomass is carbon) and LAI have been collected for years 2001–2005 for the various plant segments, including leaf (LEAFC), stem (STEMC), and grain (GRAINC), which are reported on the AmeriFlux website (<http://public.ornl.gov/ameriflux>). The frequency of biomass measurements is generally every seven days, beginning a few weeks after planting and continuing through the harvest. We chose to calibrate against the Bondville AmeriFlux site because of the availability of unique biomass data

collected. By performing the calibration against site data that includes crop rotation, we hope to indirectly include the effects of crop rotation on GPP and NEE in the model. Finally, in order to assess the transferability of the calibrated parameters across sites, we perform one more validation experiment using observations from the Mead, NE, AmeriFlux site, located at 41.1741 N, -96.4396 W, is similar to the Bondville, IL site growing a corn-soybean rotation under no-till conditions. Although there are three fields, only one is under rainfed conditions. The site was initialized in 2001, a description can be found in (Verma et al., 2005). The data collected is the same as the Bondville, IL Ameriflux site. Soybeans were planted in 2002 and 2004.

The time-dependent observations are denoted by $z(t) = \{z_1(t), \dots, z_6(t)\}$, where the indices correspond to GPP, NEE, GRAINC, LEAFC, STEMC, and TLAI. Because of uncertainties in fertilization use and measured data, we focused on the peak observed values, as well as the growth slope for GPP, NEE, LEAFC, and STEMC. To remove the atmospheric induced noise in the NEE and GPP measurements we filtered the time series by applying a moving average operator with a width of 30 days. These operations are denoted by the map

$$y = [y_1, \dots, y_{10}]^T = [\max(\bar{z}_1(t)), \text{slope}(\bar{z}_1(t)), \max(\text{abs}(\bar{z}_2(t))), \text{slope}(\bar{z}_2(t)), \max(z_3(t), \max z_4(t)), \text{slope} z_4(t), \max(z_5(t), \text{slope}(z_5(t)), \max(z_6(t))]^T, \quad (3)$$

where \bar{z}_i represents the filtered z_i and the slope is calculated in the beginning of the plant emergence phase, resulting in one maximum and one slope per variable per year. The observed GPP and NEE slopes were computed as the slope between the 208th day and 188th day for 2002 and between the 180th day and 160th day for 2004. The observed LEAFC and STEMC slopes were computed based on observed values on 7/16-8/13 and 7/23-9/10 for 2002, and on 6/8-7/27 and 6/8-8/10 for 2004, respectively.

2.3 Initial conditions and spin-up

CLM requires a spin-up to obtain balanced soil carbon and nitrogen pools, which are responsible for driving decomposition and turnover. A global spin-up of the model is provided with the model, using the below-ground biogeochemistry and spin-up method provided by Koven et al. (2013). Crops are then interpolated to a higher resolution over the Bondville, IL, site.

The meteorological forcing data used for the calibration procedure (post spin-up) is from the Bondville, IL flux tower site. The atmospheric data covers the years 1996-2007, but we focus on 2002 and 2004 for this experiment. The model is run in point mode, meaning only one grid cell is simulated at a resolution of roughly $0.1^\circ \times 0.1^\circ$.

3 Calibration strategy

We represent the CLM-Crop model output relevant to Eq. (3) by $f(\theta) = (f_1(\theta), \dots, f_q(\theta))$, where $\theta = (\theta_1, \dots, \theta_d)$ are the d time-independent parameters that we wish to calibrate and $q = 10$ is the number of outputs. The slopes estimated from numerical simulations were computed as the variable slopes between the date when the fraction of growing degree days to maturity reaches 0.3 and 20 days prior to this point ahead of it, where growing degree days are accumulated each day by subtracting the minimum temperature for growth (10°C for soybean) from the average daily temperature; see (Oleson et al., 2013).

We consider a set of ten calibration parameters that were indicated by the model as being highly uncertain. This set consists of: xl, slatop, leafcn, frootcn, livewdcn, grperc, grpnw, graincn, fleafcn, and fstemcn. See Table 1 and Sect. 2.1 for details.

The model calibration strategy aims to merge model predictions that depend on parameters θ with observational datasets. We assume that the relationship between observation data and the true process follows a relationship of type

$$Y = f(\theta^*) + \varepsilon, \quad (4)$$

where θ^* are the perfectly calibrated parameters and ε represents the observational errors. This holds under the assumption that the model is a perfect representation of reality (Kennedy and O'Hagan, 2001). The problem statement can be extended to account for imperfect models, but then the statistical description of ε tends to become much more complicated. Therefore, for this study we start by considering a perfect model assumption.

Following a Bayesian approach, we assume a prior distribution on the calibration parameters:

$$p(\theta) \propto C(\theta) \prod_{i=1}^d 1_{[\theta_{i,\min}, \theta_{i,\max}]}(\theta_i), \quad (5)$$

where $\theta_{i,\min}$ and $\theta_{i,\max}$ are the minimum and maximum allowed values for the parameter θ_i , respectively; $1_A(x)$ is the indicator function of a set A (i.e., $1_A(x)$ is one if $x \in A$ and zero otherwise); and $C(\theta)$ models any physical constraints that are known a priori. For the parameters we are considering, the constraints are as follows:

$$C(\theta) = 1_{[\theta_{\text{leafcn}} > \theta_{\text{leafcn}}]}(\theta) 1_{[\theta_{\text{fstemcn}} > \theta_{\text{livewdcn}}]}(\theta). \quad (6)$$

We define the likelihood as

$$p(y|\theta) \propto \mathcal{N}(y|f(\theta), \Sigma_{\text{obs}}), \quad (7)$$

where $\mathcal{N}(x|\mu, \Sigma)$ is the Gaussian probability density with mean μ and covariance matrix Σ . The covariance matrix Σ_{obs} is taken to be diagonal, namely, $\Sigma_{\text{obs}} = \text{diag}(\sigma_{\text{obs},1}^2, \dots, \sigma_{\text{obs},q}^2)$, with each diagonal component $\sigma_{\text{obs},i}^2$

being the square of 10% of the corresponding observed value. This choice of Σ_{obs} is equivalent to a priori assuming 10% observational noise.

Our state of knowledge about the parameters θ after observing y (see Sec. 2.2) is captured by the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (8)$$

4 Approximating the posterior

We are going to construct a *particle approximation* of Eq. (8) $\{(w^{(i)}, \theta^{(i)})\}_{i=1}^N$, in the sense that

$$p(\theta|y) \approx \sum_{i=1}^N w^{(i)} \delta(\theta - \theta^{(i)}), \quad (9)$$

where $\sum_{i=1}^N w^{(i)} = 1$, and $\delta(\cdot)$ is Dirac's delta function. This is achieved by using a combination of MCMC (Metropolis et al., 1953; Hastings, 1970) and SMC methodologies (Doucet et al., 2001). For more details on the methodological aspects, we refer the reader to the work of Del Moral et al. (2006); Koutsourelakis (2009) and Bilionis and Koutsourelakis (2012). Here we present the material briefly, focusing only on the novel aspect of our approach that concerns automatically tuning the MCMC proposals.

Let us define a sequence of *bridging distributions*:

$$p(\theta|y, \gamma_t) \propto p(y|\theta)^{\gamma_t} p(\theta) =: \pi_t(\theta), \quad (10)$$

where $0 = \gamma_0 < \gamma_1 < \dots < \gamma_t < \dots \leq 1$. Notice that for $\gamma_t = 0$ we obtain the prior and for $\gamma_t = 1$ the posterior. The key idea of SMC is to start from a particle representation of the prior ($\gamma_t = 0$), which is easy to obtain, and gradually increase γ_t until it reaches 1, adjusting the weights along the way. We will show later how this sequence can be determined on the fly by taking into account the degeneracy of the particle representations.

4.1 Sequential importance sampling

Let $\{(w_t^{(i)}, \theta_t^{(i)})\}_{i=1}^N$ be a particle representation of $p(\theta_t|y, \gamma_t)$,

$$p(\theta_t|y, \gamma_t) \approx \sum_{i=1}^N w_t^{(i)} \delta(\theta - \theta_t^{(i)}), \quad (11)$$

with the weights being normalized (i.e., $\sum_{i=1}^N w_t^{(i)} = 1$). We now examine how this particle representation can be updated to a particle representation corresponding to $\gamma_{t+1} > \gamma_t$. Toward this goal, we introduce a fictitious probability density on the joint space of θ_t and θ_{t+1} by

$$q_t(\theta_t, \theta_{t+1}) = p(\theta_{t+1}|y, \gamma_{t+1}) L_t(\theta_t|\theta_{t+1}) \propto \pi_{t+1}(\theta_{t+1}) L_t(\theta_t|\theta_{t+1}), \quad (12)$$

where L_t is a *backward* transition density (i.e., $L_t(\theta_t|\theta_{t+1})$ is the probability of θ_t given θ_{t+1}) properly normalized, that is, $\int L_t(\theta_t|\theta_{t+1}) d\theta_t = 1$. In addition, we introduce an *importance sampling* density,

$$\eta_t(\theta_t, \theta_{t+1}) = p(\theta_t|y, \gamma_t) K_t(\theta_{t+1}|\theta_t) \propto \pi_t(\theta_t) K_t(\theta_{t+1}|\theta_t), \quad (13)$$

where K_t is a *forward* transition density (i.e., $K(\theta_{t+1}|\theta_t)$ is the probability of θ_{t+1} given θ_t) properly normalized, that is, $\int K(\theta_{t+1}|\theta_t) d\theta_t = 1$. Notice that

$$\begin{aligned} p(\theta_{t+1}|y, \gamma_t) &= \int p(\theta_{t+1}|y, \gamma_{t+1}) L(\theta_t|\theta_{t+1}) d\theta_t \\ &= \int q_t(\theta_t, \theta_{t+1}) d\theta_t \\ &= \int \frac{q_t(\theta_t, \theta_{t+1})}{\eta_t(\theta_t, \theta_{t+1})} \eta_t(\theta_t, \theta_{t+1}) d\theta_t \\ &= \int \frac{q_t(\theta_t, \theta_{t+1})}{\eta_t(\theta_t, \theta_{t+1})} p(\theta_t|y, \gamma_t) K_t(\theta_{t+1}|\theta_t) d\theta_t \\ &\approx \sum_{i=1}^N w_t^{(i)} \frac{q_t(\theta_t^{(i)}, \theta_{t+1})}{\eta_t(\theta_t^{(i)}, \theta_{t+1})} K_t(\theta_{t+1}|\theta_t^{(i)}). \end{aligned}$$

This observation immediately suggests that to move the γ_t particle representation of Eq. (11) to a γ_{t+1} representation $\{(w_{t+1}^{(i)}, \theta_{t+1}^{(i)})\}_{i=1}^N$,

$$p(\theta_{t+1}|y, \gamma_{t+1}) \approx \sum_{i=1}^N w_{t+1}^{(i)} \delta(\theta - \theta_{t+1}^{(i)}),$$

with $\sum_{i=1}^N w_{t+1}^{(i)} = 1$, we must sample

$$\theta_{t+1}^{(i)} \sim K_t(\theta_{t+1}|\theta_t^{(i)}), \quad (14)$$

compute the *incremental weights*,

$$\hat{w}_{t+1}^{(i)} = \frac{\pi_{t+1}(\theta_{t+1}^{(i)}) L_t(\theta_t^{(i)}|\theta_{t+1}^{(i)})}{\pi_t(\theta_t^{(i)}) K_t(\theta_{t+1}^{(i)}|\theta_t^{(i)})} \left(\propto \frac{q_t(\theta_t^{(i)}, \theta_{t+1}^{(i)})}{\eta_t(\theta_t^{(i)}, \theta_{t+1}^{(i)})} \right), \quad (15)$$

get the unnormalized γ_{t+1} -weights,

$$W_{t+1}^{(i)} = w_t^{(i)} \hat{w}_{t+1}^{(i)}, \quad (16)$$

and get the normalized γ_{t+1} -weights:

$$w_{t+1}^{(i)} = \frac{W_{t+1}^{(i)}}{\sum_{j=1}^N W_{t+1}^{(j)}}. \quad (17)$$

4.2 Convenient choices for L_t and K_t

445 The preceding remarks hold for any backward and forward transition densities L_t and K_t , respectively. We now seek a convenient choice that will simplify the form of the incremental weights given in Eq. (15). Suppose for the **moment moment** that K_t is given and let us look for the optimal choice of L_t . Since q_t is the target distribution and η_t is the importance sampling density, the best choice of L_t is the one that attempts to bring the two densities as close together as possible. This is easily seen to be the conditional of η_t on θ_t , in other words, the optimal choice is

$$450 \quad L_t^*(\theta_t|\theta_{t+1}) = \frac{\eta(\theta_t, \theta_{t+1})}{\int \eta(\theta'_t, \theta_{t+1}) d\theta'_t} = \frac{\pi_t(\theta_t) K_t(\theta_{t+1}|\theta_t)}{\int \pi_t(\theta'_t) K_t(\theta_{t+1}|\theta'_t) d\theta'_t}. \quad (18)$$

From a computational point of view, however, it is more convenient to work with the *suboptimal* choice,

$$L_t^{*,s}(\theta_t|\theta_{t+1}) = \frac{\pi_{t+1}(\theta_t) K_t(\theta_{t+1}|\theta_t)}{\int \pi_{t+1}(\theta'_t) K_t(\theta_{t+1}|\theta'_t) d\theta'_t}, \quad (19)$$

460 which is motivated by the expectation that consecutive densities are similar (i.e. $\pi_t \approx \pi_{t+1}$). For this choice the incremental weights of Eq. (15) become

$$\hat{w}_{t+1}^{(i)} = \frac{\pi_{t+1}(\theta_{t+1}^{(i)}) \pi_{t+1}(\theta_t^{(i)})}{\pi_t(\theta_t^{(i)}) \int \pi_{t+1}(\theta'_t) K_t(\theta_{t+1}^{(i)}|\theta'_t) d\theta'_t}. \quad (20)$$

To get rid of the integral in the denominator, we pick K_t to be invariant with respect to π_{t+1} :

$$465 \quad \int \pi_{t+1}(\theta'_t) K_t(\theta_{t+1}|\theta'_t) d\theta'_t = \pi_{t+1}(\theta_{t+1}). \quad (21)$$

This can always be achieved with a suitable choice of a Metropolis-Hastings transition kernel (see below). For this case, the incremental weights simplify to

$$470 \quad \hat{w}_{t+1}^{(i)} = \frac{\pi_{t+1}(\theta_t^{(i)})}{\pi_t(\theta_t^{(i)})} = p(y|\theta_t^{(i)})^{\gamma_{t+1}-\gamma_t}. \quad (22)$$

4.3 Metropolis-Hastings-based K_t .

As shown in the previous paragraph, it is convenient to select K_t to be invariant with respect to π_{t+1} . The easiest way to achieve this is to associate K_t with one or more steps of the Metropolis-Hastings algorithm. Let $h_t(\theta'|\theta)$ be any proposal density (e.g., a simple random walk proposal). The single-step Metropolis-Hastings forward transition density is

$$K_t^1(\theta_{t+1}|\theta_t) = h_t(\theta_{t+1}|\theta_t) a(\theta_{t+1}, \theta_t), \quad (23)$$

where

$$a(\theta_{t+1}, \theta_t) := \min \left\{ 1, \frac{\pi_{t+1}(\theta_{t+1}) h_t(\theta_t|\theta_{t+1})}{\pi_{t+1}(\theta_{t+1}) h_t(\theta_{t+1}|\theta_t)} \right\}. \quad (24)$$

480 Samples from Eq. (14) may be obtained by performing one step of the well-known Metropolis-Hastings algorithm. The forward kernel corresponding to $M > 1$ Metropolis-Hastings steps is given recursively by

$$K_t^M(\theta_{t+1}|\theta_t) = \int K_t^{M-1}(\theta_{t+1}|\theta') K_t^1(\theta'|\theta_t) d\theta'. \quad (25)$$

The number of Metropolis-Hastings steps, M , at each γ_t is a parameter of SMC. This is the forward kernel we use in all numerical examples. Theoretically, $M = 1$ is enough, since the number of particles $N \rightarrow \infty$. Therefore, we will use $M = 1$ in our numerical examples.

4.4 Resampling

As SMC moves to higher values of γ_t , some of the particles might find themselves in low probability regions. Consequently, their **corresponding correspondings** weights will be small. This *degeneracy* of the weights can be characterized by the **effective effectice** sample size (ESS) metric, defined by

$$ESS(\gamma_t) = \frac{1}{\sum_{i=1}^N \left(w_t^{(i)} \right)^2}. \quad (26)$$

Notice Notice that ESS is equal to N when the particles are equally important (i.e., $w_t^{(i)} = 1/N$) and equal to 1 when only one particle is important (e.g., $w_t^{(1)} = 1$ and $w_t^{(i)} = 0$ for $i \neq 1$) and in general takes values between 1 and N for arbitrary weights. Resampling is triggered when the ESS falls below a prespecified threshold ($\frac{1}{2}N$ in our numerical examples). The idea is to kill particles that have very small weights and let the particles with big weights replicate. This process must happen in a way that the resulting particle ensemble remains a valid representation of the current target probability density. This can be achieved in various ways. Perhaps the most straightforward way is to use multinomial sampling.

490 Let the resulting particles be denoted by $\left\{ \left(\tilde{w}_t^{(i)}, \tilde{\theta}_t^{(i)} \right) \right\}_{i=1}^N$. In multinomial resampling the final weights are all equal:

$$\tilde{w}_t^{(i)} = 1/N. \quad (27)$$

The sequence $\left\{ \tilde{\theta}_t^{(i)} \right\}_{i=1}^N$ is found by sampling a sequence of integers $j_i \subset \{1, \dots, N\}$ with probabilities $\{w_t^{(1)}, \dots, w_t^{(N)}\}$ and by setting

$$\tilde{\theta}_t^{(i)} = \theta_t^{(j_i)}, \quad (28)$$

for $i = 1, \dots, N$.

4.5 Choosing γ_{t+1} on the fly

We note that the incremental weights of Eq. (22) do not depend on the γ_{t+1} -samples obtained in Eq. (14). They depend

only on the likelihood of the γ_t -samples. In this part, we exploit this observation in order to devise an effective way of selecting γ_{t+1} based on the ESS. The idea is to pick the new γ_{t+1} so that the resulting particles do not become too degenerate. Their degeneracy is characterized by $\text{ESS}(\gamma_{t+1})$ given in Eq. (26). From Eqs. (22), (16), and (17), evaluation of $\text{ESS}(\gamma_{t+1})$ does not require any new likelihood evaluations. We select the new γ_{t+1} by requiring that

$$\text{ESS}(\gamma_{t+1}) = \zeta \text{ESS}(\gamma_t), \quad (29)$$

where ζ is the percentage of the degeneracy we are willing to accept ($\zeta = 0.99$ in our numerical examples). It is fairly easy to show that $\text{ESS}(\gamma_{t+1})$ is a strictly decreasing function for $\gamma_{t+1} \in (\gamma_t, 1]$. Therefore, Eq. (29) has a unique solution that can be easily found by using a bisection algorithm.

4.6 Adapting the K_t^M on the fly

Based on the discussion above, we expect that $p(\theta_t|y, \gamma_t)$ should be similar to $p(\theta_{t+1}|y, \gamma_{t+1})$. To exploit this fact, we pick the proposal $h_t(\theta'|\theta)$ required by the Metropolis-Hastings kernel K_t^M given in Eq. (25) to be a mixture of Gaussians that approximates $p(\theta_t|y, \gamma_t)$. In particular, we pick

$$h_t(\theta'|\theta) = \sum_{i=1}^L c_i \mathcal{N}(\theta'|\mu_{t,i}, \Sigma_{t,i}), \quad (30)$$

where the non-negative coefficients c_i sum to one, $\mu_{i,i} \in \mathbb{R}^d$, and $\Sigma_{t,i} \in \mathbb{R}^{d \times d}$ are covariance matrices. The number of components L as well as all the parameters of the mixture fitted to a resampled version of the particle approximation of $p(\theta_t|y, \gamma_t)$ (see Eq. (28)) using the procedure of Blei and Jordan (2005) as implemented by Pedregosa et al. (2011).

4.7 Parallelization

SMC is embarrassingly parallelizable. Basically, each CPU can store and work with a single particle. Communication is required only for normalizing the weights (see Eq. (17)), finding γ_{t+1} (see Eq. (29)), and resampling. The first two have a negligible communication overhead and can be implemented easily. Implementation of the resampling step is more involved and requires more resources. However, the cost of resampling is negligible compared with the evaluation of the forward model.

4.8 The final algorithm

We now collect all the details of SMC discussed above in a single algorithm for convenience; Algorithm 1. Our implementation is in Python and is provided at: <https://github.com/ebilonis/pysmc>.

5 Results

In this section we present our calibration results for the parameters described in Sect. 2.1 by using the observations detailed in Sect. 2.2. In this study we focus only on the parameters affecting the soy crop and restrict our calibration to year 2004. With these calibrated parameters we perform a validation experiment by using the data from year 2002. In addition, we forecast 2004 through a 2002-2003-2004 simulation. We recognize that the Bondville observations include crop rotation during 2003 that will influence the sequence of output; but since the model does not support crop rotation, we plant soybean during 2003. The role of the latter experiment is to demonstrate the robustness of the proposed calibration scheme. Moreover, we perform a twin experiment that consists of generating artificial data by using some control parameter values, then applying the calibration strategy to recover the control parameters.

In all our numerical examples we fix the planting and harvest days. This approach is essential in order to avoid overfitting the physiological parameters due to offsets in the growing seasons. The planting dates for 2002 and 2004 are 06/02 and 05/07, respectively. The harvest day is controlled via the input variable “hybgdd” (growing degree days for maturity), where growing degree days are defined in Sec. 3. The values of “hybgdd” that give the right harvest days for 2002 and 2004 are 1474 and 1293, respectively.

The number of particles we use is $N = 1280$. Each particle is assigned to a different computational core, i.e., we use 1280 computational cores. A simulated year takes about 2 min to complete if data localization is used. Calibration requires approximately 100,000 simulations and completes in about 6 hours.

5.1 Validation of the method

We begin the twin experiment with the aim of validating the proposed calibration strategy. We generate artificial observations by randomly sampling θ from its prior Eq. (5). We apply the calibration strategy on the artificial observations to see whether the method can recover the ground truth. The adaptively selected γ_t sequence is shown in Fig. 1(a). In Fig. 2 we compare the posterior of each parameter with the prior. The true parameters are indicated by red dots. The fit to the artificial outputs is shown in Fig. 3. The parameters that are not specified precisely are parameters that have a small (if any) effect on the observed outputs.

5.2 Calibration using real data

In our next experiment we calibrate the parameters listed in Table 1. The observational operator (Eq. 3) is defined by taking the annual maximum of the absolute value of LEAFC, LAI, GRAINC, STEMC, GPP, and NEE; and the slope of LEAFC, STEMC, GPP, and NEE as described in Sec. 2.2. In

Fig. 4, we compare the posterior we obtain with the prior. The default parameters are indicated by red dots. The fit to the artificial outputs is shown in Figures 5. The adaptively selected γ_t sequence is shown in Fig. 1(b). In Table 2 we summarize our findings, by showing the median and the $p = 0.05$ and $p = 0.95$ quantiles of each calibrated parameter.

5.3 Validation of real data results

To validate the generalization potential of our calibration we perform a one-way validation. We use the calibrated parameters to predict the observables in 2002. In Fig. 6 we plot the median and 95% error bars of the calibrated time series and we compare the results with observations and the default parameter output of 2002. We observe a notable improvement in the ability of the model to explain the observations. One of the most important improvements is related to LAI calculations, which comes from improvements to the leaf CN ratio and the specific leaf area. The timing of maximum LAI is important for the carbon allocation; when the crops in CLM4.5 reach peak LAI, carbon allocation shifts from above and below-ground to strictly below-ground (roots). With the default parameter values, peak LAI occurred early in the growing season, resulting in large and unrealistic allocation of carbon to roots and insufficient carbon to leaf, stem, and ultimately grain. The large increase in stem carbon and the slower rate of growth and peak of GPP are clear indications that the shift in allocation to roots no longer occurs with the new calibrated parameter values. The grain carbon is still low, however, a result of the low leaf carbon and the overestimation of stem carbon, which increases the amount of carbon allocated to maintenance respiration at the expense of new growth for this year. The increase in uncertainty is likely a result of a limitation in nitrogen availability in some scenarios. When CN ratios are low, a higher demand of nitrogen from plants contributes to an increase in competition for resources with below-ground decomposition processes. When the nitrogen demand from the two sources exceeds availability, the amount of carbon that can be assimilated is downscaled, resulting in a lower GPP, increase in NEE, and so on. We continue the simulation through 2003 to 2004 and compare the calibrated time series with the observations and the default parameter output in Fig. 7. The differences between this plot and Fig. 5 are due to differences in the below ground conditions of carbon and nitrogen that drive the dynamics for plant competition with below-ground decomposition processes. Crops in CLM4.5 tend to be sensitive to variation in carbon and nitrogen pools; and since we ran the calibration over one year and didn't consider variability in previous years' carbon and nitrogen pools, demand for nitrogen is likely different when the model is run for multiple years. Since the change in pools is minor, the resulting change in output by the model is also small. This is also likely responsible for the increased uncertainty in GPP and NEE, which occurs from competition for resources as dis-

cussed above. Finally, in Fig. 8 show the results of the same validation experiment at the Mead, NE, site.

6 Discussion

In this paper, we sought to improve CLM-Crop model performance by parameter calibration of a subset of model parameters governing, mostly, the carbon and nitrogen allocation to the plant components. By using a Bayesian approach, we were able to improve the model-simulated GPP, NEE, and carbon biomass to leaf, stem, and grain with the new parameter values. In addition, we demonstrated that the calibrated parameters are applicable across alternative years and not solely representative of one year.

This study does have a few limitations stemming from a lack of observation data. Currently our results are suitable at one site across multiple years; testing at multiple sites would give a better indication of how well the model can perform globally or even across a region. However, the limited data over agricultural sites constrains our ability to determine parameter values that are relevant at a global scale. In addition, our use of actual planting dates is not a typical approach with CLM4.5, which generally uses temperature thresholds to trigger planting. Thus, the model may plant earlier or later compared with observations, which, if significant, could influence the growth cycle and resulting carbon fluxes. In addition, CLM-Crop does not have crop rotation, which is common across agricultural landscapes, including in the observation dataset. Crop rotation can modify below-ground carbon and nitrogen cycling that would have an impact on crop productivity through nutrient availability as well as NEE. While we would like to include crop rotation, CLM does not currently have the capability to support this function. Therefore, we tried to include the effects indirectly by calibrating against data that includes crop rotation. As more sophisticated crop representation is introduced into the model, we will revisit the calibration to improve model parameters. Moreover, we considered the initial litter, carbon, and nitrogen pools fixed by the values of the prior parameters because a direct spin-up calculation would have made sampling prohibitively expensive. We will address this issue in a future study by including these pools in the calibration procedure.

Our approach has focused on one crop type, soybean, with the intent of determining the effectiveness of the proposed calibration method. We consider the results promising and, as part of future work, hope to expand this research to additional years, crop types, and other parameters. Many other variables are of interest, including fertilization rate, timing of the growth stages, and a few other parameters related to photosynthesis. As the model continues to evolve with the addition of new or improved processes, we also may need to revisit the parameter choices and evaluate their appropriateness. Moreover, a calibration procedure carried for such complex models with relatively little data and a few calibra-

720 tion parameters has the potential to lead to overfitting. To assess this effect, we performed a validation experiment, which 775 provides good confidence, albeit not proof, of a robust calibration of the parameters. Richer datasets will likely sharpen the results and enhance the confidence intervals.

725 The introduction of new datasets documenting agriculture productivity or carbon mass will also allow us to determine 780 the applicability of our new parameter values across regions. In general, the calibration results depend on an accurate specification of the observational errors. In this study we did not have access to any information regarding the measurement 785 process and, therefore, assumed a certain observational noise. These calibration results can be sharpened by annotating the observational data with levels of confidence. The calibration strategy presented in this study has the potential to improve model performance by helping modelers define 790 parameters that are not often measured or documented.

Acknowledgements. This work was supported by the Office of Biological and Environmental Research, U.S. Department of Energy, 795 under Contract DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided on “Fusion,” a 320-node computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. 800

References

- Annan, J., Hargreaves, J., Edwards, N., and Marsh, R.: Parameter estimation in an intermediate complexity Earth System Model using an ensemble Kalman filter, *Ocean Modelling*, 8, 135–154, 2005.
- Bender, F.: A note on the effect of GCM tuning on climate sensitivity, *Environ. Res. Lett.*, 3, 014 001, doi:[10.1088/1748-9326/3/1/014001](https://doi.org/10.1088/1748-9326/3/1/014001), 2008. 810
- Bilionis, I. and Koutsourelakis, P. S.: Free energy computations by minimization of Kullback–Leibler divergence: An efficient adaptive biasing potential method for sparse representations, *Journal of Computational Physics*, 231, 3849–3870, 2012.
- 755 Bilionis, I. and Zabaras, N.: Solution of inverse problems with limited forward solver evaluations: A fully Bayesian perspective, *Inverse Problems*, 30, 015 004, 2014.
- Blei, D. M. and Jordan, M. I.: Variational inference for Dirichlet process mixtures, *Bayesian Analysis*, 1, 121–144, 2005.
- 760 Del Moral, P., Doucet, A., and Jasra, A.: Sequential Monte Carlo samplers, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 411–436, 2006.
- Doucet, A., De Freitas, N., and Gordon, N., eds.: *Sequential Monte Carlo Methods in Practice (Statistics for Engineering and Information Science)*, Springer, 2001. 825
- Drewniak, B., Song, J., Prell, J., Kotamarthi, V., and Jacob, R.: Modeling agriculture in the Community Land Model, *Geoscientific Model Development*, 6, 495–515, doi:[10.5194/gmd-6-495-2013](https://doi.org/10.5194/gmd-6-495-2013), 2013.
- 770 Evensen, G.: *Data assimilation: The ensemble Kalman filter*, Springer, 2009.
- Hargreaves, J., Annan, J., Edwards, N., and Marsh, R.: An efficient climate forecasting method using an intermediate complexity Earth System Model and the ensemble Kalman filter, *Clim. Dynam.*, 23, 745–760, 2004.
- Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97–109, 1970.
- Hourdin, F., Grandpeix, J., Rio, C., Bony, S., Jam, A., Cheruy, F., Rochetin, N., Fairhead, L., Idelkadi, A., Musat, I., et al.: LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection, *Clim. Dynam.*, pp. 1–30, doi:[10.1007/s00382-012-1343-y](https://doi.org/10.1007/s00382-012-1343-y), 2012.
- Kaipio, J. and Somersalo, E.: *Statistical and Computational Inverse Problems (Applied Mathematical Sciences) (v. 160)*, Springer, 2004.
- Kennedy, M. and O’Hagan, A.: Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 425–464, 2001.
- Koutsourelakis, P. S.: A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters, *Journal of Computational Physics*, 228, 6184 – 6211, 2009.
- Koven, C., Riley, W., Subin, Z., Tang, J., Torn, M., Collins, W., Bonan, G., Lawrence, D., and Swenson, S.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, *Biogeosciences*, 10, 7109–7131, 2013.
- Kucharik, C. and Brye, K.: Integrated Biosphere Simulator (IBIS) yield and nitrate loss predictions for Wisconsin maize receiving varied amounts of nitrogen fertilizer, *Journal of Environmental Quality*, 32, 247–268, 2003.
- Lawrence, D., Oleson, K., Flanner, M., Fletcher, C., Lawrence, P., Levis, S., Swenson, S., and Bonan, G.: The CCSM4 land simulation, 1850–2005: Assessment of surface climate and new capabilities, *Journal of Climate*, 25, 2240–2260, 2012.
- Levis, S., Bonan, G., Kluzek, E., Thornton, P., Jones, A., Sacks, W., and Kucharik, C.: Interactive crop management in the Community Earth system model (CESM1): seasonal influences on land–atmosphere fluxes, *Journal of Climate*, 25, 4839–4859, doi:[10.1175/JCLI-D-11-00446.1](https://doi.org/10.1175/JCLI-D-11-00446.1), 2012.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Junglauss, J., Klocke, D., Matei, D., et al.: Tuning the climate of a global model, *Journal of Advances in Modeling Earth Systems*, 4, M00A01, 2012.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, 21, 1087–1092, 1953.
- Meyers, T. and Hollinger, S.: An assessment of storage terms in the surface energy balance of maize and soybean, *Agricultural and Forest Meteorology*, 125, 105–115, 2004.
- Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., Thornton, P., Bozbiyik, A., Fisher, R., Heald, C., Kluzek, E., Lamarque, J.-F., Lawrence, P., Leung, L., Lipscomb, W., Muszala, S., Ricciuto, D., Sacks, W., Sun, Y., Tang, J., and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), p. 420, doi:[10.5065/D6RR1W7M](https://doi.org/10.5065/D6RR1W7M), 2013.
- Pauwels, V., Verhoest, N., De Lannoy, G., Guissard, V., Lucau, C., and Defourny, P.: Optimization of a coupled hydrology–crop growth model through the assimilation of observed soil moisture

- and leaf area index values using an ensemble Kalman filter, *Water Resources Research*, 43, W04421, 2007.
- 835 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830, 2011.
- 840 Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, 11, 1424–1439, 2005.
- 850 Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, 2004.
- Verma, S. B., Dobermann, A., Cassman, K. G., Walters, D. T., 855 Knops, J. M., Arkebauer, T. J., Suyker, A. E., Burba, G. G., Amos, B., Yang, H., Ginting, D., Hubbard, K. G., Giltson, A. A., and Walter-Shea, E. A.: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, *Agricultural and Forest Meteorology*, 131, 77 – 96, doi:<http://dx.doi.org/10.1016/j.agrformet.2005.05.003>, <http://www.sciencedirect.com/science/article/pii/S0168192305000973>, 2005.
- 860 Yang, B., Qian, Y., Lin, G., Leung, R., and Zhang, Y.: Some issues in uncertainty quantification and parameter tuning: A case study of convective parameterization scheme in the WRF regional climate model, *Atmospheric Chemistry & Physics*, 12, 2409–2427, 2012.

Data: N the desired number of particles, M the desired number of Metropolis-Hastings steps per SMC step, $\zeta \in (0, 1)$ the desired effective sample size (ESS) reduction and $\epsilon \in (0, 1)$ the resampling threshold.

Result: A particle approximation $\left\{ \left(w^{(i)} = w_1^{(i)}, \theta^{(i)} = \theta_1^{(i)} \right) \right\}_{i=1}^N$ of Eq.(8).

$t \leftarrow 0$;

$\gamma_t \leftarrow 0$;

$w_0^{(i)} \leftarrow 1/N, i = 1, \dots, N$;

Sample $\theta_0^{(i)} \sim p(\theta|y, \gamma_0 = 0) = p(\theta), i = 1, \dots, N$;

while $\gamma_t < 1$ **do**

 Find $\gamma' \in (\gamma_t, 1]$ s.t.:

$$\text{ESS}(\gamma') = \frac{1}{\sum_{i=1}^N \left(w_{\gamma'}^{(i)} \right)^2} = \zeta \text{ESS}(\gamma_t), \quad (31)$$

 where $w_{\gamma'}^{(i)}$ is the normalized version of $W_{\gamma'}^{(i)} = w_{\gamma_t}^{(i)} \hat{w}_{\gamma'}^{(i)}$ with $\hat{w}_{\gamma'}^{(i)} = \frac{p(\theta_{\gamma_t}^{(i)}|y, \gamma')}{p(\theta_{\gamma_t}^{(i)}|y, \gamma_t)}$.

if $\text{ESS}(\gamma') < \epsilon N$ **then**

 | Resample according to Sec. 4.4;

end

 Adjust the MCMC proposal $h_t(\theta'|\theta)$ according to Sec. 4.6;

 Perform M MCMC steps (see Sec. 4.3);

 Adjust the next proposal $h_t(\theta'|\theta)$;

$\gamma_{t+1} \leftarrow \gamma'$;

$t \leftarrow t + 1$;

end

Algorithm 1: Sampling from the posterior using sequential Monte Carlo.

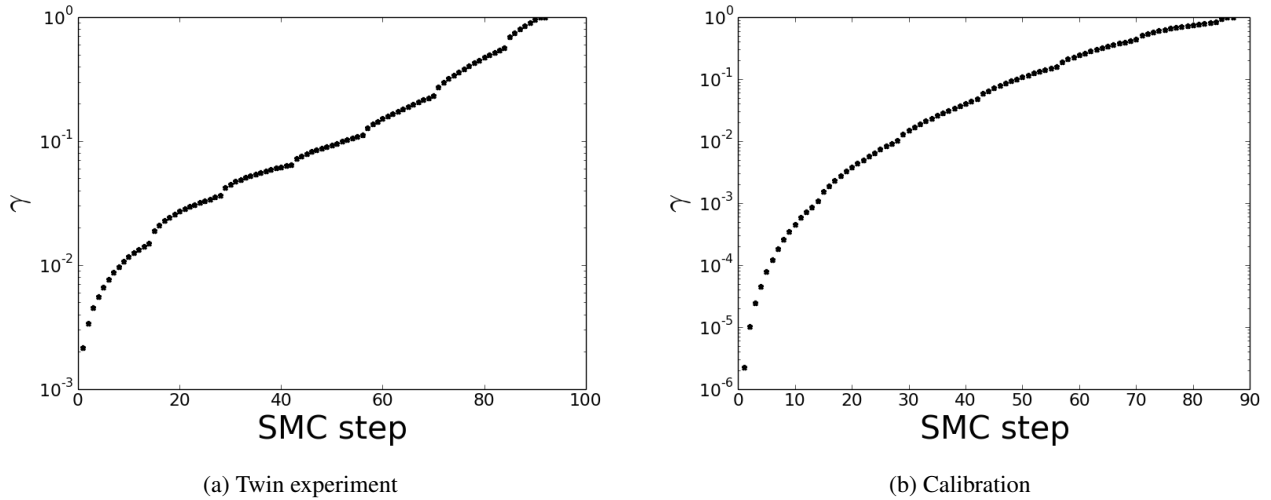


Fig. 1: Adaptation of γ_t for the twin experiment (a) and the calibration (b). The small jumps indicate the locations where resampling occurs.

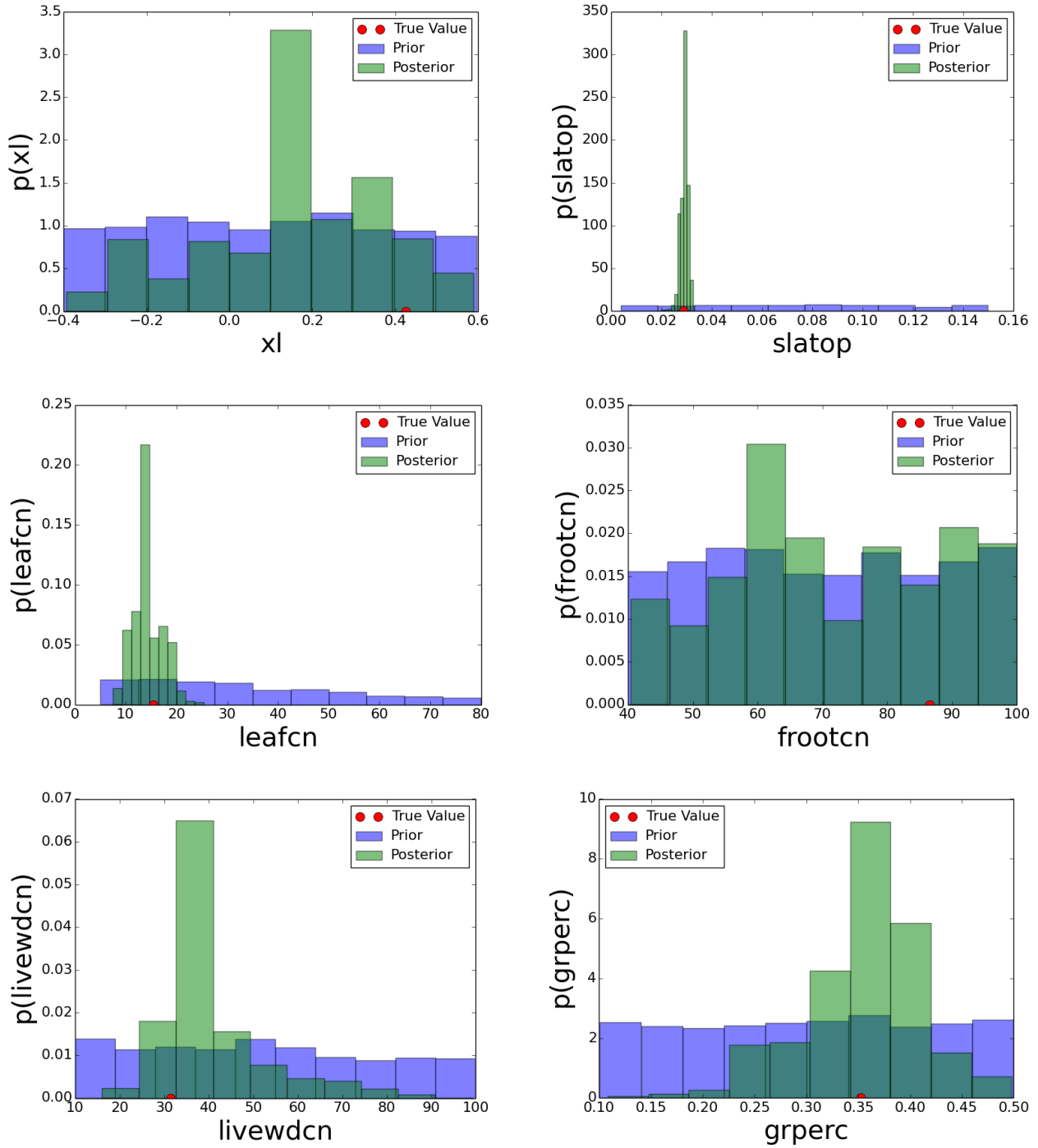


Fig. 2: Twin experiment: Comparison of the posterior with the prior. The red dot indicates the true parameter value. The figure continues on the next page.

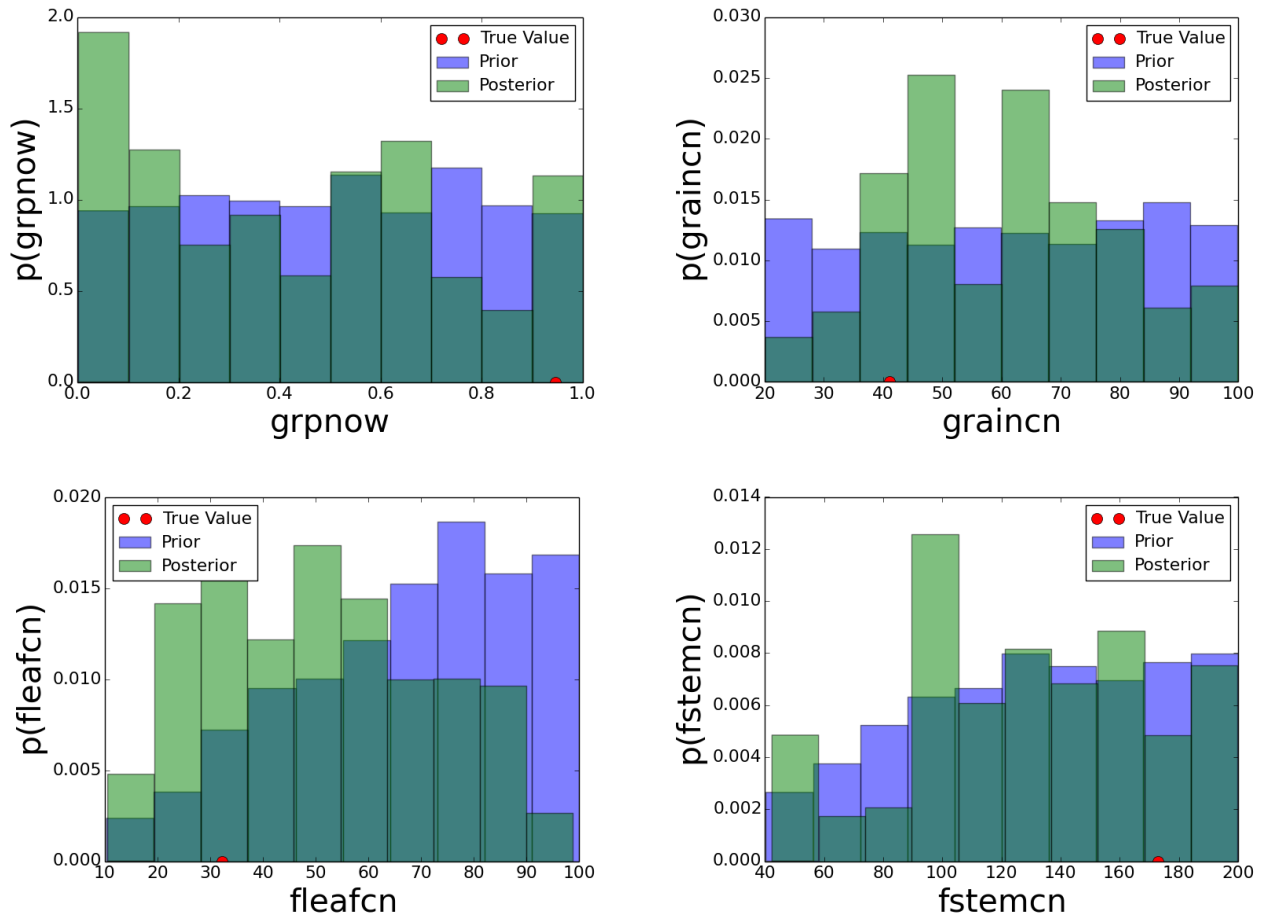


Fig. 2 continued.

Table 1: Prior information on the parameters.

Name	Description	Constraints	Min	Max	Default
xl	leaf/stem orientation index	none	-0.40	0.60	-0.40
slatop	specific leaf area at top of canopy	none	0.00	0.15	0.07
leafcn	leaf C	none	5.00	80.00	25.00
frootcn	fine root C	none	40.00	100.00	42.00
livewdcn	live wood (phloem and ray parenchyma C)	none	10.00	100.00	50.00
grperc	growth respiration factor 1	none	0.10	0.50	0.25
grpnow	growth respiration factor 2	none	0.00	1.00	1.00
graincn	organ carbon nitrogen ratio	none	20.00	100.00	50.00
fleafcn	final leaf carbon nitrogen ratio	> leafcn	10.00	100.00	65.00
fstemcn	final stem carbon nitrogen ratio	> livewdcn	40.00	200.00	130.00

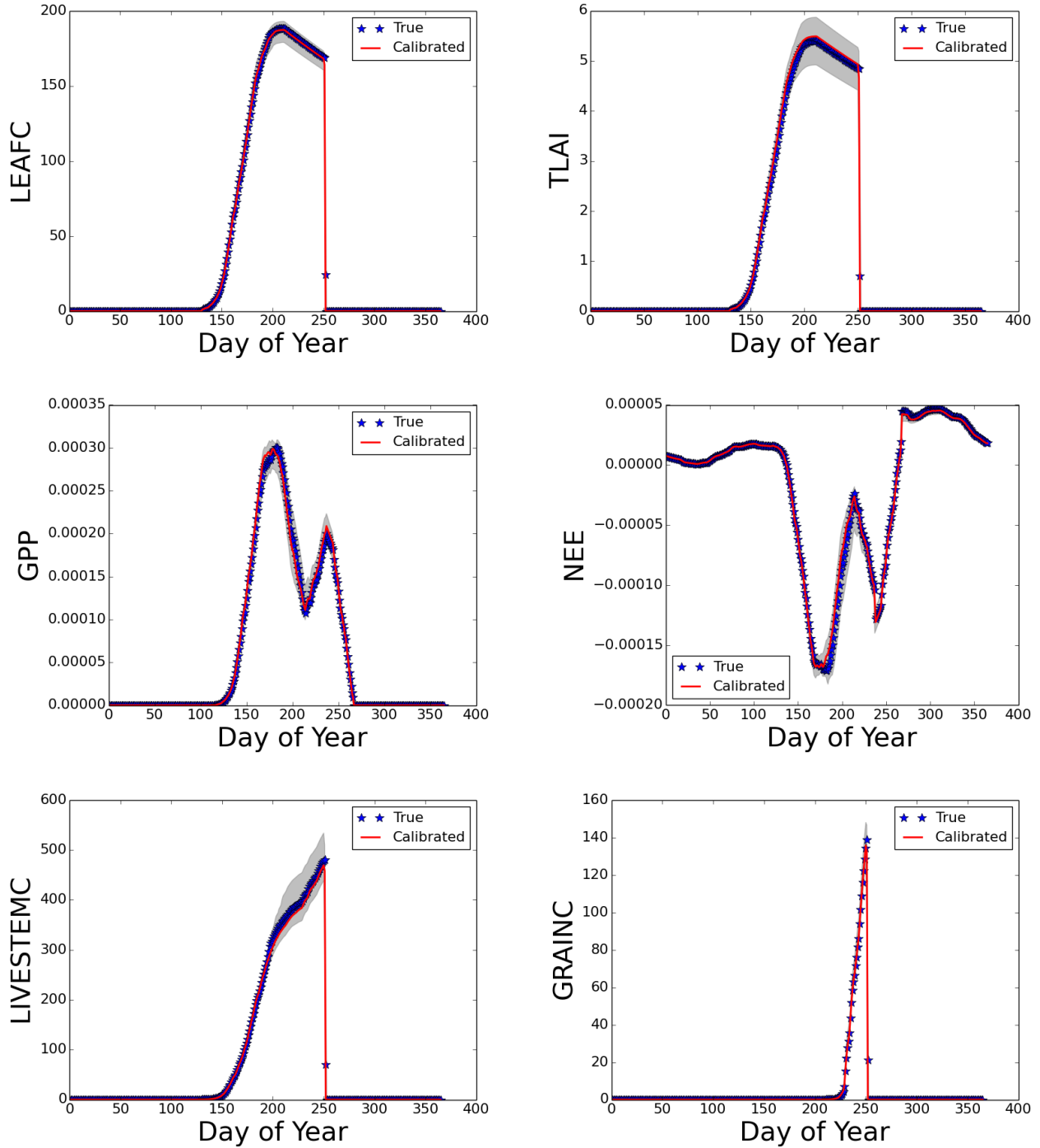


Fig. 3: Twin experiment: Comparison of the true model output with samples from the posterior of the calibration.

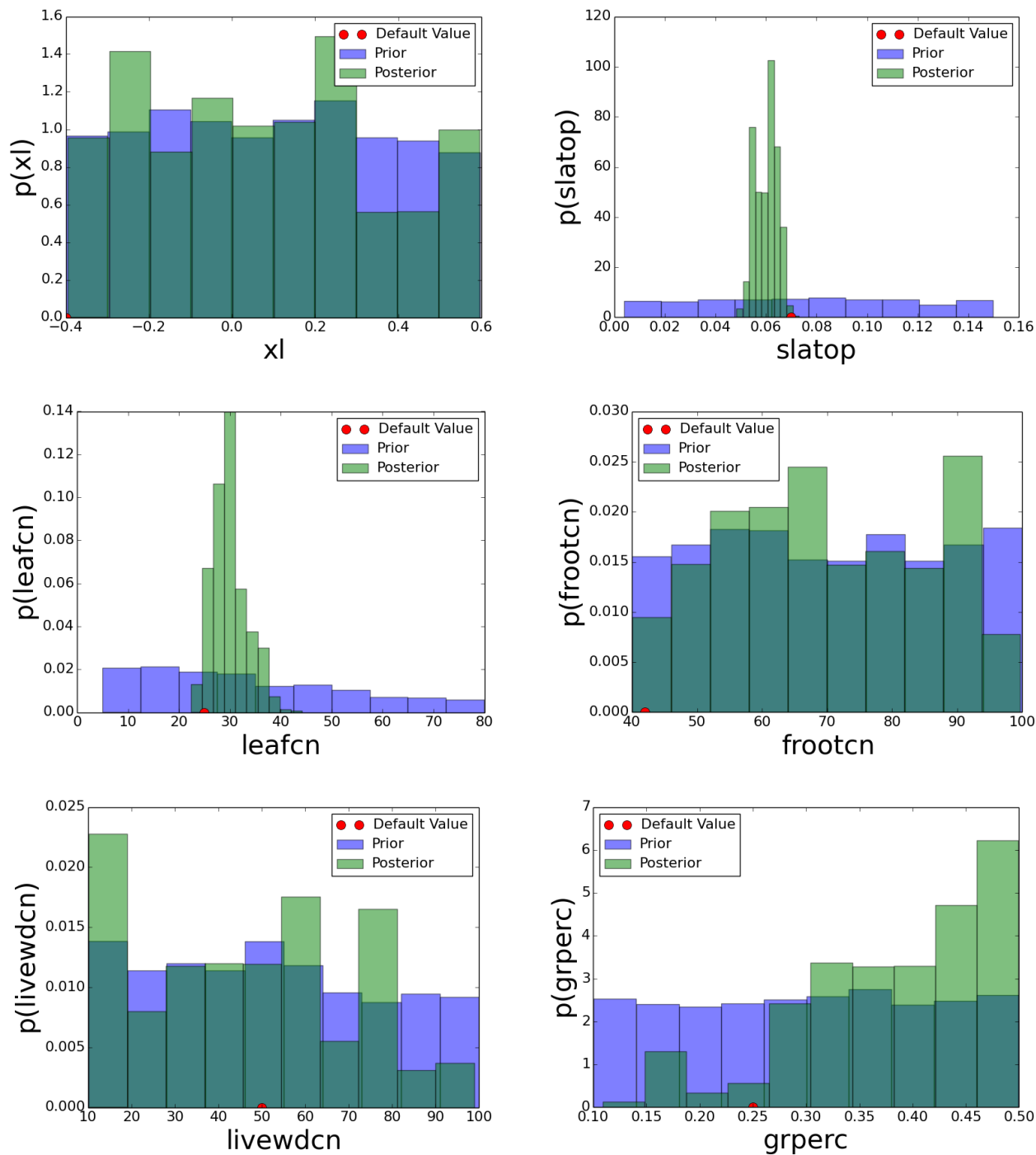


Fig. 4: Calibration experiment (Bondville, IL): Comparison of the posterior with the prior. The red dot indicates the default parameter value. The figure continues on the next page.

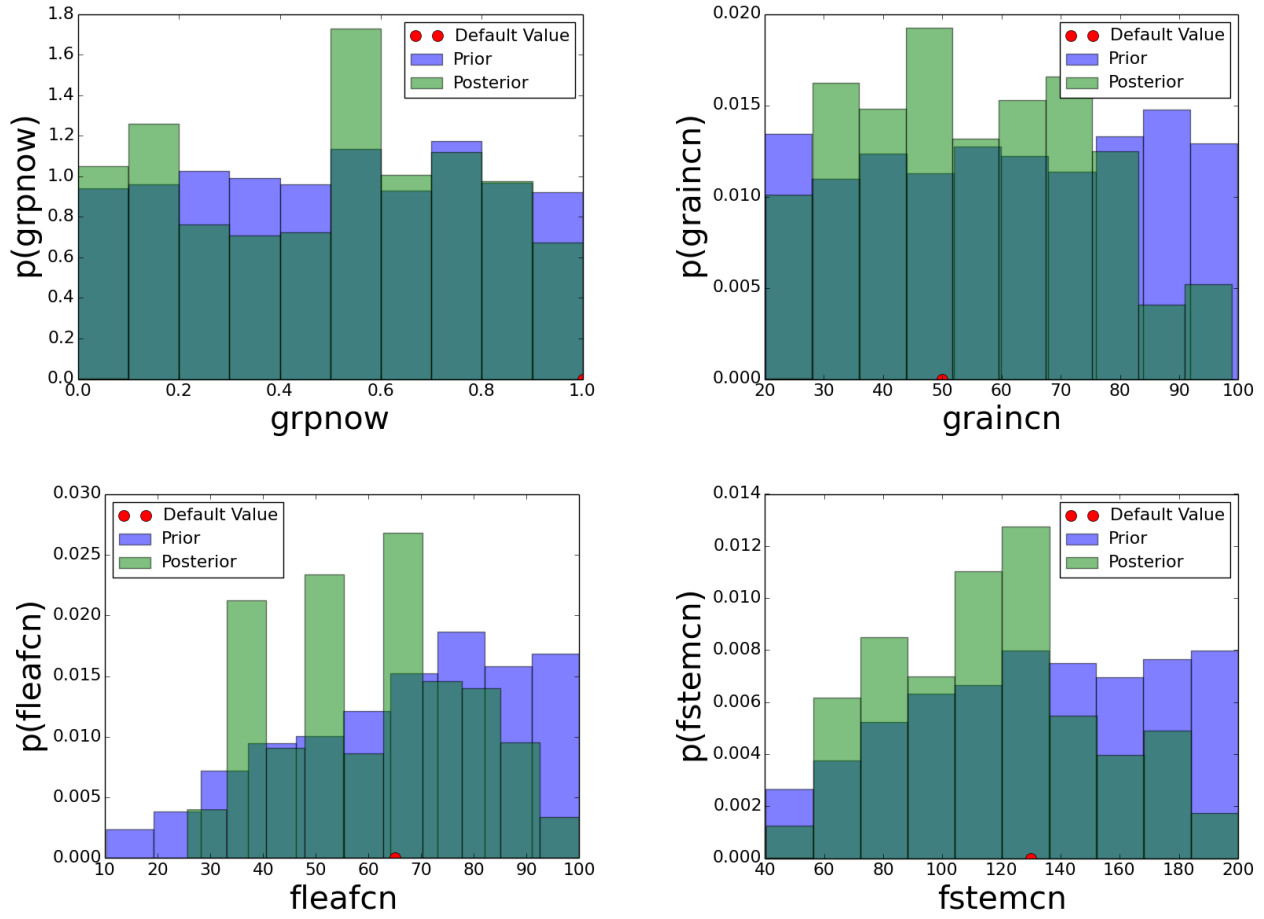


Fig. 4 continued.

Table 2: Posterior information on the parameters.

Name	Median	$p = 0.05$	$p = 0.95$
xl	0.09	-0.32	0.51
slatop	0.06	0.05	0.07
leafcn	29.51	25.26	35.79
frootcn	66.90	46.02	95.66
livewdcn	48.17	10.56	85.89
grperc	0.40	0.18	0.50
grpnw	0.54	0.05	0.91
graincn	55.69	27.18	89.14
fleafcn	64.30	33.68	89.43
fstemcn	118.02	59.47	180.88

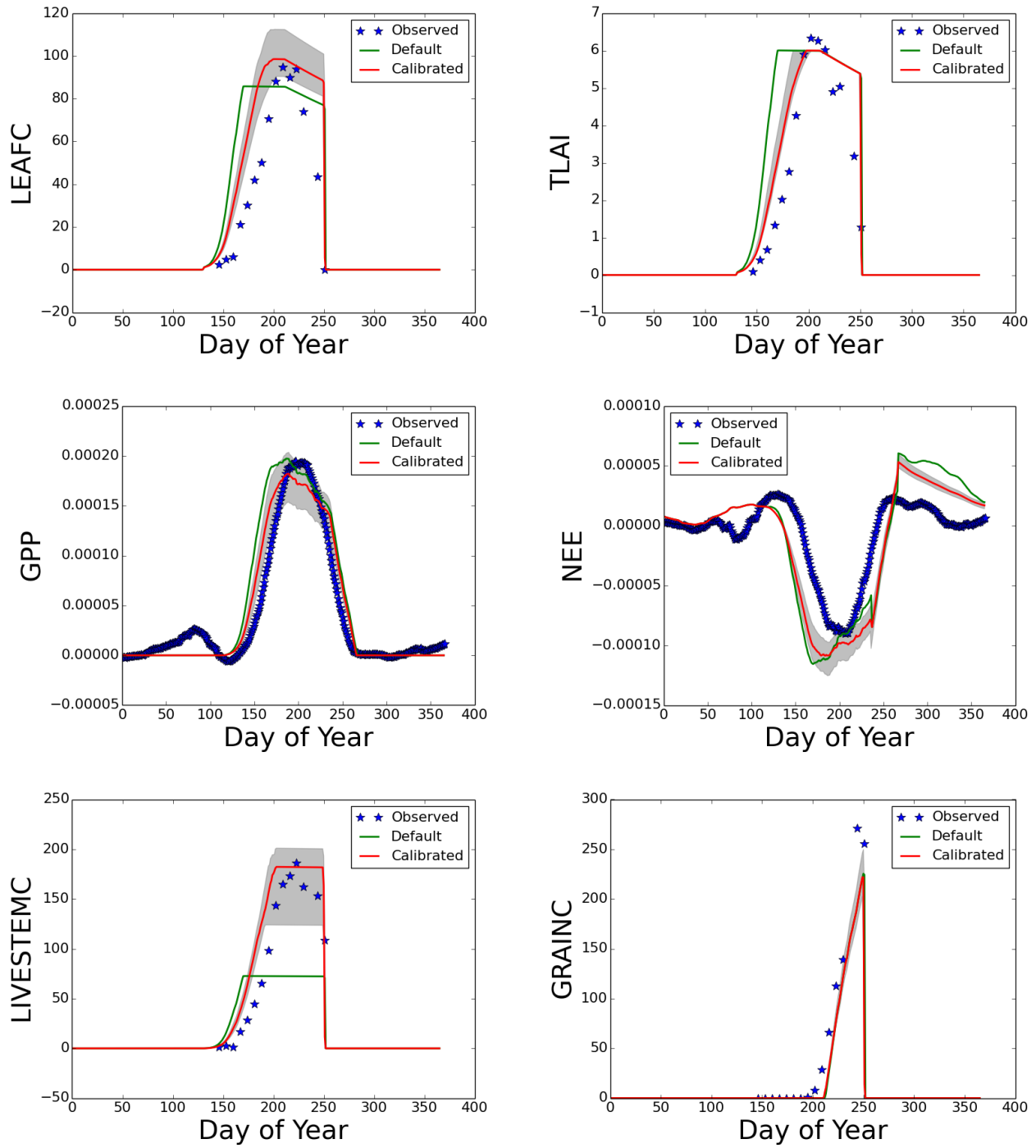


Fig. 5: Calibration experiment (Bondville, IL): Comparison of the observed data for 2004 with calibrated outputs. The grey areas correspond to 95% confidence error bars.

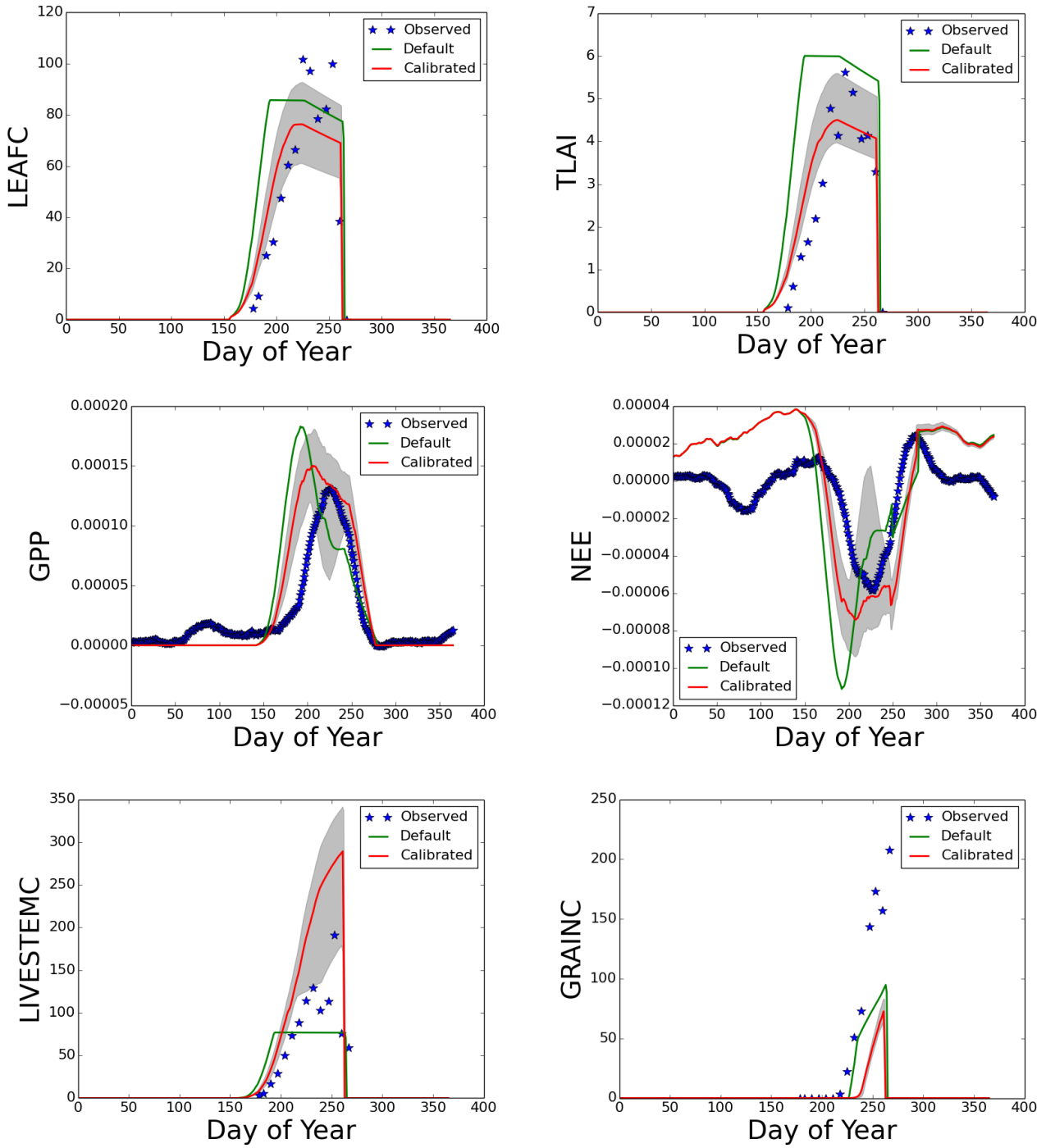


Fig. 6: Validation experiment (Bondville, IL): Comparison of the observed data for 2002 with the model. The grey areas correspond to 95% confidence error bars.

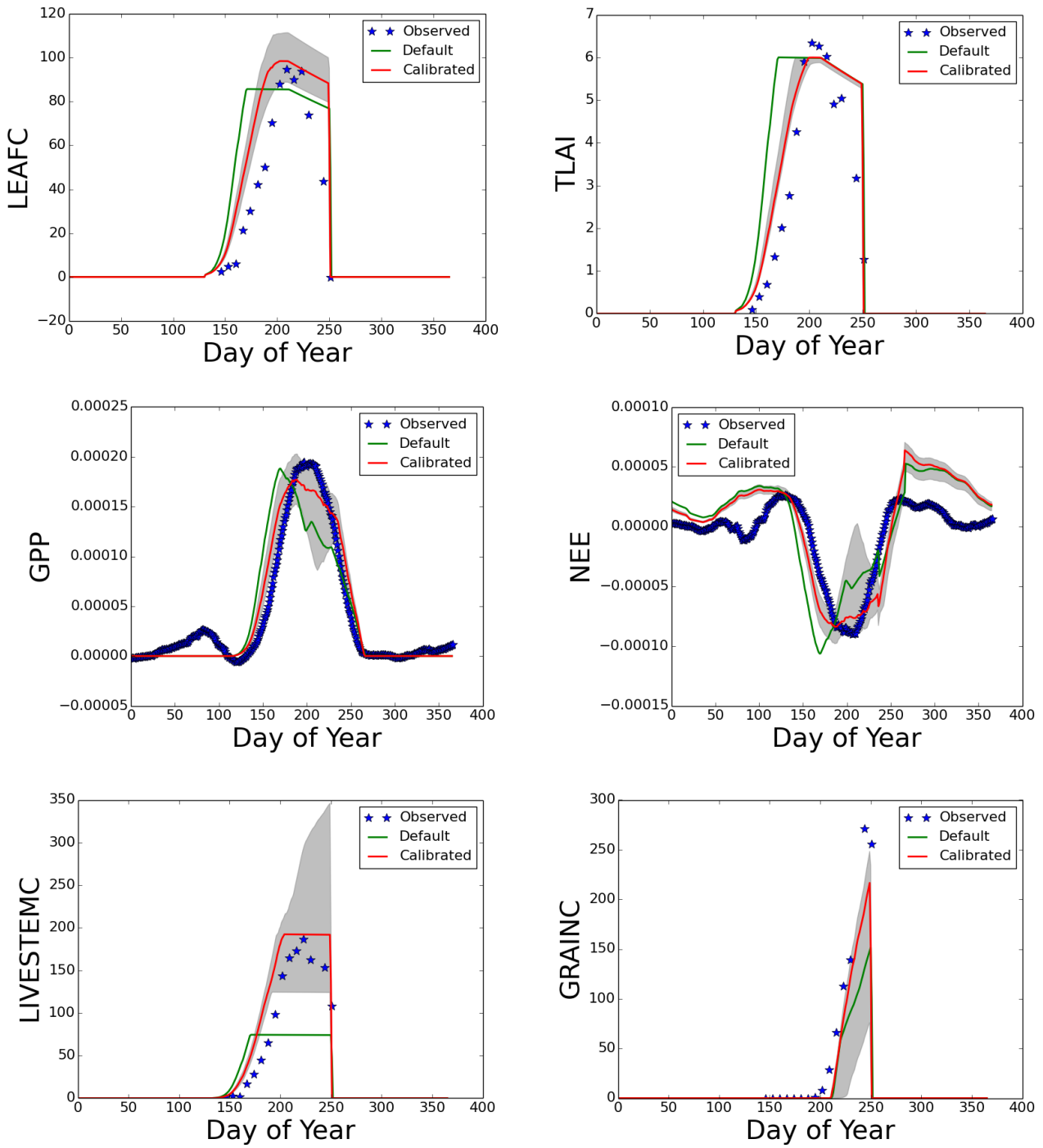


Fig. 7: Validation experiment (Bondville, IL): Comparison of the observed data for 2004 with the model started on 2002. The grey areas correspond to 95% confidence error bars.

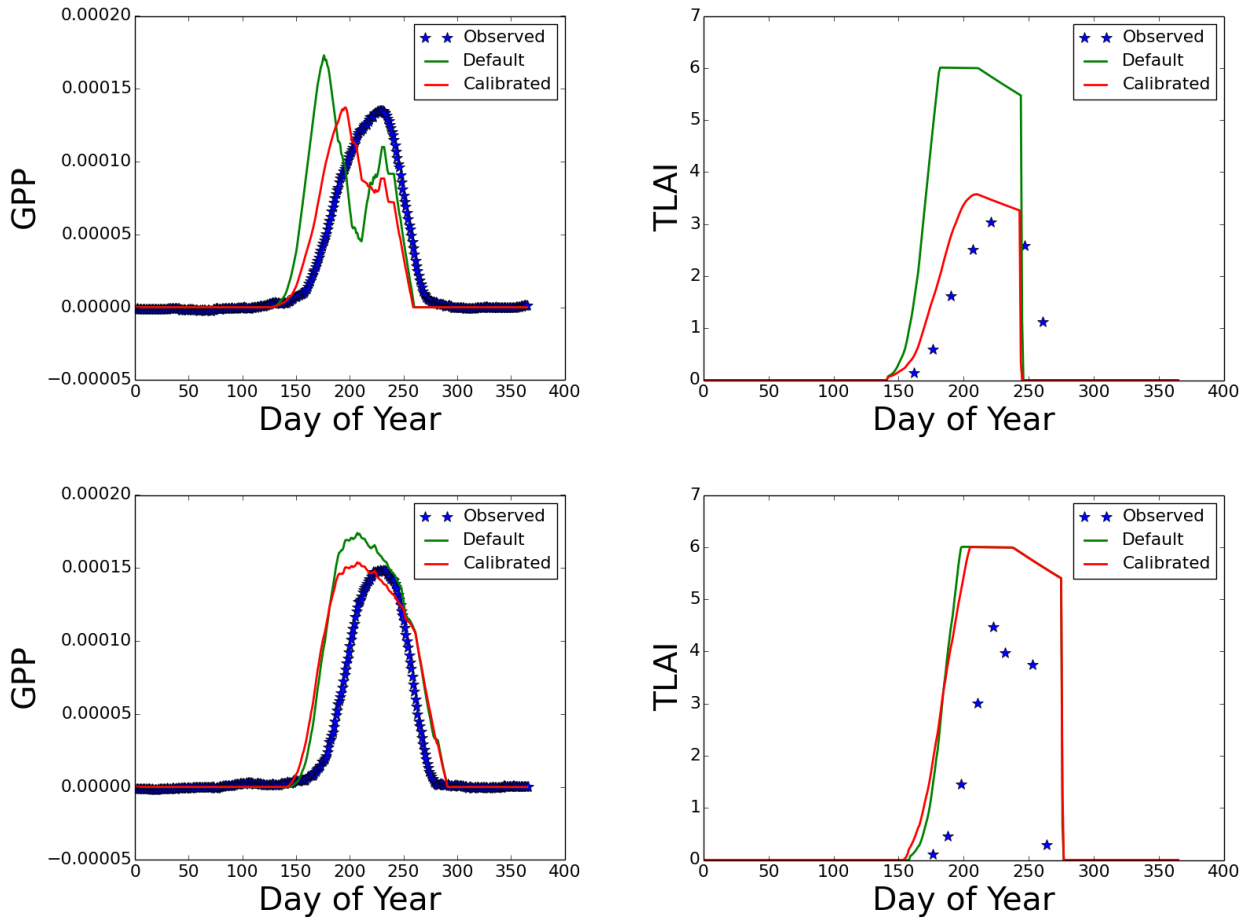


Fig. 8: Validation experiment (Mead, NE): Comparison of the observed data for 2002 with the model (top). Comparison of the observed data for 2004 with the model (bottom).

Government License The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.