Biogeosciences
Open Access
Discussions

Interactive comment on "Environmental drivers of drought deciduous phenology in the Community Land Model" *by* K. M. Dahlin et al.

K. M. Dahlin et al.

kdahlin@msu.edu

Many thanks to the reviewers for your thorough and helpful responses. We have addressed the comments within the manuscript and please find our responses to your comments and suggestions below, identified by paragraphs that start with a ##.

Major comments and suggestions: The papers central result seems to be the pragmatic "fix" to stop the unrealistic simulation of multiple onsets over much of the draught deciduous parts of the world. This fix – using a cumulative rainfall threshold - is basically designed to stop anomalous onset caused by unrealistic soil water movement in CLM. While I am not completely against using these kinds of pragmatic solutions to prevent erroneous events within a model - and have had to use them myself - to base an entire paper around such a development that has no real world basis seems to be a

bit overkill. The authors identify the actual cause of the models problems in simulating seasonal phenology (i.e. that a sudden drop in transpiration at the end of the growing season and a potentially unrealistic water table dynamics triggers leaf onset through a sudden increase in soil water potential), but fail to investigate potential changes in the simulation of soil water dynamics or a more real-world based soil water/phenology coupling. To me it seems that this would have been a much better starting position for model development and, based on the information that the authors (again, very clearly) explain in the paper, I can think of several avenues of investigation, provided that data is available for proper analysis: 1. Is the simulation of the "unconfined aquifer" and its interaction with the soil layers realistic? (to be fair, the authors do suggest may be tricky because of lack of data) 2. Is the sudden increase in water potential after the growing season realistic? If not, is there a way of "blocking" this sudden spike? A pragmatic solution here would probably be more appropriate than the one the authors implemented, as the fix is applied at the root of the problem. 3. Is phenology actually coupled to soil water potential? If not, is a more realistic decoupling than the one put forward possible, along with a demonstration that this is based on real world plant responses? 4. Can plants "pull out" of onset if soil water potential peaks but then suddenly drops again?

We agree that this 'quick fix' is not the ideal solution to a fairly significant issue in CLM, however, our goal was to correct this one problem (dry season green-up) without significantly impacting other parts of the model. Several groups are currently working on new and better soil water solutions for the CLM, but changes to the soil water algorithms will have wide reaching impacts that will need to be assessed in their own right. The CLM4.5 is widely used by the land surface modeling community, and its development is a collaborative effort that relies on identification and publication of issues as rapidly as possible, to reduce redundant development efforts. We felt that this issue was too important to wait for those soil water parameterizations to come on line, given its impacts on the global carbon cycle in the model (new soil hydrology representations are not finalized at this time). It is not our decision whether or not this modification is incorporated into the next release of CLM, but we wanted to make it

available if deeper issues in the soil water algorithms are not resolved, or to others to use prior to the release of the new model, which will likely be up to a year from now. Regarding #3 and #4, those are great questions. KMD is currently conducting a literature review to attempt to address these questions across the semi-arid parts of the globe, but so far the answer seems to be that it depends on the plant and that there is quite a bit of variation in how plants respond to soil moisture at scales finer than the CLM plant functional type groupings. We hope that this ms stimulates more focus in the land modeling community on dry ecosystems and their processes.

With a slight shift on focus, the paper may have the potential for being an interesting case study of model benchmarking and evaluation. However, model quantitative evaluation of initial and developed model is a little spare and not systematic. A more suitable evaluation of phenology would probably need to assess the models simulation of the timing of the start, peak and end of the season, magnitude and number of onsets via quantifiable metrics, rather than just RMSE over the annual cycle and visual spatial and timeseries comparisons. Also, a proper assessment of model improvement (or degradation) over a range of model outputs outside of phonology would help guide interpretation of the results impact and guide further work (as the authors hint at in the discussion on carbon and fire on page 5820). There has been a lot of work on land surface model benchmarking over the past few years (see e.g. Randerson et al. 2009; Lou et al. 2012; Kelley et al. 2013) which could serve as a good starting point for assessment. Randerson and Kelley both design metrics for assessing simulated vs observed seasonal signals, and then relate this to vegetation cover and/or productivity (Randerson by comparing LAI, Kelley by comparing faPAR), and both demonstrate full model assessment of both chosen area for development and full-model impacts.

This paper did indeed begin as a benchmarking study, however, the discovery of the strange behavior of the model phenology in drought deciduous areas quickly changed the focus of the paper. As stated above, much more work is needed in this area, but we think that a description of the current state of the model is an important contribution.

C3866

Benchmarking software following the concepts in Kelley et al. 2013 and Luo et al 2012 is in development for the CLM, and will be available soon. It should be noted, however, that many benchmarking systems used monthly data as a default output. This study highlights that a great deal of information can be lost via the use of automated assessments of monthly timeseries. There are other metrics that we might have used, but identification of onset and offset from the satellite timeseries is subject to issues surrounding the subjectivity of the criteria defining onset and offset thresholds. We considered that the direct metrics of RMSE and R2 over the whole timeseries were more appropriate measures, given the complexity of the multiple growing seasons simulated in the default model. Regarding the other benchmarking metrics we could have used, we agree that this should be addressed, and we've added the following paragraph to the last section of the Methods and a sentence emphasizing this point in the conclusions:

The recent focus on land model benchmarking has led to a number of additional suggested methods for assessing seasonality in models compared to data (e.g. Randerson et al 2009, Kelley et al 2013), however, the proposed metrics would not capture the central issue addressed in this paper – model output with two or more peaks per year, data with only one – as they begin with the unstated assumption that seasonality is unimodal over the course of a year, as do measures of the start and end of the growing season. In Randerson et al (2009) seasonality is assessed by identifying the month of peak LAI and comparing that to MODIS LAI (MOD15A2), and in Kelley et al (2014) several more complicated metrics are introduced (equations 7-9) to again produce single numbers to compare a model's seasonality to a benchmark data set. In these examples, as in other benchmarking studies, the focus is on producing a single number, which, while useful, can miss important details.

Specific comments:

Region assessment – pg 5807, line 19/20. It would be useful to provide information on how PFTs are prescribed: i.e what is the dataset, what period is the prescribed cover

based on etc.

Lawrence & Chase 2007 (referenced in the ms) describes the development of CLM's land cover data set in detail.

- pg 5808, equation 1: Should ψ offset be ψ onset?

Good catch! Looks like this error was introduced during typesetting per your later comment we've changed all of these to psi-threshold

- pg 5808, line 25: Why is the number of onset days prescribed as 30?

That's what is prescribed in CLM currently – re-written to be more clear

- pg 5809, line 4: Why is a timestep 108000s? If I've got my arithmetic right, 10800s is 30 hours or 1.25 days.

This is a typo (thanks!) – should be 30 minutes, or 1800 seconds.

- Pg 5809, line 20: why are ψ offset and ψ onset defined as different parameters despite having the same value?

Changed to be one symbol defined as 'soil water potential threshold'

- Pg 5809, line 22: as with the number of onset days, why is the offset period proscribed as 15 days?

As with onset, that's what is prescribed in CLM currently – re-written to be more clear

- 5810, lines 4-9: A little more detail on climate information would be nice: What climate variables are needed to run the model? Which of these has the largest effect on soil water potential (and therefore the phenology indices)? Are soils prescribed? Does the model require CO2 inputs? Is the 45- year run using transient or equilibrium (detrended) climate? If detrended, how is this done? What time period does the run cover? What was the climate and vegetation cover (and soil and CO2?) inputs for the

equilibrium baseline state run? How was equilibrium tested for in the baseline state?

We have included more detail in this paragraph to address these questions. It now reads as follows: The model runs used in the global simulations described here ran for 45 years, and were started from an equilibrium baseline state generated by a standard CLM spin up run, (as described in detail by Koven et al. 2013) cycling meteorological conditions of 1948-1972. The present-day run (1965 – 2010) used CRU-NCEP meteorological reanalysis data (N. Viovy, pers. comm.; data available at: http://dods.ipsl.jussieu.fr/igcmg/IGCM/BC/OOL/OL/CRU-NCEP/) and transient CO2 concentrations to drive the model. Soil type and land cover are prescribed in the model, and recent work has suggested that the soil resistance parameterization may be unrealistic in arid ecosystems (Swenson & Lawrence 2014). More details on CLM are available in Oleson & Lawrence 2013.

- Pg 5810, lines 18-22: I'm not sure I follow the spatial averaging and resampling procedure. Were observed cells just averaged to decide if a CLM cell should be excluded? And was averaging performed just using LAI3g cells falling entirely within a CLM cell, with resampling used to incorporate the rest?

We've revised the sentences to say:

First, the two LAI3g maps generated for each month were averaged, then the LAI3g pixels were aggregated (averaged) to match the size of a CLM grid cell ($\sim$165 pixels per grid cell). If more than 80% of the grid cell did not have values in LAI3g (mostly applicable at high latitudes), the entire grid cell was removed from further analysis. Finally, the aggregated LAI3g data was resampled using a nearest neighbor approach to align with the CLM grid for further analysis.

- Pg 5812, line 13: Does the Latin-hypercube approach test for equilibrium itself, and/or did you define an equilibrium position was?

We defined the equilibrium state, as clarified in the next paragraph (p. 5812, line

21)

- pg 5813, line 18: Again, why 10 days?

We investigated alternative rainfall accumulation periods between 5 and 60 days. The impact of the accumulation periods was low, (since the model is mostly sensitive to the condition that there is any rain at all, rather than to the precise definition of the threshold).

- Pg 5812 (5814, I think?), lines 3-4: I'm not sure I understand this sentence. Are the three maps of LAI for CLM4.5BGC, LAI3g and CLM-MOD?

We've modified this sentence to be more clear and it is addressed in the results.

- Pg 5815, line 9: Is there an important reason for running CLM globally here?

Not particularly – we've eliminated the word 'globally'.

- Pg 5816, line 8-9: Surely a model change such as this would require a new equilibrium baseline state? If you think it doesn't, what is the rationale?

We discuss the equilibrium conditions on P5812 L18-L22. Spin-up of the new model to an LAI equilibrium was remarkably fast in these simulations, potentially reflecting a lack of significant feedback between the soil nitrogen state and vegetation physiology for the semi-arid regions.

- Pg 5816, line 9-12: can this "closer match" be quantified?

We feel that showing the maps, including the difference maps in Figure 6, particularly 6D and 6F adequately illustrate this improvement.

- Pg 5816, line14-19: Again, can this be quantified? In figure 7, it looks like some places get better (sahel, southern and western Australia etc) whilst some places get worse (i.e, Asia, north east Aus etc).

We've added a confusion matrix table comparing counts of peak numbers between

LAI3g and the two model runs (CLM and CLM-MOD) and this paragraph now reads as follows:

To test whether the poor fit between CLM and LAI3g was due to multiple annual LAI peaks in CLM we counted the number of peaks per year in each data set (Fig 7). We found that in the observations, only areas in the humid tropics had multiple peaks in the LAI3g data ("peaks" in these cases being relatively small fluctuations), while CLM showed multiple peaks per year throughout many of the savanna regions of the world. CLM-MOD has more areas with only one peak, particularly in Sub-Saharan Africa. To quantify these changes to the model we constructed confusion matrices to compare the peak counts in LAI3g to those in CLM and CLM-MOD (Table 3) for grid cells with >50% drought deciduous cover (Figure 1). Overall, CLM-MOD had a slightly poorer performance, matching the number of peaks in the LAI3g dataset 42.5% of the time, while CLM matched LAI3g 43.7% of the time. However, these unweighted summary numbers mask improvements in CLM-MOD. CLM only correctly predicted a single peak 8.9% of the time, while CLM-MOD correctly predicted single peaks 59% of the time, and never did CLM-MOD predict more than two peaks in a year, matching the LAI3g data. The overall degradation in CLM-MOD is due to fewer correctly identified grid cells with zero or two peaks.

- Pg 5816, line 24 - pg 5817, line 2: Again, the improvement does not seem to be quantified. Introduction of a more systematic benchmark system (see major comments) would help with these last 3 points)

Here, and in the above 2 comments, it would be possible to determine a single number for the change in model performance, but, as can be discerned from the global maps, the situation is actually spatially complex, and, as highlighted by Luo et al. and Randerson et al., the generation of single number benchmarking products is fraught with subjectivity. This is why we chose to present the results in their entirety, as part of the development of process representations.

- Pg 5818, line 23-pg 5819 line 20: This part seems very out of place in the discussion. The implementation of this should be in the methods, and it would be nice to see some results, even if they go in an SI. If you cannot do this, I would take this section out.

We feel that it is important to keep this discussion in the paper as these alternative options were frequently brought up when this work was discussed among CLM users, but we've moved it to the methods (section 2.4).

References

Levis, S., Bonan, G. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., & Kucharik, C. J. (2012). Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land-atmosphere fluxes. Journal of Climate, 25(14), 4839-4859.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., ... & Swenson, S. C. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. Biogeosciences, 10(11), 7109-7131.