

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2021-244-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-244

Anonymous Referee #1

Referee comment on "Modeling perennial bioenergy crops in the E3SM land model (ELMv2)" by Eva Sinha et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2021-244-RC1>, 2021

General comments:

This study implements two perennial bioenergy crops miscanthus and switchgrass in the E3SM land model ELM. The revised phenology, carbon/nitrogen allocation, and biomass harvest parametrizations are based on the generic grass and annual crop functional types in CLM4.5 with the main distinction that the perennials are planted once and have repeated leaf onset, senescence and harvest cycle each year with a longer growing season than annuals. The manuscript focuses on sensitivity analysis and parameter optimization using an approach combining surrogate construction with polynomial regression and Bayesian calibration with MCMC. While the sensitivity analysis and calibration methods are novel, there are some issues that need to be clarified (which I list below as three questions). The manuscript can also be substantially improved by an independent validation with data not used in the sensitivity analysis and calibration.

- Are the constructed surrogate models each representing one quantity or variable of interest (QoI)?
- From Table 2, I see that each parameter has different optimized values for different QoIs (GPP, ER, LE, H). For example, the parameter 'slatop' for Miscanthus used values 0.02, 0.06, 0.04, and 0.02 for estimating GPP, ER, LE, and H, respectively, in Fig. 3ACEG, right? Does it imply that the MCMC calibration was done for each QoI independently and that the ELM model or surrogate models were run multiple times using different parameter values to simulate the four QoIs in Fig. 3 separately?

My initial understanding is 'yes' to the above questions (correct me if I misunderstood and you can skip to 3, but please then explain what is the final value used for each parameter in Table 1 & 2 and how is it determined?)

This is the problem – Model calibration should lead to one decided value for each parameter for each bioenergy crop regardless of the QoI. A land surface model should have a determined set of parameter values to be used together to simulate all the carbon, water, and energy flux and state variables, including the QoIs here simultaneously, because they are interconnected. You can not use different (optimized) values for each parameter to simulate different target variables separately. Otherwise, you will have to conduct numerous runs for the large number of variables in ELM, which is not feasible for real applications. The calibrated results presented in Fig. 3 and Table 3 essentially come from the surrogate based posterior simulations for the respective QoIs, not from the ELM. Ideally, each parameter should be calibrated against all the observed variables (e.g., GPP, ER, LE, H) at the same time based on the average performance, so that it is applicable to other variables of interest such as LAI, carbon stocks, transpiration... in the same model.

- The manuscript lacks an independent validation step to prove the model's generalizability and applicability for cross-location, regional, or global simulations. Validation usually entails applying the set of determined model parameters for each crop from calibration to one or more new sites to evaluate the model performance on all concerned variables or, if without new sites, at least reserving some observation data for evaluating some variables that are not used in the calibration. Although the 2000 ELM simulations were split to 1600 and 400 for training and testing the surrogates, it is not the same concept as validation. Fig. 3 is merely a calibration plot (and it is partial calibration/over-fitted to the input data for each individual QoI, thus lacking applicability to the ELM model as a whole). Therefore, the new model is not sufficiently validated yet.

Before the above questions are clarified, the current model development and evaluation are immature. If I did not misunderstand, the current model description and calibration are problematic and need to be revised. Essential steps to obtain a consistent set of parameter values for the ELM model (not for each QoI) and further validation of the model using independent data are needed.

Specific comments:

Line 59: "ESM crop-models often use default global parameter values rather than crop- and region-specific values" – this is partly untrue. Take CLM4.5 for example, it uses crop-

specific parameters for six major crop types, and it also considers regional differences for soy and maize by implementing tropical and subtropical cultivars for these two crops. It may be too challenging to use site-specific or region-specific parameters for a crop model in ESM if the objective of model development is for global applications.

Lines 83-86: are these distinctive characteristics of perennial crops being reflected in the current model development for miscanthus and switchgrass?

Section 2.1: The described phenology is very similar to annual crop phenology, except that annuals are planted every year, but perennials are planted once and have repeated leaf onset, growth, senescence and harvest cycle each year. The major difference of the two perennial crops from the generic grass is their longer growing season and "planting once", while the phenological cycles are essentially the same as annual grass/crop in CLM4.5/ELM. It seems the Cheng et al. 2020 study also used this strategy to simulate miscanthus and switchgrass (please verify the statement in Line 291-294).

Section 3.3: Isn't there LAI and/or biomass harvest data for the two crops? It would be beneficial to have harvest data for validation of the model, given that this is the key output of interest for bioenergy crops.

Line 208: the seasonal dynamics of H is poorly simulated by the ensemble. It will be good to see more explanation why the observed trough of H in the summer season (May-Aug) is not captured by the model.

Lines 231-232: this could be mentioned earlier in the methods section when describing the construction of surrogates.

Section 4.4, Lines 236-239: From here I realize that the authors calibrated the parameters for the four QoIs separately and obtained different optimized parameter values for the different QoIs in Table 2. Thus, I asked the above main questions. Only if the authors have determined a final set of parameter values for each crop type (not for each QoI) and validate the whole model using this parameter set against various observations of (e.g., LAI, carbon stocks, harvested biomass, etc.) in addition to the variables used for calibration (GPP, ER, LE, H), can we trust that the new model is fully developed and validated for real world applications.

Lines 243-245: these less reliable parameters are the most sensitive ones -- What is the implication for the model accuracy? This needs to be discussed.

Lines 262-263: this same reason should also lead to decreased H during the summer

season, which is not well captured by the model.

Line 264: the diurnal mismatch is likely related to plant hydraulics and soil hydraulics.

Line 265: not really captured the seasonality, esp. for Sep. and Oct. as shown in Fig. 3H

Section 5: The Discussion is focused on sensitivity analysis and calibrated parameters. While they are an important step of model development, the discussion could devote more to validation and comparison of model performance with reference to observation or the literature, and the implications for model application in potential research fields. For example, why the model failed to capture the seasonality of sensible heat? And can the model correctly simulate harvested biomass or soil carbon pools (which are important consideration for bioenergy crops)? Even if there are no direct observations at the study site, there may be published data or related information in the ecology or agricultural literature.

The manuscript can also be improved if the authors clearly list the limitations and uncertainties of the current model.

Line 300: Given the high sensitivity of the model to this parameter, even if the optimized range of slatop matches observed range, a single value of slatop should be decided for each crop functional type in order to be used in cross-site or regional simulations. Please see earlier comments about this issue.

Lines 325-326: Why water budgets but not carbon budgets and harvests? Carbon is the major concern for bioenergy crops. Again, calibrating a land surface model using spatially varying site data would lead to spatially varying parameter values, which may work for those specific sites used for calibration, but cannot guarantee their applicability to other places or to regional and global simulations. That's why I commented that the current study lacks an independent validation using the calibrated and determined parameter set against different datasets and/or different variables of interest.

Figure 1: it would be helpful to add the ensemble mean of model simulations, not just the range.

Table 2: explain what the dash "-" means.

