

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 ##.

Anonymous Referee #1

First, I am curious why the authors chose to use the LAI3g dataset rather than the MODIS LAI product (MCD15A2). While these datasets have similar agreement with respect to in situ LAI observations (RMSE \approx 0.7; Zhu et al. 2013 RemSen), both validation exercises used a limited number of sites located in drought deciduous regions (e.g., Figure 2 in Zhu et al.) and MODIS data are known to have higher fidelity than AVHRR data (Huete et al. 2002 RemSenEnv). Moreover, while the LAI3g dataset is twice as long as MCD15A2, it is not clear how additional years of data actually benefit the model assessments made in this analysis. It would be helpful if the authors provided explanation for why they chose to use LAI3g rather than MCD15A2. If this is not possible, then I recommend that the authors redo the analysis using MCD15A2 to ensure that the highest quality data are used.

We agree that the MCD15A2 data set is more robust, but we felt that the length of the LAI3g data set was important, as semi-arid ecosystems have inconsistent rainfall and green-up patterns, making more data useful to cover a broader range of weather patterns. In addition, since we were focused on broad issues in CLM, not pinning the LAI values exactly to the satellite data, we expect that switching data sets would not significantly change the overall results, given that the discrepancies in the original model were so profound.

Second, while it is certainly necessary to assess model performance over the entire annual cycle, I believe it is equally important to consider how well models capture timing of seasonal metrics such as start and end of season since they largely control annual carbon uptake (e.g., Ma et al. 2007 AgForMet). For example, in Figure 2a it is apparent that the LAI3g growing season across NH C3 grasses is shifted early by an entire month. Despite the relatively high R2 and low RMSE across this region as shown in Table 2, this result suggests that the model does not incorporate the mechanistic controls for triggering leaf onset or autumn senescence. Therefore, for regions with stronger seasonality in LAI (mainly temperate grasslands), I suggest the authors perform a quantitative assessment of model bias in start and end of growing season dates.

We agree that this would be an additional interesting assessment, however, we do not think it would change the overall results of this paper, and algorithms like those designed to assess start and end of growing season, growing season length, or amplitude do not perform well when trying to assess areas with more than one growing season per year, which are a common part of savanna-type ecosystems. We have added comments regarding this in the methods section of the manuscript (see response to reviewer 2, below). Further, our study is focused on semi-arid systems, and not on cold-deciduous phenology, which has been studied more extensively in the past (e.g. Levis et al. 2012).

Third, perhaps this is outside of the context of this study but if the authors have access to in situ observations of LAI data from a semi-arid/drought deciduous region it would be interesting and

worthwhile to validate the authors' novel cumulative rainfall model using these data. This is briefly mentioned on Page 5821, Lines 23-24.

We agree! There is lots of room for more assessment of drought deciduous phenology, however, there are only a small number of in situ data sets (e.g. Phenocams) in drought deciduous areas, as far as we know. Adding more data sources (both on the ground and space-borne) is a focus of ongoing proposals, etc.

Finally, while the authors describe and perform parameter sensitivity analyses in the methods and results sections (research question #2), there does not appear to be any significant discussion or formal conclusions made regarding these results later in the manuscript. Moreover, in Figures 4 and 5 the relationship between each model run and the magnitude of the varied parameter is unclear. Overall, I believe this is an important aspect of the analysis and, therefore, suggest that the authors make appropriate changes to resolve these issues.

Latin hypercube types of analysis are difficult to illustrate, however, we agree that the discussion of the sensitivity analysis could be more detailed, and appropriate changes have been made to the manuscript explaining the results of the sensitivity analyses and what is shown in Figures 4 and 5.

We've added the following paragraph to explain this challenge in the methods, as well as described this issue elsewhere in the manuscript:

To assess the performance of the different models in the Latin hypercube test we originally plotted the coefficients of determination between the different models' LAI values and the LAI3g data at those points. However, this result did not illustrate any clear optimum in model performance either for the parameters of the existing model, nor for the rainfall threshold. We illustrate this using the time-series data in Figure 4, which highlight the unusual behavior of the model, and to assess whether the extra green-up period during the dry season had been eliminated in any of the parametric permutations. We ascribe the lack of a clear parametric signal to two effects. First, the LAI3g data were necessarily aggregated to monthly values, meaning that the primarily sub-monthly variation between ensemble members was masked. Second, the timing of the secondary leaf-on period in the dry season was the emergent property of the oscillatory (and thus somewhat chaotic) dynamics of the soil-vegetation feedback on soil moisture.

Regarding the rainfall threshold, we've added the following statement in the results:

While this new rainfall threshold improved model performance both at our points and globally (see below), we note that, except in a few exceptionally dry areas, the model did not appear to be particularly sensitive to the rainfall threshold, as long as some rain did fall, but this threshold, and the drought deciduous algorithm as a whole, deserves more research into seasonal drivers.

Minor Comments and Suggestions:

Page 5804, Line 23: Not sure that quotes are necessary here.

Removed

Page 5807, Line 7: Please explain what BGC stands for.

BGC stands for Biogeochemistry and designates a particular version of CLM that includes an active carbon and nitrogen cycle, but is different from the CLM-CN version (also with active C and N) which was used primarily in earlier versions of CLM. Since CLM4.5BGC is now the

default version, we have included an explanation in the beginning of the paper but changed further mentions of CLM4.5BGC to CLM for simplicity.

Page 5808, Equation 1: offset → onset

Good catch! Looks like this error was introduced during typesetting – per reviewer 2’s comment we’ve changed all of these to psi-threshold

Page 5810, Line 6: Please explain what CRU-NCEP stands for.

CRU-NCEP = Climate Research Unit (University of East Anglia) – National Centers for Environmental Prediction (NOAA), we did not define this following other papers published in Biogeosciences, but will defer to the Editor’s preference.

Page 5810, Line 19 (and instances afterward): gridcell → grid cell

Changed

Page 5816, Section 3.3: The CLM naming convention (e.g., CLM4.5BGC, CLM, CLMMOD) gets a little confusing here. Perhaps it makes sense to only use CLM and CLMMOD?

See comment above – we’ve changed CLM4.5BGC to just CLM

Page 5822, Line 12: phonological → phenological

Changed. Whoops!

Figure 2: Need to show letters in each panel.

Letters are in the bottom right corners of each panel – e.g. “A. NH C3 Grasses”

Figure 3: This figure is somewhat busy, although I am not sure what makes the most sense to remove. Perhaps it is okay but I suggest the authors consider alternative representations, if possible.

We agree that these figures (and Figure 10) are challenging, however, we opted for this presentation as it shows all of the relevant details, and we felt that stacking all of the data allowed for easier comparisons than plotting things in different panels.

Figure 4: If the authors choose not to explain differences between each model run, it may also make sense to just plot a mean curve with +/- 1 standard deviation or some other indicator of variance.

Related to the comment above, we’ve added a more clear description of the Latin hypercube analysis and discussion of these figures to better explain how to interpret them.

Figures 8/9: Perhaps it would make sense to mask out grid cells with less than some fraction of drought deciduous land cover so that it is easier to depict spatial patterns in the improvement of model performance.

We've opted to leave all of the data clearly represented (instead of masked) with the hope that these figures will be useful to researchers interested in study regions beyond those we have focused on.

Anonymous Referee #2

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 sparse 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 phenology 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. 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 (~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.

1 Environmental drivers of drought deciduous phenology in 2 the Community Land Model (MARKUP VERSION)

3 K.M. Dahlin^{1,2}, R.A. Fisher², & P.J. Lawrence²

4 [1] Department of Geography, Michigan State University, East Lansing, Michigan, USA

5 [2] Climate and Global Dynamics Division, National Center for Atmospheric Research*, Boulder,
6 Colorado, USA

7 * NCAR is sponsored by the National Science Foundation.

8 Correspondence to: K.M. Dahlin (kdahlin@msu.edu)

9 10 **Abstract**

11 Seasonal changes in plant leaf area have a substantial impact on global climate. If and when leaves
12 are present affects surface roughness and albedo, and the gas exchange occurring between leaves
13 and the atmosphere affects carbon dioxide concentrations and the global water system. Thus,
14 correct predictions of plant phenological processes are important for understanding the present and
15 future states of the Earth system. Here we compare plant phenology as estimated in the Community
16 Land Model (CLM) to that derived from satellites in drought deciduous regions of the world. We
17 reveal a subtle but important issue in the CLM: anomalous green-up during the dry season in many
18 semi-arid parts of the world owing to rapid upwards water movement from wet to dry soil layers.
19 We develop and implement a solution for this problem by introducing an additional criterion of
20 minimum cumulative rainfall to the leaf-out trigger in the drought deciduous algorithm. We discuss
21 some of the broader ecological impacts of this change and highlight some of the further steps that
22 need to be taken to fully incorporate this change into the CLM framework.

24 1 Introduction

25 Ecosystems change with the seasons in response to environmental cues. Some of those cues
26 are fixed, like day length, while others are climate-driven and therefore vary from year to year. The
27 combination of fixed and climate-driven phenological cues poses an interesting problem in the face
28 of climate change – climate related drivers of phenology (temperature and rainfall patterns) are likely
29 to change (Lau et al., 2013), while fixed cues will remain unchanged. Phenological shifts due to
30 climate change have already been identified (e.g. Parmesan & Yohe, 2003). “Phenology” can refer to
31 a large number of patterns and behaviors in plants and animals that shift with the seasons. Here,
32 however, because we are focused on land surface model simulations, we use phenology specifically
33 to refer to intraannual variation in leaf area index (LAI). Leaf area can vary significantly within a
34 year and is, therefore, a critical control on land-atmosphere feedbacks (Lawrence et al., 2012).

35 Recent advances have greatly improved our ability to predict seasonal patterns in northern
36 temperate deciduous forests (Richardson et al., 2012), but our understanding of phenological
37 patterns in stress or drought deciduous plants (also called ‘raingreen’) remains weak (Guan et al.,
38 2014; Jenerette et al., 2010; Ma et al., 2013). The semi-arid ecosystems that host the majority of
39 drought deciduous woody plants have relatively low biomass but make up a large fraction of global
40 land area (~30%; Scholes & Hall, 1996). Their extensiveness alone makes them important to global
41 radiation budgets, but additionally these systems are likely very sensitive to climate change given
42 their apparent bistability (Scholes & Hall, 1996; Staver *et al.*, 2011). In semi-arid ecosystems leaf-out
43 is typically thought to be a function of water availability (Reich, 1995; White et al., 1997), however,
44 some woody plants do leaf-out several weeks before the first rains of the season (Archibald and
45 Scholes, 2007).

46 In an Earth system modeling context, the timing and magnitude of plant phenology, and
47 how these processes may change, is critical for approximating the energy and carbon balances of the
48 planet. Prognostic phenology has only recently been incorporated in to Earth system models,
49 however, and its fidelity, particularly in semi-arid regions, remains poorly tested (Blyth et al., 2011;
50 Lawrence et al., 2011; Randerson et al., 2009). Lawrence *et al.* (2012) found that the prognostic
51 phenology in the Community Land Model version 4 (CLM4(CN)) degraded estimates of latent heat
52 flux and other biophysical properties in comparison to using prescribed, satellite-derived phenology
53 (CLM4SP). Wang *et al.* (2013) compared intraannual variation in the fraction of absorbed
54 photosynthetically active radiation (fAPAR) in CLM4CN to satellite-derived estimates and found
55 substantial differences in regional averages, zonal means, and interannual trends. It is difficult,
56 however, to isolate the impact of the drought deciduous phenology algorithm using these regional
57 and zonal estimates.

58 Satellite-derived estimates of greenness, fAPAR, and LAI have greatly improved our ability
59 to study the environmental drivers of phenology (Reed et al., 2009), however, the majority of studies
60 have focused on northern deciduous and boreal forests (e.g. Delbart *et al.*, 2006; White *et al.*, 2009;
61 Yang *et al.*, 2012). While fewer studies have focused on remote sensing of phenology in semi-arid
62 systems, Zhang *et al.* (2005) found a strong relationship between greenness onset and the start of the
63 rainy season across the semi-arid parts of Africa. They found a weaker relationship, however,
64 between dormancy and the end of rainy seasons, and they attribute this weakness to differences in
65 soil properties. Similarly, Ma *et al.* (2013) found a strong relationship between greenness and rainfall
66 in northern Australia in both seasonal timing and amplitude and Bradley *et al.* (2011) found a close
67 relationship between rainfall and seasonality in Amazonian savannas. Interestingly, in Africa Zhang
68 *et al.* (2005) also showed a strong relationship between latitude and both green-up and dormancy
69 onset, even in the narrow band of the Sahelian and Sub-Saharan region, suggesting a possible link

70 between phenology and subtle changes in photoperiod at least in northern Africa. Recently Guan *et*
71 *al.* (2014) showed a relationship between woody plant cover and phenological timing in African
72 savannas.

73 In this study we address three questions related to the representation of drought deciduous
74 phenology in the CLM. (1) How well does the CLM capture phenological patterns of LAI among
75 different drought deciduous plant functional types (PFTs) as compared to satellite-derived
76 estimates?; (2) Which parameters in the current version of the CLM have the most leverage on
77 ~~woody plant~~drought deciduous phenology?; and (3) Do changes in the phenology algorithms in the
78 CLM improve the model's representation of seasonal cycles regionally?

79

80 2 Methods

81 2.1 Model Description

82 The CLM is the terrestrial component of the Community Earth System Model (CESM;
83 Lawrence *et al.*, 2011); it simulates biogeophysical and biogeochemical processes including radiation
84 interactions with vegetation and soil, heat transfer in soil and snow, hydrology, and plant
85 photosynthesis and respiration. In this paper we use the most recent release of the Community Land
86 Model with active biogeochemistry, CLM4.5BGC (Oleson *et al.*, 2013). Henceforth, references in
87 this paper to the “CLM” will refer to CLM4.5BGC.

88 The CLM is run here on a 1.25° x 0.9375° grid, and each `grid_cell` is, where applicable,
89 divided into fractions representing vegetated land, lakes, glaciers, and urban areas. Within the
90 vegetated fraction of a `grid_cell` there may be multiple PFTs representing a coarse division of
91 biodiversity along its major axes of variation: trees/shrubs/grass, broadleaf/needleleaf, C3/C4
92 photosynthesis mechanisms and phenological habit (evergreen, cold deciduous and stress/drought

93 deciduous). There are currently 15 non-crop PFTs in the CLM, four of which follow the drought
94 deciduous phenology algorithm (Oleson et al., 2013). **Figure 1** shows where these different PFTs
95 dominate the globe. Over time the relative cover of the PFTs may shift, as may the overall fraction
96 of vegetation, depending on shifts in land use, though these shifts are minor in recent decades. In
97 the simulations used in this paper these shifts in PFT fractions and cover are prescribed from
98 satellite observations (Lawrence and Chase, 2007) as opposed to emerging from vegetation
99 competition (Bonan et al., 2003).

100 In the CLM, drought deciduous plants are represented by the ‘stress deciduous’ phenology
101 type, as distinct from the evergreen or ‘seasonal’ (cold) deciduous phenology types. This designation
102 allows for plants to lose their leaves either via the impact of cold, via the impact of drought, or via
103 the onset of short days thus allowing the model to simulate, for example, grass vegetation growing in
104 an environment that is both seasonally cold and seasonally dry. If the triggers for offset are not
105 reached in a given year, drought deciduous vegetation will follow the evergreen phenology
106 algorithm, gaining and losing fixed fractions of carbon with each time step. This stress deciduous
107 algorithm, described in more detail below and in Oleson *et al.* (2013), was developed in part from
108 White *et al.* (1997), though that study was particularly focused on grass phenology.

109 The deciduous algorithms are hierarchical, such that plants classified as ‘stress deciduous’ but
110 growing at high latitudes or in cold climates will follow the same onset/offset rules as
111 cold/seasonally deciduous plants. From the beginning of a dormant period a ‘freezing day
112 accumulator’ is activated whereby time steps with a temperature below freezing (0° C) are summed
113 and if this sum exceeds 15 days then the plants will follow both the winter deciduous and drought
114 deciduous algorithms. Leaf onset can only be triggered if day length is greater than 6 hours, a
115 latitude-specific sum of growing degree days has been reached (described in Oleson *et al.* (2013)) and
116 the soil wetness criteria described below have been met.

117 In seasonally dry, warm regions (the focus of this paper) where day length is never less than
 118 6 hours, leaf onset for the stress deciduous phenology type is determined by soil wetness. At the end
 119 of the previous offset period an accumulated soil water index (SWI) is set to zero and accumulation
 120 is calculated as:

$$121 \quad SWI^n = \begin{cases} SWI^{n-1} + f_{day} & \text{for } \psi_{soil\ 3} \geq \psi_{onsetthreshold} \\ SWI^{n-1} & \text{for } \psi_{soil\ 3} < \psi_{onsetthreshold} \end{cases} \quad (1)$$

122 Where n and $n-1$ refer to the values in the previous and current time steps, $\Psi_{soil\ 3}$ is the soil water
 123 potential (MPa) in the third soil layer (6.23 cm – 9.06 cm), $\Psi_{thresholdonset}$ is -2 MPa, and f_{day} is a time
 124 step (30 minutes in CLM) as a fraction of a day. Onset is triggered when SWI exceeds 15 days.

125 The rate of leaf onset (fraction of onset per time step), which in the CLM is represented as
 126 the transfer of C and N from a storage pool to the ‘display’ leaf pool, is determined by the number
 127 of days prescribed for onset, here fixed at 30 days. The rate (r_{onset}) at each time step is defined as:

$$128 \quad r_{onset} = \begin{cases} \frac{2}{t_{onset}} & \text{for } t_{onset} \neq \Delta t \\ \frac{1}{\Delta t} & \text{for } t_{onset} = \Delta t \end{cases} \quad (2)$$

129 where t_{onset} is time remaining in the current onset period in seconds and Δt is the length of a time step
 130 (~~108000 seconds~~ 1800 seconds). The flux of C out of the storage pool is then defined as the amount
 131 in the C storage pool at that time step multiplied by r_{onset} . These functions result in a linearly
 132 decreasing flux out of the transfer pool, so the rate of increase in LAI over the onset period steadily
 133 decreases as C moves from the storage pool to the display pool (see Fig 14.1 in Oleson *et al.* (2013)).
 134 During the onset period C and N are also transferred from storage pools for fine roots, live and
 135 dead stem, and live and dead coarse roots into these components’ respective displayed growth pools.
 136 During the growing season, C and N taken up by the plant are accumulated in transfer pools, to be
 137 used in the next growing season.

138 As long as the leaf onset period is complete, leaf offset can be triggered by short (<6 hr) day
139 length, a period of cold temperatures ~~has been reached~~ (described in Oleson *et al.* (2013)) or if the
140 soil dryness criteria described below has been met.

141 The offset soil wetness index (OSWI) can potentially start accumulating time steps once the
142 previous leaf onset phase is complete. The algorithm differs slightly from the onset trigger, in that
143 OSWI can increase or decrease as described below.

$$144 \quad OSWI^n = \begin{cases} OSWI^{n-1} + f_{day} & \text{for } \psi_{soil\ 3} \leq \psi_{thresholdoffset} \\ \max(OSWI^{n-1} - f_{day}, 0) & \text{for } \psi_{soil\ 3} > \psi_{thresholdoffset} \end{cases} \quad (3)$$

145 where $\psi_{thresholdoffset}$ is -2 MPa, and leaf offset is triggered when OSWI equals 15 days.

146 Similar to the rate of leaf onset, leaf offset rate is a function of the amount of time left in the
147 offset period, ~~here~~ fixed at 15 days:

$$148 \quad r_{offset} = \frac{2\Delta t}{t_{offset}^2} \quad (4)$$

149 Carbon fluxes into the litter pool are only calculated for leaves and fine roots (stems and coarse
150 roots cannot shrink). Nitrogen fluxes into the litter pool reflect retranslocation of N prior to offset.
151 See Oleson *et al.* (2013) for more details.

152 The model runs used in the global simulations described here ran for 45 years, and were
153 started from an equilibrium baseline state generated by a standard CLM_4.5BGC spin up run; ~~as~~
154 ~~described in detail by Koven *et al.*, 2013), cycling meteorological conditions of 1948-1972. The~~
155 ~~present-day run (1965 – 2010) used as documented in (Oleson *et al.*, 2013) and using~~ CRU-NCEP
156 meteorological reanalysis data (N. Vivoy, pers. comm.; data available at:
157 <http://dods.ipsl.jussieu.fr/igcmg/IGCM/BC/OOL/OL/CRU-NCEP/>) ~~and transient CO₂~~
158 ~~concentrations~~ to drive the model. ~~Soil type and land cover are prescribed in the model, and recent~~

159 [work has suggested that the soil resistance parameterization may be unrealistic in arid ecosystems](#)
160 (Swenson and Lawrence, 2014). [More details on CLM are available in](#) Oleson et al. (2013).

161 2.2 Satellite derived LAI

162 We compared the model-derived estimates of LAI to those estimated from the Advanced
163 Very High Resolution Radiometer sensors (AVHRR) onboard the National Oceanic and
164 Atmospheric Administration satellites. These data are available bimonthly and span from 1981 to
165 2011. They are supplied at 1/12 degree resolution. A detailed description of the development of the
166 LAI product (hereafter LAI3g) is in Zhu *et al.* (2013).

167 To ensure the most appropriate comparison possible, the LAI3g dataset was rescaled to
168 match the mean monthly LAI output from the CLM. [First, the two LAI3g maps generated for each](#)
169 [month were averaged, then the LAI3g pixels were aggregated \(averaged\) to match the size of a CLM](#)
170 [grid cell \(~165 pixels per grid cell\). If more than 80% of the grid cell did not have values in LAI3g](#)
171 [\(mostly applicable at high latitudes\), the entire grid cell was removed from further analysis. Finally,](#)
172 [the aggregated LAI3g data was resampled using a nearest neighbor approach to align with the CLM](#)
173 [grid for further analysis.](#)~~First, the two LAI3g maps generated for each month were averaged, then~~
174 ~~the LAI3g pixels within a CLM gridecell were averaged. If more than 80% of the gridecell did not have~~
175 ~~values in LAI3g (mostly applicable at high latitudes), the entire gridecell was removed from further~~
176 ~~analysis. Finally, the resized LAI3g data was resampled using a nearest neighbor approach to align~~
177 ~~with the CLM grid.~~ All spatial and statistical analyses were performed in R (R Core Team, 2013)
178 using the ncd (Pierce, 2011), raster (Hijmans and van Etten, 2013) and rgdal (Bivand et al., 2013)
179 packages.

180 2.3 Comparing LAI3g to CLM LAI

181 We first compared LAI3g to the CLM output for 1982 (the first full year of LAI3g) to 2010
182 (the last available year of the CRU-NCEP forcing dataset for CLM), aggregating values by zones
183 based on dominant PFT and hemisphere. To aggregate into PFT classes, we only considered grid
184 cells dominated by a single drought deciduous PFT (>50% cover), permitting six possible
185 comparisons with a sufficient number of grid_cells for comparison: northern hemisphere (NH)
186 temperate C3 grasses (n = 180), NH C4 grasses (n = 234), tropical deciduous trees (n = 242),
187 southern hemisphere (SH) deciduous shrubs (n = 68), SH C3 grasses (n = 160), and SH C4 grasses
188 (n = 271). Note that the actual number of grid_cells compared from year to year varied slightly with
189 changes in land cover and, in the case of LAI3g, available data. The counts listed here are the
190 averages for each PFT. To visually assess the comparison between LAI3g and CLM we plotted the
191 monthly means and standard deviations for these seven regions. We also computed the R^2 and root
192 mean squared error (RMSE) across all 29 years' monthly values to assess CLM's ability to fit the
193 seasonality and the magnitude of the LAI3g values. Due to the temporal coarseness of these data
194 and the irregular seasonal patterns found in many of our areas of interest, we did not fit a
195 continuous function to these data.

196 **2.4 Point Simulations and Parameter Sensitivity Tests**

197 Given the observed mismatches between LAI3g seasonality and CLM predictions among
198 drought deciduous woody plants (see Discussion), we conducted an analysis of how the parameters
199 determining phenology in the model affect the model outcome. The use of global or regional
200 simulations to assess the sensitivity of models to their structural and parametric assumptions is
201 problematic on account of both the computational requirements to do so and the high-
202 dimensionality of the model outputs, which can hinder understanding. To avoid these issues, we
203 conducted a sensitivity test to the major model parameters that control seasonality in LAI at the
204 point scale. In these point simulations we focused on low-latitude drought-deciduous ecosystems,

205 and selected six locations dominated by tropical deciduous trees, grasses, or a combination of the
206 two (**Table 1**).

207 The phenology model contains three empirical parameters that collectively describe the leaf
208 onset and offset algorithms.

- 209 1. Critical soil water potential ($\Psi_{\text{onset/offsetthreshold}}$) values for leaf onset and offset (default = -2
210 MPa)
- 211 2. Soil water potential days to onset/offset (SWI/OSWI) threshold (default = 15 days)
- 212 3. Quantity of carbon assimilation which is directed to current leaf growth rather than
213 storage (F_{current}) (default = 0.0)

214 To determine what impact the choice of these parameters has on the model outcome, we
215 conducted a Latin-hypercube analysis (McKay et al., 1979), beginning the model from a spun-up
216 state with default parameters. We then perturbed the parameters and ran the model forwards until a
217 new LAI equilibrium condition was detected. Because the nitrogen cycle is active in CLM, soil
218 biogeochemical equilibrium can in some circumstances (particularly at high latitudes) take many
219 decades or even centuries achieve, therefore we set a threshold for the new equilibrium state as the
220 absence of a trend in LAI resulting in a 2% increase over a five year period. Given the high
221 temperature and relatively low biomass and productivity of the ecosystems in question, in our
222 simulations LAI equilibrium was in practice typically reached after the first 5 year period of the
223 simulation. As with the global simulations, we used the CRU-NCEP reanalysis forcing data,
224 extracted at the six points of interest, to drive the point simulations.

225 The Latin hypercube methodology is a particularly efficient means of investigating a multi-
226 dimensional parameter space, because each run of the model perturbs every parameter, and the
227 algorithm ensures that the distribution of sampled points is distributed efficiently (but not

228 ~~uniformly) through parameter space. For this study, since we were comparing the model output to~~
229 ~~monthly satellite observations, we did not manipulate the number of days for onset (30) or offset~~
230 ~~(15) to occur. We also did not consider the growing degree days and day length parameters because~~
231 ~~the focus of this study was on tropical and subtropical regions where these components of the~~
232 ~~algorithm are not active. To determine what impact the choice of these parameters has on the model~~
233 ~~outcome, we conducted a Latin-hypercube analysis (McKay *et al.*, 1979), beginning the model from~~
234 ~~a spun-up state with default parameters. We then perturbed the parameters and ran the model~~
235 ~~forwards until a new LAI equilibrium condition was detected. The Latin hypercube methodology is~~
236 ~~a particularly efficient means of investigating a multi-dimensional parameter space, because each run~~
237 ~~of the model perturbs every parameter, and the algorithm ensures that the distribution of sampled~~
238 ~~points is distributed efficiently (but not uniformly) through parameter space.~~

239 ~~Because the nitrogen cycle is active in CLM, soil biogeochemical equilibrium can in some~~
240 ~~circumstances (particularly at high latitudes) take many decades or even centuries to achieve, therefore~~
241 ~~we set a threshold for the new equilibrium state as the absence of a trend in LAI resulting in a 2%~~
242 ~~increase over a five year period. Given the high temperature and relatively low biomass and~~
243 ~~productivity of the ecosystems in question, in our simulations LAI equilibrium was in practice~~
244 ~~typically reached after the first 5 year period of the simulation. As with the global simulations, we~~
245 ~~used the CRU-NCEP reanalysis forcing data, extracted at the six points of interest, to drive the~~
246 ~~point simulations.~~

247 ~~For this study, since we were comparing the model output to monthly satellite observations,~~
248 ~~we did not manipulate the number of days for onset (30) or offset (15) to occur. We also did not~~
249 ~~consider the growing degree days and day length parameters because the focus of this study was on~~
250 ~~tropical and subtropical regions where these components of the algorithm are not active.~~

251 For the critical soil moisture potential threshold, $\Psi_{\text{onset/offsetthreshold}}$, we investigated values from
252 0 to -3.5MPa (where the default value is -2_MPa). The upper end of this range is the maximum
253 possible value for saturated soils, whereas the bottom end of the range was determined from a prior
254 set of sensitivity tests which determined that sensitivity below this range was very low (i.e. the soil
255 moisture potential in the third layer rarely drops below -3.5_MPa, and so leaves remain on
256 continuously for those simulations at our locations of interest). For the number of days of onset,
257 $\text{SWI}-\Psi_{\text{days}}$, we followed a similar protocol, and found that the range of sensitivity was focused
258 between 5 and 35 days (where the default is 15 days). For the fraction of displayed assimilated
259 carbon (F_{current}) we varied the values between 0 and 0.5, (where the default is zero). Sensitivity of
260 average LAI to this parameter was low in all cases, but it has an impact on the intraannual cycle,
261 since the LAI is unchanging through a single growing season if $F_{\text{current}} = 0$.

262 To assess the performance of the different models in the Latin hypercube test we originally
263 plotted the coefficients of determination between the different models' LAI values and the LAI3g
264 data at those points. However, this result did not illustrate any clear optimum in model performance
265 either for the parameters of the existing model, nor for the rainfall threshold. We illustrate this using
266 the time-series data in Fig. 4, which highlight the unusual behavior of the model and to assess
267 whether the extra green-up period during the dry season had been eliminated in any of the
268 parametric permutations. We ascribe the lack of a clear parametric signal to two effects. First, the
269 LAI3g data were necessarily aggregated to monthly values, meaning that the primarily sub-monthly
270 variation between ensemble members was masked. Second, the timing of the secondary leaf-on
271 period in the dry season was the emergent property of the oscillatory (and thus somewhat chaotic)
272 dynamics of the soil-vegetation feedback on soil moisture.

273 Once we determined that we could not eliminat the dry season green-up by changing the
274 existing model parameters, we considered four possible additions to the model. The first three are
275 described here but, for brevity, are not quantified in the results. First, we considered that using the
276 third soil layer in CLM may be an arbitrary choice of soil depth, and that usage of the soil moisture
277 potential derived drought index ('BTRAN', (Oleson et al., 2013)), which is weighted by vertical root
278 fraction across the whole rooting depth profile, might provide a more physiologically relevant metric
279 and be less prone to increases due to upwards moisture diffusion in the dry season. However, since
280 the exponential root profile in the CLM weights the top soil layers (including layer 3) more strongly
281 than the lower layers with fewer roots, this metric was just as prone to increasing water potential
282 during the dry season as soil water potential in the third soil layer.

283 Second, we implemented leaf onset as function of a total column soil moisture content
284 threshold rather than soil moisture potential. We postulated that the redistribution of water causes
285 the erroneous behavior and that this would not impact total column moisture. However, the
286 establishment a single global threshold for total soil moisture is challenging, as a number of different
287 variables impact soil moisture, including the variation in soil water retention capacities between
288 different land points, and by the interaction between leaf area, evaporation rate and deep soil
289 moisture content. Variation in rainfall and evaporation rates affects the equilibrium water content of
290 deep soils, which changes the total column soil moisture content between locations and years, but
291 not the physiologically relevant upper soil moisture potential. Therefore, we abandoned this metric
292 of phenological trigger.

293 Third, we considered a metric of triggering leaf flush by the rate of change of total column
294 soil moisture, rather than soil moisture potential. However, this methodology also generates
295 erroneous behavior, on account of the ability of the CLM hydrology model to extract water from the
296 water table or aquifer along a water potential gradient. Thus, when water potential is low in the

297 bottom soil layer in the dry season, the rate of change of total soil moisture can be positive without
298 any input from rainfall.

299 2.5 Rainfall Model

300 To correct biases uncovered in the model output (described below) we introduced a simple
301 trigger into the model, that time-averaged 10-day precipitation must exceed a given threshold before
302 leaf onset is triggered. This approach requires the addition of a new parameter, *rain_threshold*, into the
303 model, which is the threshold over which ~~10-day precipitation~~ the sum of precipitation over 10 days
304 must be for leaf-on to occur. Leaf onset is thus triggered if 10 day rain is higher than *rain_threshold*
305 and if the SWI is greater than 15 days ~~Ψ_{days}~~ .

306 We then used a Latin hypercube approach again to determine the sensitivity of the model to
307 *rain_threshold* at our six chosen geographical points. We considered a range of rainfall rates, requiring
308 that it rain 20 mm over the course of 5 to 60 days in order for plants to begin growing
309 leaves. Requiring that it rain 20 mm over the course of 10 days in order for plants to begin growing
310 leaves best fit the LAI3g data in our six points. To test the global impact of these parameter changes
311 we ran CLM~~4.5BGC~~ with the new rainfall-based trigger and compared the results both at several
312 points and globally.

313 2.6 Global Simulations

314 We used a number of different metrics to globally compare CLM~~4.5BGC~~ to the LAI3g data
315 and, later, to the modified version of the model (CLM-MOD). First we compared maps of
316 maximum annual LAI and differences between the three maps. We also developed an algorithm to
317 count the number of LAI peaks per year in all three data sets on grid_cells with a range greater than
318 one, by counting the number of times per year that the difference between one month's LAI and the
319 next was negative, then taking the mode across all 29 years. Finally, we calculated the coefficient of

320 determination (R^2) in each grid_cell, comparing the monthly LAI3g data to CLM4.5BGG and CLM-
321 MOD to identify areas with strong agreement between the remotely sensed data and the models,
322 and areas with weak relationships.

323 The recent focus on land model benchmarking has led to a number of additional suggested
324 methods for assessing seasonality in models compared to data (e.g. Randerson *et al.*, 2009, Kelley *et*
325 *al.*, 2013), however, none of the proposed metrics would capture the central issue addressed in this
326 paper – model output with two or more peaks per year, data with only one – as they begin with the
327 unstated assumption that seasonality is unimodal over the course of a year, as do measures of the
328 start and end of the growing season. In Randerson *et al.* (2009) seasonality is assessed by identifying
329 the month of peak LAI and comparing that to MODIS LAI (MOD15A2), and in Kelley *et al.* (2014)
330 several more complicated metrics are introduced (equations 7-9) to again produce single numbers to
331 compare a model’s seasonality to a benchmark data set. In these examples, as in other benchmarking
332 studies, the focus is on producing a single number, which, while useful, can miss important details.

333

334 3. Results

335 3.1 Seasonal Patterns in CLM

336 We found generally good agreement between LAI3g and CLM averaged across grass-
337 dominated regions. In a comparison of monthly values from 1982 to 2010 for the single PFT
338 dominated regions in **Fig. 1**, R^2 values ranged from 0.54 to 0.9 (**Table 2**) with the majority of the
339 grass R^2 greater than 0.7. **Figure 2** shows the monthly values across all years, and we see similar
340 results – generally good correspondence, especially in seasonal pattern, between LAI3g and the CLM
341 runs in the grass-dominated regions. The root mean squared error (RMSE) values in **Table 2**, as

342 well as **Fig. 2B** and **C** show that CLM does not always capture the appropriate LAI values in
343 grasslands, but the seasonal cycle is reasonably correct.

344 In contrast, CLM does not successfully capture phenological patterns or values in areas
345 dominated by woody drought deciduous vegetation. Among tropical deciduous trees CLM
346 predicted LAI appears to be both too high and out of phase with the satellite observations (**Fig. 2E**)
347 while CLM shows no apparent seasonality among deciduous shrubs in the southern hemisphere
348 (**Fig. 2F**), while LAI3g shows a slight cycle ranging from 0.4 to 0.7 LAI.

349 3.2 Point simulations & sensitivity tests

350 To look more closely at seasonal patterns in drought deciduous locations we selected six
351 points around the globe across a range of latitudes dominated by a mixture of broadleaf deciduous
352 tropical trees, C3 and C4 grasses (**Table 1**) (all of which use the same stress deciduous phenology
353 algorithm). To better understand the phenological patterns, we re-ran CLM **globally** using the same
354 methods as described above but recording daily outputs of relevant parameters (including LAI, soil
355 water potential, rainfall, and others). Plots of the seasonal cycles at these specific points using daily
356 model output (solid green lines in **Fig. 3**) revealed **an apparent** pattern whereby CLM appears to put
357 leaves on during the “brown season” in the LAI3g data in some of the points in addition to during
358 the LAI3g green season. We note, however, that some areas in reality do have two separate growing
359 seasons per year (e.g. **Fig. 3E**). Despite the lack of rainfall, soil water potential in the third soil layer
360 **in CLM** rises during the dry season and is extremely variable in the dry season, on account of
361 periods of high transpiration when plants leaf out (blue dot-dashed line in **Fig. 3**).

362 We used the output from the Latin hypercube approach at these six points to vary the
363 parameters of interest (days to onset/offset, critical soil water potential, carbon assimilation) to
364 assess whether modification of parameter values could ameliorate the problem of plants leafing out

365 during the dry season in CLM. We found, however, that simply varying the parameters of the
366 existing model within the parameter space investigated (and assuming no large non-linearities in the
367 model response surface) did not remove the dry season leaf out in the model (**Fig. 4**).

368 In order to address this issue, we considered a number of structural perturbations to the leaf-
369 on and leaf-off algorithms (described in the discussion below), but ultimately decided on adding a
370 new parameter, *rain_threshold*, to the model. We then used the same Latin hypercube approach to
371 determine the best fitting values for this parameter (**Fig. 5**). This additional leaf-on criterion, set so
372 that 20 mm of rain must accumulate over 10 days in order for leaf onset to occur, led to a removal
373 of the “brown season” leaf out in CLM (dashed green line in **Fig. 3**) without preventing two green
374 seasons per year, as is possible in some semi-arid regions (e.g. parts of Ethiopia, **Fig. 3E**). While this
375 new rainfall threshold improved model performance both at our points and globally (see below), we
376 note that the model did not appear to be particularly sensitive to the amount of rain that fell, as long
377 as some rain did fall, but this threshold, and the drought deciduous algorithm as a whole, deserves
378 more research into seasonal drivers.

379 3.3 Global simulations

380 To test how well the additional rainfall parameter performed globally, we ran CLM with the
381 new rainfall parameter for 45 years (CLM-MOD) from the same equilibrium baseline state as was
382 used in the first run described. Measures of maximum LAI (**Fig. 6**) and mean LAI (data not shown)
383 in CLM-MOD showed closer matches to LAI3g than CLM4.5BGC. While CLM values remain far
384 too high in the evergreen tropics, the maximum LAI values in deciduous savanna regions did
385 increase appropriately in CLM-MOD to better match the LAI3g data in deciduous savanna regions.

386 To test whether the poor fit between CLM and LAI3g was due to multiple annual LAI peaks
387 in CLM we counted the number of peaks per year in each data set (**Fig. 7**). We found that in the

388 observations, only areas in the humid tropics had multiple peaks in the LAI3g data (“peaks” in these
389 cases being relatively small fluctuations), while CLM showed multiple peaks per year throughout
390 many of the savanna regions of the world. CLM-MOD has more areas with only one peak,
391 particularly in Sub-Saharan Africa. To quantify these changes to the model we constructed
392 confusion matrices to compare the peak counts in LAI3g to those in CLM and CLM-MOD (Table
393 3) for grid cells with >50% drought deciduous cover (Fig. 1). Overall, CLM-MOD had a slightly
394 poorer performance, matching the number of peaks in the LAI3g dataset 42.5% of the time, while
395 CLM matched LAI3g 43.7% of the time. However, these unweighted summary numbers mask
396 improvements in CLM-MOD. CLM only correctly predicted a single peak 8.9% of the time, while
397 CLM-MOD correctly predicted single peaks 59% of the time, and never did CLM-MOD predict
398 more than two peaks in a year, matching the LAI3g data. The overall degradation in CLM-MOD is
399 due to fewer correctly identified grid cells with zero or two peaks.

400 We compared monthly data and mapped the point-wise coefficients of determination (R^2)
401 globally to consider how well CLM LAI seasonality matched the LAI3g dataset (**Fig. 8A**). There
402 were moderate to good relationships ($R^2 > 0.4$) in the higher latitudes for the standard model, but
403 notably poorer relationships in the lower latitudes, particularly in savanna regions. In contrast, a
404 comparison between LAI3g and CLM-MOD showed improvements in savanna regions, with the
405 most dramatic improvements in Sub-Saharan Africa (**Fig. 8B**). Predicting the phenology of the
406 Brazilian Cerrado continues to be a challenge in CLM-MOD, shifting from two peaks in
407 CLM4.5BGC to no peaks in CLM-MOD, but the heterogeneity revealed in the LAI3g dataset
408 suggests that this region may need closer consideration and a separate phenology algorithm.

409

410 4. Discussion

411 4.1 Comparing LAI3g to CLM LAI

412 By comparing the satellite LAI3g data to output from CLM4.5BGC we found that while the
413 model performed reasonably well in temperate grasslands both in seasonal cycle and in magnitude, it
414 performed poorly in areas of tropical grasslands, mixed grassland and drought deciduous trees
415 (savannas) and areas dominated by drought deciduous trees. Closer examination of individual grid
416 cells in tropical regions revealed that these points often have a leaf flush during the dry season in the
417 model, which is not the case in the satellite data, or in reality. This additional leaf flush not only
418 impacts the phenological cycle, but also affects the overall amount of carbon stored in plants and
419 their maximum LAI, as plants spend their stored carbon unnecessarily in the dry season, leaving less
420 carbon available during the wet season for growing leaves. This addition of leaf carbon in the dry
421 season also may affect the fire cycle in varying ways around the dry tropics. While these runs of the
422 model were not coupled to a dynamic atmosphere, we expect that this dry season leaf flush could
423 also impact the climate, potentially having an unrealistic cooling effect by moving more water in to
424 the atmosphere during what should be a very dry time of year, but also darkening the land surface,
425 possibly leading to a slight warming.

426 The mechanism behind the dry season leaf flush is an increase in soil water potential in the
427 dry season to levels above the prescribed leaf-out threshold. These increases derive from the
428 assumption that all of the land surface in CLM sits on top of an unconfined aquifer. In most cases
429 this aquifer is either irrelevant because plenty of soil water is available or it is essential to plant
430 survival in areas where aquifers do exist in the real world. In semi-arid systems, however, this extra
431 pool of soil water becomes problematic in the dry season. The top soil layers dry out due to soil
432 evaporation and, when plants are active, evapotranspiration, establishing a water potential gradient,
433 which causes water to be transferred by mass flow from the aquifer up through the soil column to
434 the shallow soil layers, until eventually the moisture potential reaches the trigger for plants to leaf

435 out. Per the drought deciduous phenology algorithm, once leaf out is triggered it must be completed,
436 so plants begin to grow leaves but then the increased evapotranspiration rate quickly draws the soil
437 moisture down below leaf off threshold levels, so leaf drop begins again, typically as soon as the leaf
438 out period (30 days) has ended. The degree to which aquifers in reality contribute to dry season
439 evapotranspiration is largely unconstrained because there are no global data sets for depth to water
440 table, making it impossible to non-arbitrarily define where plants should have access to ground water
441 and where they should not. Refinements of the soil water algorithms in CLM and access to new data
442 sources like the NASA Soil Moisture Active Passive mission (SMAP; Entekhabi *et al.*, 2014) will
443 likely improve this part of the model, but like many aspects of the CLM, more global-scale data is
444 needed.

445 4.2 Soil water and rainfall in CLM

446 To address the erroneous dry season leaf flush we tested a number of different model
447 alterations, beginning with the least invasive – adjusting existing parameters – and ending with
448 adding an additional rule to the ~~existing~~ drought deciduousness algorithm. We experimented with
449 four alternative methodologies for triggering leaf onset, [described in the methods \(section](#)
450 [2.4\)described below](#), but for brevity we have only shown results from the last and most effective
451 approach.

452 ~~First, we considered that using the third soil layer in CLM may be an arbitrary choice of soil~~
453 ~~depth, and that usage of the soil moisture potential derived drought index ('BTRAN', (Oleson *et al.*,~~
454 ~~2013)), which is weighted by vertical root fraction across the whole rooting depth profile, might~~
455 ~~provide a more physiologically relevant metric and be less prone to increases due to upwards~~
456 ~~moisture diffusion in the dry season. However, since the exponential root profile in the CLM~~
457 ~~weights the top soil layers (including layer 3) more strongly than the lower layers with fewer roots,~~

458 ~~this metric was just as prone to increasing water potential during the dry season as soil water~~
459 ~~potential in the third soil layer.~~

460 ~~Second, we implemented leaf onset as function of a total column soil moisture content~~
461 ~~threshold rather than soil moisture potential. We postulated that the redistribution of water causes~~
462 ~~the erroneous behavior and that this would not impact total column moisture. However, the~~
463 ~~establishment a single global threshold for total soil moisture is challenging, as a number of different~~
464 ~~variables impact soil moisture, including the variation in soil water retention capacities between~~
465 ~~different land points, and by the interaction between leaf area, evaporation rate and deep soil~~
466 ~~moisture content. Variation in rainfall and evaporation rates affects the equilibrium water content of~~
467 ~~deep soils, which changes the total column soil moisture content between locations and years, but~~
468 ~~not the physiologically relevant upper soil moisture potential. Therefore, we abandoned this metric~~
469 ~~of phenological trigger.~~

470 ~~Third, we considered a metric of triggering leaf flush by the rate of change of total column~~
471 ~~soil moisture, rather than soil moisture potential. However, this methodology also generates~~
472 ~~erroneous behavior, on account of the ability of the CLM hydrology model to extract water from the~~
473 ~~water table or aquifer along a water potential gradient. Thus, when water potential is low in the~~
474 ~~bottom soil layer in the dry season, the rate of change of total soil moisture can be positive without~~
475 ~~any input from rainfall.~~

476 ~~These hydrological issues in CLM are complex, and derive from the need to operate an~~
477 ~~internally consistent global model of the water cycle in the absence of critical data at the appropriate~~
478 ~~scale (depth to water table, the unsaturated hydraulic conductivity of deep soils, etc). In an ideal~~
479 ~~case, improvements in hydrology might allow the existing phenological model to operate correctly.~~

480 ~~However, here we took a more pragmatic approach and so for a fourth method we partially~~

481 decoupled the soil hydrology and the phenology models, allowing rainfall inputs to directly impact
482 on leaf phenology without interacting with the assumptions of the [soil-hydrology](#) model. Leaving the
483 condition of soil water potential in the third soil layer in place, we then added an additional
484 condition which was that the rainfall accumulated over the last 10 days should be higher than a
485 threshold value (20 mm). Thus, if soil moisture rose above the threshold level, but little or no rain
486 had fallen, plants would not put on leaves. The new model performs better both for the point
487 simulations and in global simulations, both in terms of the seasonal cycle of LAI, where the average
488 point-wise coefficient of determination (R^2) between modelled and observed monthly satellite LAI
489 of drought deciduous dominated points (>50% drought deciduous cover) is significantly higher for
490 the new model (0.31 vs 0.13). [While there was no substantial change in the overall peak count](#)
491 [accuracy \(Table 3\), CLM-MOD had zero drought deciduous dominated grid cells with >2 peaks and](#)
492 [a substantial improvement in the identification of single-peak grid cells \(8.9 to 59%\). The added](#)
493 [rainfall trigger did, however, reduce the number of zero peak and two peak grid cells correctly](#)
494 [identified. This result highlights the need for more research into the diversity of drought deciduous](#)
495 [phenology drivers around the world.](#)

496 4.3 Impacts of modifications to the model

497 This relatively small change to the drought deciduous phenology algorithm had wide ranging
498 impacts within [the](#) CLM. Because carbon was not being unnecessarily spent to grow leaves during
499 the dry season, which was then not replenished since there was not enough water to maintain
500 photosynthesis, CLM-MOD showed substantially higher overall carbon stores in savanna regions
501 (**Fig. 10**, blue lines). Over time, this increase in vegetation carbon could lead to more realistic soil
502 carbon levels, which have been shown to be too low in savanna regions in CLM (Wieder et al.,
503 2013).

504 Fire is a critical component of savanna ecology and has been a focus of recent efforts to
505 improve the CLM (Li et al., 2014). Our change to the drought deciduous phenology algorithm does
506 have an-slight impact on the fire cycle, but unfortunately, though not surprisingly, it degrades the fire
507 model's performance relative to a global fire data set (GFED4; Giglio *et al.*, 2013). Comparing the
508 average total annual fire fractions for each grid_cell with drought deciduous cover greater than 50%
509 across the time period for which we have both GFED4 data and CLM output (1996-2010) we find a
510 correlation between GFED4 and CLM 4.5BGC of 0.35 (global correlation = 0.44), and a correlation
511 of 0.23 with CLM-MOD (global correlation = 0.33). This degradation of fire model performance is
512 not surprising, however, given that the fire model was developed using CLM4.0CN with the
513 erroneous dry season green up and a different forcing dataset. As shown in **Fig. 9**, fraction of area
514 burned per grid_cell decreases in many areas in CLM-MOD, likely due to the fact that less fuel is
515 being produced in the dry season, and seasonality shifts as well (**Fig. 10**, red lines). Future work will
516 include exploring the impacts of this change to CLM on fire and other ecosystem properties.

517 Two other outstanding questions about LAI in CLM remain. First, in savanna regions in the
518 CLM LAI drops to zero during the dry season, implying that across an entire grid_cell all vegetation
519 is perfectly drought deciduous (**Figs. 3-6**). Reality is, of course, far more complex, as reflected in the
520 LAI3g dataset which rarely drops below 1.0 in savanna regions when aggregated to the CLM grid.
521 The focus of this study was on improving the timing and magnitude of peak LAI, however,
522 improving dry season values is also a concern. This is a deeper question in the CLM, as it relates to
523 the overall land cover data. It is possible, for example, that there is evergreen vegetation in these grid
524 cells, while the land cover classification determines that all grid_cells with a significant seasonal LAI
525 signal are 100% 'drought deciduous'. Even if drought deciduous and evergreen vegetation types did
526 not co-exist in the same ecosystem type, within grid_cell spatial heterogeneity might also allow their
527 coexistence within a whole grid cell (e.g. riparian areas or areas with shallow ground water that are

528 able to stay green year round). Second, though also not the focus of this study, it is clear from **Fig. 6**
529 that across the mesic regions of the terrestrial biosphere CLM is dramatically overestimating LAI.
530 This issue, often masked when only mean annual values or zonal means are considered, deserves
531 more attention, and it is likely that recent detailed studies of carbon allocation (e.g. Doughty *et al.*,
532 2014) could improve this part of the model.

533 A question also remains as to whether our new representation of leaf phenology, in spite of
534 its improved performance, constitutes a better predictive model of current and future ecosystem
535 behavior. In general, we hope to construct ecosystem models that represent hypotheses of how
536 plants function, that we might test against observations. In this case, we find that the existing
537 hypothesis - that plants respond to the soil water potential of the upper soil - does not adequately
538 represent phenological patterns. However, this explanation is complicated because the predictions
539 depend also on the properties of soil water in the model. Given a perfect representation of soil
540 moisture, we might find that the existing leaf-on hypothesis is a good approximation of average
541 vegetation behavior. However, at present the coupling of these two complex systems produces
542 unexpected results. By tying the vegetation behavior to the actual climate drivers we are reducing the
543 complexity of the problem, however, we are also reducing the capacity of the model to be
544 responsive to the nuances of climate drivers. For example, the same rainfall amount in high and low
545 humidity regimes will have different impacts on net soil moisture.

546 Ideally, models should represent the mechanisms by which ecological processes operate in as
547 much fidelity as we understand. The representation of drought phenology is interesting; however, as
548 we suspect that there are many different phenological strategies in the tropics that the CLM classifies
549 with the same algorithm (e.g. Archibald & Scholes, 2007). This means that, in the absence of the
550 representation of these numerous phenological strategies in the model, we are really representing
551 the net behavior of ecosystems, rather than the exact mechanisms pertaining to a single species. The

552 fact that CLM-MOD improved model performance most significantly in Africa and less so in
553 Australia and South America by some metrics (**Fig. 7**) suggests that evolutionary differences
554 between plants could play a significant role in determining phenological patterns between
555 continents. In a higher-fidelity land surface model, we might ideally allow numerous phenological
556 algorithms to compete for light and water resources, and the ecosystem LAI profile would reflect
557 the net behavior of the successful algorithms. This type of modelling is now theoretically possible
558 (e.g. Fisher et al., 2010, 2015), and will be investigated in future versions of the CLM.

559

560 **5. Conclusions**

561 By comparing satellite derived estimates of LAI to LAI values produced by the latest version
562 of the CLM we revealed a small but significant issue in the CLM – the tendency for leaves to flush
563 during the dry season in drought deciduous PFTs due to unrealistic upwards movement of water
564 through the soil column. We tested a number of different approaches to address this issue, however
565 we found that tying leaf flushing to rainfall directly produced results that best-better matched the
566 satellite data. While this change to the drought deciduous phenology algorithm does not reflect our
567 understanding of how plants respond to their environment in the real world, without better data on
568 soil water movement at scales relevant to global land surface modeling it is difficult to rely on the
569 soil water model to drive plant physiology. Changing the drought deciduous phenology algorithm to
570 remove dry season leaf flushes improved overall LAI values in savanna systems as well as changed
571 the amount of carbon stored in these systems and altered the fire cycle. We also emphasize that this
572 issue would have been impossible to detect with a standard ‘benchmarking’ type of metric for
573 measuring seasonality and was difficult to identify until daily model outputs were reported and
574 analyzed (i.e. Fig. 3). Future work will include exploring different drought deciduous phenology

575 algorithms for different PFTs and testing the importance of this change in a coupled Earth system
576 model.

577

578 **Acknowledgements**

579 The authors thank the members of the Terrestrial Sciences Section at NCAR for helpful
580 discussions of this work, and we thank Dr. Ranga Myneni and his group for providing the LAI3g
581 dataset. KMD was funded by an Advanced Study Program Postdoctoral Fellowship at NCAR. We
582 would also like to acknowledge high-performance computing support from Yellowstone
583 ([ark:/85065/d7wd3xhc](https://doi.org/10.7554/85065/d7wd3xhc)) provided by NCAR's Computational and Information Systems Laboratory,
584 sponsored by the National Science Foundation.

585

586 **REFERENCES**

- 587 Archibald, S. and Scholes, R. J.: Leaf green-up in a semi-arid African savanna – separating tree and
588 grass responses to environmental cues, *J. Veg. Sci.*, 18, 583–594, 2007.
- 589 Bivand, R., Keitt, T. and Rowlingson, B.: *rgdal*: Bindings for the geospatial data abstraction library,
590 [online] Available from: <http://cran.r-project.org/package=rgdal>, 2013.
- 591 Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M. and Sitch, S.: A
592 comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and
593 carbon at both the global and seasonal scale, *Geosci. Model Dev.*, 4(2), 255–269, doi:10.5194/gmd-
594 4-255-2011, 2011.
- 595 Bonan, G. B., Levis, S., Sitch, S., Vertenstein, M. and Oleson, K. W.: A dynamic global vegetation
596 model for use with climate models : concepts and description of simulated vegetation dynamics,
597 *Glob. Chang. Biol.*, 9, 1543–1566, doi:10.1046/j.1529-8817.2003.00681.x, 2003.
- 598 Bradley, A. V., Gerard, F. F., Barbier, N., Weedon, G. P., Anderson, L. O., Huntingford, C., Aragão,
599 L. E. O. C., Zelazowski, P. and Arai, E.: Relationships between phenology, radiation and
600 precipitation in the Amazon region, *Glob. Chang. Biol.*, 17(6), 2245–2260, doi:10.1111/j.1365-
601 2486.2011.02405.x, 2011.
- 602 Delbart, N., Le Toan, T., Kergoat, L. and Fedotova, V.: Remote sensing of spring phenology in
603 boreal regions: A free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–
604 2004), *Remote Sens. Environ.*, 101(1), 52–62, doi:10.1016/j.rse.2005.11.012, 2006.
- 605 Doughty, C. E., Malhi, Y., Arujo-Murakami, A., Metcalfe, D. B., Silva-Espejo, J. E., Arroyo, L.,
606 Heredia, J. P., Pardo-Toledo, E. and Mendizabal, L. M.: Allocation trade-offs dominate the response
607 of tropical forest growth to seasonal and interannual drought, , 95(8), 2192–2201, 2014.
- 608 Entekhabi, D., Yueh, S., O’Neill, P. E., Kellogg, K. H., Allen, A., Bindlish, R., Brown, M., Chan, S.,
609 Colliander, A. and Crow, W. T.: *SMAP Handbook*, 2014.
- 610 Fisher, R. a., Muszala, S., Vertenstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G.,
611 Koven, C., Holm, J., Rogers, B. M., Lawrence, D. and Bonan, G.: Taking off the training wheels: the
612 properties of a dynamic vegetation model without climate envelopes, *Geosci. Model Dev. Discuss.*,
613 8(4), 3293–3357, doi:10.5194/gmdd-8-3293-2015, 2015.
- 614 Fisher, R., McDowell, N., Purves, D., Moorcroft, P., Sitch, S., Cox, P., Huntingford, C., Meir, P. and
615 Ian Woodward, F.: Assessing uncertainties in a second-generation dynamic vegetation model caused
616 by ecological scale limitations, *New Phytol.*, 187, 666–681, doi:10.1111/j.1469-8137.2010.03340.x,
617 2010.
- 618 Giglio, L., Randerson, J. T. and Van Der Werf, G. R.: Analysis of daily, monthly, and annual burned
619 area using the fourth-generation global fire emissions database (GFED4), *J. Geophys. Res.*
620 *Biogeosciences*, 118(November 2012), 317–328, doi:10.1002/jgrg.20042, 2013.

621 Guan, K., Wood, E. F., Medvigy, D., Kimball, J., Pan, M., Caylor, K. K., Sheffield, J., Xu, X. and
622 Jones, M. O.: Terrestrial hydrological controls on land surface phenology of African savannas and
623 woodlands, *J. Geophys. Res. Biogeosciences*, 119, 1652–1669, doi:10.1002/2013JG002572. Received,
624 2014.

625 Hijmans, R. J. and van Etten, J.: raster: Geographical data analysis and modeling, [online] Available
626 from: <http://cran.r-project.org/package=raster>, 2013.

627 Jenerette, G. D., Scott, R. L. and Huete, A. R.: Functional differences between summer and winter
628 season rain assessed with MODIS-derived phenology in a semi-arid region, *J. Veg. Sci.*, 21(1), 16–30,
629 doi:10.1111/j.1654-1103.2009.01118.x, 2010.

630 Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B. and Willis, K. O.: A
631 comprehensive benchmarking system for evaluating global vegetation models, *Biogeosciences*, 10(5),
632 3313–3340, doi:10.5194/bg-10-3313-2013, 2013.

633 Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B.,
634 Lawrence, D. M. and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and
635 alternate soil C and N models on C dynamics of CLM4, *Biogeosciences*, 10(11), 7109–7131,
636 doi:10.5194/bg-10-7109-2013, 2013.

637 Lau, W. K.-M., Wu, H.-T. and Kim, K.-M.: A canonical response of precipitation characteristics to
638 global warming from CMIP5 models, *Geophys. Res. Lett.*, 40(12), 3163–3169,
639 doi:10.1002/grl.50420, 2013.

640 Lawrence, D. M., Oleson, K. W., Flanner, M. G., Fletcher, C. G., Lawrence, P. J., Levis, S.,
641 Swenson, S. C. and Bonan, G. B.: The CCSM4 Land Simulation, 1850–2005: Assessment of Surface
642 Climate and New Capabilities, *J. Clim.*, 25(7), 2240–2260, doi:10.1175/JCLI-D-11-00103.1, 2012.

643 Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J.,
644 Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B. and Slater, A. G.: Parameterization
645 improvements and functional and structural advances in Version 4 of the Community Land Model,
646 *J. Adv. Model. Earth Syst.*, 3(3), M03001, doi:10.1029/2011MS000045, 2011.

647 Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the
648 Community Land Model (CLM 3.0), *J. Geophys. Res.*, 112(G1), G01023,
649 doi:10.1029/2006JG000168, 2007.

650 Li, F., Bond-Lamberty, B. and Levis, S.: Quantifying the role of fire in the Earth system – Part 2:
651 Impact on the net carbon balance of global terrestrial ecosystems for the 20th century,
652 *Biogeosciences*, 11(5), 1345–1360, doi:10.5194/bg-11-1345-2014, 2014.

653 Ma, X., Huete, A., Yu, Q., Coupe, N. R., Davies, K., Broich, M., Ratana, P., Beringer, J., Hutley, L.
654 B., Cleverly, J., Boulain, N. and Eamus, D.: Spatial patterns and temporal dynamics in savanna
655 vegetation phenology across the North Australian Tropical Transect, *Remote Sens. Environ.*, 139,
656 97–115, doi:10.1016/j.rse.2013.07.030, 2013.

- 657 McKay, M. D., Beckman, R. J. and Conover, W. J.: Comparison of Three Methods for Selecting
658 Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21(2),
659 239–245, doi:10.1080/00401706.1979.10489755, 1979.
- 660 Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C., Levis, S., Li,
661 F., Riley, W., Subin, Z., Swenson, S., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C., Kluzek, E.,
662 Lamarque, J.-F., Lawrence, P., Leung, L., Lipscomb, W., Muszala, S., Ricciuto, D., Sacks, W., Sun,
663 Y., Tang, J. and Yang, Z. L.: Technical description of version 4.5 of the Community Land Model
664 (CLM), NCAR Tech. Note, 503+STR(June), doi:10.5065/D6RR1W7M, 2013.
- 665 Parmesan, C. and Yohe, G.: A globally coherent fingerprint of climate change impacts across natural
666 systems., *Nature*, 421(6918), 37–42, doi:10.1038/nature01286, 2003.
- 667 Pierce, D.: ncdf: Interface to unidata netCDF files, [online] Available from: [http://cran.r-](http://cran.r-project.org/package=ncdf)
668 [project.org/package=ncdf](http://cran.r-project.org/package=ncdf), 2011.
- 669 Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y.-H.,
670 Nevison, C. D., Doney, S. C., Bonan, G., Stöckli, R., Covey, C., Running, S. W. and Fung, I. Y.:
671 Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, *Glob.*
672 *Chang. Biol.*, 15(10), 2462–2484, doi:10.1111/j.1365-2486.2009.01912.x, 2009.
- 673 RCoreTeam: R: A language and environment for statistical computing, 2013.
- 674 Reed, B. C., Schwartz, M. D. and Xiao, X.: Remote sensing phenology: Status and the way forward,
675 in *Phenology of Ecosystem Processes: Applications in global change research*, edited by A.
676 Noormets, pp. 231–246, Springer., 2009.
- 677 Reich, P. B.: Phenology of tropical forests: Patterns, causes, and consequences, *Can. J. Bot.*, 73, 164–
678 174, 1995.
- 679 Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M.,
680 Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant,
681 R., Hollinger, D. Y., Margolis, H. a., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W.,
682 Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R.,
683 Verbeeck, H., Xiao, J. and Xue, Y.: Terrestrial biosphere models need better representation of
684 vegetation phenology: results from the North American Carbon Program Site Synthesis, *Glob.*
685 *Chang. Biol.*, 18(2), 566–584, doi:10.1111/j.1365-2486.2011.02562.x, 2012.
- 686 Scholes, R. J. and Hall, D. O.: The carbon budget of tropical savannas, woodlands, and grasslands,
687 in *SCOPE 56 - Global Change: Effects on Coniferous forests and grasslands*, edited by A. I.
688 Breymer, D. O. Hall, J. M. Melillo, and G. I. Agren, Wiley., 1996.
- 689 Staver, A. C., Archibald, S. and Levin, S. A.: The global extent and determinants of savanna and
690 forest as alternative biome states., *Science*, 334(6053), 230–2, doi:10.1126/science.1210465, 2011.

691 Swenson, S. C. and Lawrence, D. M.: Assessing a dry surface layer-based soil resistance
692 parameterization for the Community Land Model using GRACE and FLUXNET-MTE data, *J.*
693 *Geophys. Res. Atmos.*, 119(10), 299–312, doi:10.1002/2014JD022314. Received, 2014.

694 Wang, K., Mao, J., Dickinson, R., Shi, X., Post, W., Zhu, Z. and Myneni, R.: Evaluation of CLM4
695 Solar Radiation Partitioning Scheme Using Remote Sensing and Site Level FPAR Datasets, *Remote*
696 *Sens.*, 5(6), 2857–2882, doi:10.3390/rs5062857, 2013.

697 White, M. a., de BEURS, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P.,
698 O’Keefe, J., Zhang, G., Nemani, R. R., van LEEUWEN, W. J. D., Brown, J. F., de WIT, A.,
699 Schaepman, M., Lin, X., Dettinger, M., Bailey, A. S., Kimball, J., Schwartz, M. D., Baldocchi, D. D.,
700 Lee, J. T. and Lauenroth, W. K.: Intercomparison, interpretation, and assessment of spring
701 phenology in North America estimated from remote sensing for 1982-2006, *Glob. Chang. Biol.*,
702 15(10), 2335–2359, doi:10.1111/j.1365-2486.2009.01910.x, 2009.

703 White, M. a., Thornton, P. E. and Running, S. W.: A continental phenology model for monitoring
704 vegetation responses to interannual climatic variability, *Global Biogeochem. Cycles*, 11(2), 217–234,
705 doi:10.1029/97GB00330, 1997.

706 Wieder, W. R., Bonan, G. B. and Allison, S. D.: Global soil carbon projections are improved by
707 modelling microbial processes, *Nat. Clim. Chang.*, 3(10), 909–912, doi:10.1038/nclimate1951, 2013.

708 Yang, X., Mustard, J. F., Tang, J. and Xu, H.: Regional-scale phenology modeling based on
709 meteorological records and remote sensing observations, *J. Geophys. Res.*, 117(G3), G03029,
710 doi:10.1029/2012JG001977, 2012.

711 Zhang, X., Friedl, M. A., Schaaf, C. B. and Strahler, A. H.: Monitoring the response of vegetation
712 phenology to precipitation in Africa by coupling MODIS and TRMM instruments, *J. Geophys. Res.*,
713 110(D12), D12103, doi:10.1029/2004JD005263, 2005.

714 Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. and Myneni,
715 R.: Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically
716 Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies
717 (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2, *Remote*
718 *Sens.*, 5(2), 927–948, doi:10.3390/rs5020927, 2013.

719

720

721 **Table 1.** List of locations for point simulations and percent cover of plant functional types (PFTs).
 722 PFTs with no coverage at any point are not listed.

| point.name | Latitude | Longitude | Bare ground | Broadleaf Evergreen Tree Tropical | Broadleaf Evergreen Tree Temperate | Broadleaf Deciduous Tree Tropical | C3 Grasses | C4 Grasses | Crops |
|------------------|----------|-----------|-------------|-----------------------------------|------------------------------------|-----------------------------------|------------|------------|-------|
| Brasilia | -15 | -51 | 0.46 | 1.69 | 0 | 16.52 | 8.83 | 62.35 | 10.15 |
| western Brazil | -6 | -39 | 2.66 | 0 | 0 | 35.4 | 9.04 | 40.6 | 12.3 |
| South Chad | 11 | 18 | 1.34 | 0 | 0 | 34.26 | 0.03 | 60.22 | 4.16 |
| eastern Zambia | -13 | 32 | 0.22 | 0.56 | 1.27 | 26.4 | 37.81 | 26.39 | 7.35 |
| south Ethiopia | 5.5 | 40 | 8.75 | 0.13 | 0.02 | 63.1 | 19.12 | 5.42 | 3.47 |
| Darwin Australia | -15 | 130.5 | 15.94 | 0 | 0 | 35.73 | 0 | 48.33 | 0 |

723

724 **Table 2.** R² and RMSE of AVHRR LAI_{3g} v. CLM_{4.5BGC} monthly data for all 29 yrs.

| | | NH C3 grass | NH C4 grass | tropical DT | SH DS | SH C3 grass | SH C4 grass |
|-----------------------|----------------|-------------|-------------|-------------|-------|-------------|-------------|
| CLM _{4.5BGC} | R ² | 0.79 | 0.72 | 0.08 | 0.05 | 0.41 | 0.85 |
| | RMSE | 0.15 | 0.29 | 0.18 | 0.10 | 0.27 | 0.13 |

725 * PFT = plant functional type, NH = northern hemisphere, DS = deciduous shrub, DT =
 726 deciduous tree, SH = southern hemisphere

727

728 Table 3. Confusion matrices comparing grid cell peak counts between LAI3g and the two model
 729 data sets. “%” rows and columns are the percent of the correct values (diagonal) compared to the
 730 sums for the respective rows and columns.

| <u>A. LAI3g vs. CLM</u> | | | | | | | <u>B. LAI3g vs CLM-MOD</u> | | | | | | |
|-------------------------|-----------|--------------|------------|-------------|-----------|-------------|----------------------------|------------|--------------|-------------|-------------|-----------|-------------|
| | | <u>LAI3g</u> | | | | | | | <u>LAI3g</u> | | | | |
| | | <u>0</u> | <u>1</u> | <u>2</u> | <u>≥2</u> | <u>%</u> | | | <u>0</u> | <u>1</u> | <u>2</u> | <u>≥2</u> | <u>%</u> |
| <u>CLM</u> | <u>0</u> | <u>436</u> | <u>164</u> | <u>145</u> | <u>0</u> | <u>58.5</u> | <u>CLM-MOD</u> | <u>0</u> | <u>365</u> | <u>279</u> | <u>277</u> | <u>0</u> | <u>39.6</u> |
| | <u>1</u> | <u>28</u> | <u>74</u> | <u>130</u> | <u>0</u> | <u>31.9</u> | | <u>1</u> | <u>242</u> | <u>439</u> | <u>407</u> | <u>0</u> | <u>43.2</u> |
| | <u>2</u> | <u>196</u> | <u>555</u> | <u>499</u> | <u>0</u> | <u>39.9</u> | | <u>2</u> | <u>60</u> | <u>63</u> | <u>125</u> | <u>0</u> | <u>50.4</u> |
| | <u>≥2</u> | <u>7</u> | <u>42</u> | <u>35</u> | <u>0</u> | <u>0</u> | | <u>≥ 2</u> | <u>0</u> | <u>0</u> | <u>0</u> | <u>0</u> | <u>NA</u> |
| | <u>%</u> | <u>65.4</u> | <u>8.9</u> | <u>61.7</u> | <u>NA</u> | <u>43.7</u> | | <u>%</u> | <u>54.7</u> | <u>59.0</u> | <u>15.5</u> | <u>NA</u> | <u>42.5</u> |

731

732

733

734 **FIGURE CAPTIONS**

735 **Figure 1.** (A) Areas of the globe dominated by a single PFT (>50%) grouped where appropriate
736 (e.g. there are actually three grass PFTs). (B) Percent cover of drought deciduous PFTs within the
737 natural vegetation component of each grid_cell; gray areas have zero percent cover of drought
738 deciduous vegetation. For visual clarity, grid_cells with < 50% natural vegetation (e.g. grid_cells that
739 are mostly water) are not shown in both maps.

740 **Figure 2.** Annual LAI cycles for LAI3g and CLM4.5BGC averaged for 1982-2010; shaded areas
741 represent one standard deviation. Each plot is averaged across a region as shown in Fig. 1. (A)
742 Northern hemisphere (NH) C3 grasses; (B) NH C4 grasses; (C) Southern hemisphere (SH) C3
743 grasses; (D) SH C4 grasses; (E) Tropical deciduous trees; (F) SH broadleaved deciduous shrubs.

744 **Figure 3.** Seasonal cycles of rainfall (mm day⁻¹, gray bars); leaf area index (LAI, green lines and black
745 dots) and soil water potential in the third layer (MPa, blue lines) in CLM4.5BGC and CLM-MOD
746 for one year (2001).

747 **Figure 4.** Illustration of Latin hypercube (LH) variable exploration analysis results – here each line
748 represents one simulation all from one year of the LH analysis without the additional rainfall trigger.
749 Each line is from a model run with slightly different values for the variables considered. In actuality
750 100 simulations were performed, but for visual clarity we are showing a selection of 10 simulations.

751 **Figure 5.** Illustration of Latin hypercube (LH) variable exploration analysis results as with Fig. 4 –
752 here each line represents one simulation all from one year of the LH analysis with the additional
753 rainfall trigger.

754 **Figure 6.** Maximum annual LAI averaged across the 29 year time period (1982-2010) in LAI3g,
755 CLM4.5BGC, and CLM-MOD, and the differences between these three maps

756 **Figure 7.** Mode of annual peak count analysis for the three simulations. (A) LAI3g; (B)
757 CLM4.5BGC; (C) CLM-MOD

758 **Figure 8.** Coefficients of determination (R^2) between LAI3g and the two model versions.

759 **Figure 9.** Average burned area fraction per year across the time period where data was available
760 (1996-2010) for GFED4, CLM4.5BGC, and CLM-MOD, and the differences between these maps.

761 **Figure 10.** Seasonal cycles of rainfall (mm day^{-1} ; gray bars), total vegetation carbon (gC m^{-2} ; blue
762 lines), and grid cell burned fraction (day^{-1} ; red lines) in CLM4.5BGC and CLM-MOD for one year
763 (2001).

764

765