Biogeosciences Discuss., 12, C3274–C3278, 2015
www.biogeosciences-discuss.net/12/C3274/2015/

Biogeosciences
Discussions

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

Anonymous Referee #2

Received and published: 3 July 2015

Dahlin et al. use remote sensed observations of LAI to assess the simulation of draught deciduous phenology in the Community Land Model (CLM). They identify weaknesses in the simulation of seasonality, mainly due to anomalous extra greening up periods. In order to fix this, they go on to optimize several phenology related parameters and implement a "pragmatic" semi-decoupling of phenology from soil moisture by introducing a cumulative threshold of recent rainfall.

The authors do a good job of explaining draught deciduous phenology, how phenology in CLM works, and where problems in CLMs simulation lie. I found all this very easy to understand despite no previous experience with this particular model and only limited knowledge about phenology. I do, however, have some major concerns about the model development and evaluation which may require some significant changes to the

m/s. I also have a number of specific comments, some of which may just need brief clarification in the author's response.

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?

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.

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.

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

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

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

- Pg 5809, line 20: why are ψ offset and ψ onset defined as different parameters despite

having the same value?

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

- 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?

- 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?

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

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

- Pg 5812, lines 3-4: I'm not sure I understand this sentence. Are the three maps of

LAI for CLM4.5BGC, LAI3g and CLM-MOD?

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

- 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?

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

- 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).

- Pg 5816, line 24 0 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)

- 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.

References

Luo, Y. Q., et al. "A framework for benchmarking land models." Biogeosciences 9 (2012): 3857-3874.

Kelley, D. I., et al. "A comprehensive benchmarking system for evaluating global vegetation models." Biogeosciences 10 (2013): 3313-3340.

Randerson, James T., et al. "Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models." Global Change Biology 15.10 (2009): 2462-2484.

Interactive comment on Biogeosciences Discuss., 12, 5803, 2015.