

Biogeosciences Discuss., referee comment RC2
<https://doi.org/10.5194/bg-2021-238-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-238

Justin Sexton (Referee)

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-RC2>, 2021

The authors apply a Bayesian Sequential Updating approach to calibrate phenological parameters of a maize crop model. I haven't seen this type of approach applied to a crop model before and was very keen to learn more about it. In particular, given that long term and high quality agronomic data is often difficult to come by, the idea that the model could be iteratively updated as new data became available is very interesting. In general the quality of the writing is quite high. The manuscript was easy to read and mostly well explained. However, there for a few areas that I was unsure of and need to be further explained in order to make sure the methodology and results are valid.

I was confused as to why the authors chose to calibrate phenology parameters using data from different cultivars, which in several cases were known to have different phenologies. (i.e. early vs mid maturity). This isn't a standard practice unless perhaps you are trying to calibrate a model to match regional yields when you know a range of cultivars are used in the region. Or is there an assumption that Phenology doesn't differ between cultivars? I apologize if I've missed this, maize and this model are not my area of expertise. The authors do point out that this may have contributed to the decrease to model skill in the discussion but since it was known beforehand that there were different cultivars grown in different seasons I think the reasoning for this methodological decision should be defined (e.g. what was the objective in calibrating phenology parameters if not to capture differences in phenology?)

I was confused by the synthetic model runs and how the validation was assessed. There were two synthetic scenarios, an 'ideal' and a 'cultivar-environment' scenario. Both seem to be based on the 6_2010 site-year and simply to have had random noise added to them based on observed variance. So I'm not sure what the effective difference ended up being. Figure 3 seems to suggest that there isn't really any difference in the uncertainty between the two scenarios and there doesn't seem to be a difference between how they are simulated.

Furthermore, the 'validation' set contains 10 'site-years' which are treated as independent (i.e. figure 6). From the methodology, these site-years are simply 10 random samples from the noise distribution. To me this seems that the mean and variance of these 10 would be more meaningful than treating them as individuals. Why are they treated separately?

Specific Comments

Figure 1: What data are the boxplots based on? Is this 30 points per box (5 replicates by 6 locations) or 6 points per box (mean of 5 replicates for 6 locations)

Line 130: Probably need a bit of clarifying here. Was the simulation set to measured soil-water conditions at the start of each season? There is some mention of a 'burn-in' period presumably to settle the soil nutrient and water levels. My concern here was that some of the seasonal effects may come from what crops were grown beforehand and if there was any fertilization etc. that occurred at the start of the season.

Equation 12-13: Performance is based on skill throughout the season, for different site-years. Was there any analysis of skill at different times of the year across site-year combinations? I know this wouldn't make sense for the final sequence (only 1 validation site-year). But for the second site in particular where you could have run for 3 site-seasons and have 3-site-seasons for validation. It would be interesting to know if including the extra years had improved the ability across site-years (since this is what the calibration is actually trying to do). i.e. calibrate on site-year 1,2 and 3 tries to fit phenology to best explain variance across these three site-year combinations.

Figure 3: The boxplots should be side-by-side. It's difficult to read the overlapping boxes. In general across all figures, it's impossible to read the figures if they are not printed in colour. It would be worth checking with the journal if there are any requirements around this.

Line 285: 'structural model errors': What constitutes a structural model error? Maybe some examples for this particular model would be informative.

Figure 4: I would like to see the PDFs for the sequential calibrations as a supplementary material. I think this would be more informative than say the entropy. For example, the first calibration is going to shift the distribution towards the specific cultivar. If the next calibration add a different cultivar (especially a different phenology group) the calibration

has to try and shift somewhere between the two. This would either spread out the PDF (e.g. equi-finality as mentioned later by the authors) OR it may produce a bimodal distribution. (e.g. two separate but equally likely solutions that the calibration switches between while trying to match the two different cultivars). The results wouldn't change but it may help with interpretation.

Figure 6: as mentioned above, if the 10 validation sites are just random permutations I don't think showing each of the ten is meaningful. The average and perhaps the variance would suffice.

Figure 7: This figure was difficult to read. As the validation results are what is important, I don't think the calibration (red box) is necessary. I think the meaning of the boxes needs to be more clearly explained as well. The boxes represent different MCMC initializations? Line 376 suggests that only the median value is plotted. The specifications of the MCMC process aren't clearly defined in the Methodology. (e.g. number of initializations, burn-in period etc.).

Line 460: "it is common practice to determine cultivar-specific parameters in crop modelling", Because of this I think some explanation is needed as to why phenology was calibrated across cultivars.

Line 470: "While the collection of cultivar and maturity group information...." I'm not sure of the link here to the hierarchical Bayes method? Are you trying to say that ideally cultivar values should be collected but an alternative is to calibrate using Hierarchical Bayes? Why not do that using your technique?