Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-238-RC1, 2021

Comment on bg-2021-238

Anonymous Referee #1

Referee comment on "A Bayesian sequential updating approach to predict phenology of silage maize" by Michelle Viswanathan et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-238-RC1, 2021

The authors present a very interesting approach to Bayesian model calibration that has been under-exploited within the crop modeling community. I very much enjoyed reading it. The topic and its treatment in this manuscript are compelling and likely of interest to the Biogeosciences readership and crop modeling community more generally. The results and discussion presented are interesting, but the sampling approach and checks for convergence were not documented well enough for me to evaluate whether the results presented were valid. Further documentation is needed before the manuscript can be reconsidered for publication.

Multiple details of the sampling approach used in this study remain unclear. The authors provide equations 3 & 4 as a formal expression of Bayesian sequential updating (BSU) in which the prior is defined based on a priori beliefs and the likelihood is derived from first site-year of data. Equation 4 indicates that the prior for the second site-year would then be the posterior distribution sampled using equation 3. The prior for the third site-year would be the posterior of the second site year, and so on. However, if I understand correctly, BSU is not the approach used in this study. Instead, the prior remains fixed across all site-year combinations and only the quantity of data used for the likelihood calculation increases with each subsequent site-year. This approach is fine and perhaps is better than the BSU approach in that over time the evidentiary basis for posterior inference is broader and increasingly more likely to encompass the full range of environments over which prediction can be accurately performed. However, can this second approach be accurately termed BSU? I would suggest using an alternate term for this approach (at least something like "approximate BSU") and adjusting the title accordingly. Still, that is a relatively minor point. The greater issue is the number of questions remaining on the how this general approach was implemented. For example:

- How were chains initialized? Randomly sampling the prior? (The effectiveness of the Gelman-Rubin diagnostic generally depends on the starting points for multiple chains be overdispersed with respect to the posterior.)
- How many iterations were used for adapting the jump-size/transition kernel?

- When adding a new site-year, how were the chains handled? Were they re-initialized (along with retuning the transition kernel)? Was new data simply added to the dataset and chains allowed to adapt?
- How long was the warmup/burn-in? Was this variable?
- How many samples were generated after warmup? I see a number of 500 in Appendix B. That seems very low.
- How was parameter mixing evaluated?
- How did you check for auto-correlation in samples?
- How did the traceplots look? (Consider including representative traceplots in manuscript or supplementary methods)
- What were the numbers of effective samples (e.g. see https://mc-stan.org/docs/2_28/reference-manual/effective-sample-size.html)?

It is essential that these questions be addressed and I think doing so should not require more than adding a paragraph or two of text and possibly a supporting figure.

I also have several other specific suggestions that I think would improve the manuscript:

- line 140 Please indicate the identity of the expert (possibly in the Table 2 caption?) Citing as personal communication?
- line 176-189 A flow chart to show sequence of steps described would be very helpful.
- line 210-215 A figure visualizing the shape of eq 10 and 11 would be very helpful.
- In Figure 4(ii) What is meant by the term "Generative"? Is that referring to the "Reproductive" phase of growth (i.e. post-flowering)?
- I suggest adding some more discussion of the posterior distribution of parameter values presented in Figure 5. For example, why are there differences in parameter values across the two sites? Why do PDD1 and DELTMAX1 both decrease when the sequential years are added? Also, the shifts in distribution from prior to posterior indicate learning from the data. What do those shifts tell you about the cropping/soil system that was not known beforehand?