

Below are outlined our responses to the comments from the two anonymous reviewers for our paper entitled: "*A model inter-comparison study to examine limiting factors in modelling Australian tropical savannas.*"

Reviewer comments are numbered below and given verbatim (as received), and we have answered each comment to the best of our ability. We would like to thank the reviewers and the editor for taking the time to examine this work and provide valuable feedback.

Anonymous Reviewer #1

1.1 Comment: Page 19014 line 9 -11 'For dry season LE, half the models (BIOS2, MAESPA, and SPA) were able to consistently outperform the emp2 benchmark, and come close to meeting the same number of metrics as the emp3 benchmark particularly at the drier sites.' I don't see that MAESPA fits into the described. Three out of five sites are very close to emp1 and only two are slightly better than emp2. Did you maybe mean BESS? Or MAESPA for GPP? Please clarify.

Response: The reviewer is quite correct here; we are indeed referring to the BESS model and not MAESPA. We have amended the text appropriately.

1.2 Comment: A figure including the seasonal cycle of the drivers used for the benchmarking would be nice. Sturt Plains is most likely a water limited site also during the wet season while Howard springs and Adelaide River might also be energy limited especially during the wet season. This could help to understand under which driver-conditions the BTMs are actually predicting better.

Response: We have added the suggested figure to the supplementary information and reference it in the text.

"The seasonal behaviour of the empirical benchmark drivers along the NATT can be referred to in the supplementary information"

1.3 Comment: [Figure 2] please use the same y-axis limits on all plots. Details of differences might get lost but the amplitude of the seasonality and its prediction becomes clearer between the sites.

Response: Done.

1.4 Comment: [Figure 4] correlation plots: most of the correlations are clearly not linear, as far as it can be identified from the plot. The color scheme with the faded / pale colors makes it hard to distinguish individual sites from each other. To me it seems that there are mostly individual clusters and large scatter around it. Maybe you can redo the figure with brighter colors.

Response: There is likely a misunderstanding with the sub-figures showing model residuals vs. model prediction. These sub-figures do not show correlation but a model's bias, i.e. it's propensity to over- or under-predict a flux at low and high values. This is explicitly stated in Section 3.2 of the Results in the first paragraph. We do agree that the scatter does not show any clear linearity, but the use of the trend lines is to visually help the reader see the direction of the bias as the amount of data points makes this difficult to see otherwise. Regarding the colour scheme, we opted for the *Viridis* colour map as a preventative measure, so that colour-blind readers can make out features of the plots.

1.5 Comment: Please write LPJGUESS or LPG-GUESS throughout the text including the figures.

Response: Done.

1.6 Comment: Page 19006 Line 15 TBM is introduced but not explained.

Response: The definition for TBM is in the abstract (P19001, line 7) and then again in the introduction (P19003, lines 2-3).

1.7 Comment: Page 19006 Line 28 ESMs is introduced but not explained.

Response: The abbreviation for earths system models (ESM) has been dropped, as the term appears only twice throughout the paper. In both instances, the full name is now used.

1.8 Comment: Page 19008 line 8-9 should be Donohue et al. (2009) not Donohue et al. (Donohue et al., 2009).

Response: Done.

1.9 Comment: Page 19014 line 14 '...(LPJGUESS, BESS.....' instead of '...(LPJGUESS. BESS...'

Response: Done.

1.10 Comment: Page 19016 line 27 please add 'to' between 'expected' and 'perform'

Response: Done.

1.11 Comment: Page 19017 line 12 should be...(adjustments of stomatal ...

Response: Done.

1.12 Comment: Page 19018 line 6 maybe better to write ... observations of understory LE...

Response: Done.

1.13 Comment: Page 19021 line 4 DGVMs first time introduced please explain

Response: The definition for DGVM is given in Section 2.2 of the Methods (corresponds to P19007, lines 8-9).

1.14 Comment: Page 19021 line 16-17 should be: Such formulations are therefore not

mechanistic, do not respond to actual season dynamics....

Response: Done.

Anonymous Reviewer #2

Comment 1a: Functional composition: How is tree/grass fraction determined from MODIS (for those models that use it), or from the DGVM? What are those fractions? MAESPA and SPA have time-varying tree/grass fraction (following Donohue); does that mean the other models do not? Do those other models follow the work of Hansen, Bucini, etc? What is their partitioning? Are they similar to ground-based observations?

Response: Information regarding how the C₃/C₄ fraction in each model is determined is displayed in Table 2. Regarding the validity of this method, we find that for the BIOS2, MAESPA and SPA model these fractions are in agreement with other studies such as that of Murphy and Bowman (2007) and Williams et al. (1996). For the CABLE and BESS models, these fractions will be an emergent property, and Sea et al. (2011) has shown the MODIS LAI product for these sites to be in good agreement with the ground-based observations. Although these points are covered in Section 2.3 of the Methods (Experimental protocol), we have added the following additional text to make it clearer for the reader:

“Each TBM was setup to describe a C₃ evergreen overstorey with an underlying C₄ grass understorey, and conforms well with the characteristics of savannas in Northern Australia (Bowman and Prior, 2005). All TBMs (excluding LPJGUESS) prescribed LAI as an input, to characterise the phenology of vegetation at each site. In these cases LAI was determined from MODIS derived approximations that were well matched to ground-based estimations of LAI at the SPECIAL sites (Sea et al., 2011). The fraction of C₃ to C₄ vegetation was handled differently by each model and was determined for each as follows. For MAESPA and SPA, the models allowed for time-varying tree and grass fractions to be assigned as direct inputs, and these time-varying fractions were determined using the method of Donohue et al. (2009). BIOS2 similarly used the same method to extract time-varying fractions, while CABLE used a static fraction that did not change. The BESS model derived the C₃:C₄ fraction from the C₃ and C₄ distribution map of Still et al. (2003), while for LPJGUESS this fraction is a prognostic determination resulting from the competition between trees and grasses (see Smith et al., 2001).”

Comment 1b: Section 4.3 says “(p)rescribing LAI can be problematic depending on the time-scale and how it is partitioned between tree and grass layers.” This is critical to our ability to simulate savannas. How are the models treating it, and what is our evaluation of this treatment?

Response: Referring back to the previous comment, we have added additional text to Section 2.3 of the Methods to make the model’s treatment of this clearer [quoted in the above response]. However, there is currently no data available (at least to our knowledge) to adequately evaluate whether these assigned fractions (and their evolution through time) are correct. In light of this, this shows there is a critical need for “pheno-cam” type studies and others (e.g. as Moore et al. *in review*) that can help with such evaluation.

Comment 2: Grass Phenology: We know that C4 grasses follow a seasonal cycle of wet-season greenup followed by senescence as the soil dries. Are the models consistent in their representation of mean seasonal cycles, in terms of timing/maximum LAI? Why or why not? Which model has the most realistic grass phenology based on observational data?

Response: Our results show that very few of the models are able to consistently represent the seasonality of the fluxes; whether this be in capturing the wet season peak amplitude or the wet-to-dry (and vice versa) transition periods. Those models that were able to meet the benchmarks more frequently, tended to be those models where the Donohue et al. (2009) method was used to construct the LAI/phenology input (i.e. BIOS2, MAESPA and SPA). However, we feel we would be misleading the reader in saying that a correct approximation of the C4 phenology is the major reason for capturing the underlying seasonal signal, as for example, MAESPA is unable to meet the higher benchmarks in predicting LE, despite doing so in predicting GPP due to the model being forced with prescribed C3:C4 fractions. Failure to capture seasonality is just as likely to be due to a misrepresentation of hydrological processes in the models, either through inadequate rooting depths, or assumptions on root-water extraction. We believe it is likely a combination of processes (phenology + water uptake) that are causes for failure in this type of environment and have addressed this in the discussion.

Referring back to our previous response, there are no ground-based datasets on C4 grass phenology currently available for these sites on which we can validate modelled tree and grass phenology. However, we would like to point out that the Donohue et al. (2009) method has been validated previously and approximates reasonably well with recurrent and non-recurrent vegetation cover maps, such as that of Murphy and Bowman (2007).

Comment 3a: Tree phenology: Many (most?) of the readers will not be immediately familiar with the trees from the latin. Are the trees at all sites evergreen? Deciduous? Semi- deciduous? Mixed? How are they mixed? How is this heterogeneity represented, if at all, in the models?

Response: Section 2.1 of the Methods states that the tree species at the sites along the NATT (and mentioned in Table 1) are evergreen. This is also covered in the discussion, where we touch on the evergreen nature of the savanna canopy and its ability to maintain transpiration throughout the dry season. There is a small presence of semi- and brevi-deciduous trees along the NATT, however the savanna region of Northern Australia is largely dominated by evergreen Eucalypts, such that one can effectively model the savanna canopy as completely evergreen for these savanna sites (see Bowman and Prior, 2005). We do concede that how the models were setup to express this was lacking in the manuscript, and we have now added further text to Section 2.3 of the Methods that further describes how the savanna over- and understory was expressed for all models.

Comment 4: I appreciate seeing the partition between vegetation and soil ET, but I would really like to see a further partition of vegetation ET (and carbon flux) into tree and grass components. Are these components consistent with the fractional composition and phenology cycles? Why or why not? Are trees 'activated' quickly (like the grasses) following seasonal rains, or is there a lag time as water infiltrates into the soil? In section 4.2 the authors say "(d)uring the wet season as much as 75% of total LE arises from understory herbaceous transpiration and soil evaporation." What do the models say?

Response: Unfortunately additional outputs of which the reviewer is referring to were

not collected in this study, and would require a large effort in coordinating all partner investigators to re-run their model simulations. We concede that there are missed opportunities here in gaining a greater insight into the predicted seasonality of trees and grasses. However our aim was to examine model performance in regards to the prediction of bulk fluxes and be a foundation study on which future model experiments could be conducted for such analysis (which we are currently undertaking).

Because of the vegetation components of ET (tree and grass transpiration) are predicted at the leaf-scale and then scaled by leaf area, it should follow that the simulated water and carbon fluxes are following the phenology cycle of the site. This is illustrated in Figure 6 that shows the relative contribution of soil evaporation and total vegetation transpiration to land-surface ET. We do raise the point in the discussion (Section 4.2; paragraph 3) that the C3:C4 fraction may be too high in the wet season; that is grass LAI, and ergo grass transpiration, should be higher. However, it could also be argued that modelled grass transpiration is too conservative (e.g. modelled stomatal conductance is too low, water-use efficiency is too high, etc.) and the assumptions that govern this process may be incorrect.

Finally, the tree canopy as simulated by the models in this study was set as evergreen, reflecting the dominance of Eucalypt species in Australian savannas (Bowman and Prior, 2005). This eucalypt canopy maintains almost consistent rates of transpiration over an annual cycle, with only a minor adjustment to canopy leaf area in response to the decline in soil water content in the upper soil profile (Ma et al., 2013; Williams et al., 1996). The eucalypt canopy therefore does not respond rapidly to rainfall events (as deciduous species do in other savannas), but instead follows a long-term, conservative growth strategy. This is covered in Section 4.1 of the Discussion.

Comment 5: Two of the main take-home messages are that ‘models must treat grass as co- dominant’ and ‘models need a dynamic representation of LAI. . .’, yet current model treatment of these two processes are not explored in detail. How are the models, in their current form, failing?

Response: Generally speaking, the models used in this study do not treat grasses as an individual and independent process; that is grasses are represented as “stem-less trees”. One could therefore argue that the grass understorey is approximated rather than truly modelled; i.e. many of the same processes (and assumptions) that are used to describe

the tree canopy are also being applied to the grasses. Most of the models distinguish the separation of tree and grass life forms by relying on different parameter sets (e.g. V_{cmax} for grasses is set lower than trees) and using a slightly modified version of the same equation (e.g. light-limited rate of assimilation). Although, we have highlighted this in Section 4.2 (paragraph 3) of the Discussion, we have added the following text further to clarify this:

“Furthermore, it should be noted that the TBMs used in this study are not truly modelling grasses, but approximating them. Grasses are effectively simulated as ‘stem-less’ trees, and the distinction between the two life forms is reliant on different parameter sets (e.g. V_{cmax} , height, etc.) and slight modifications of the same process (e.g. rate of assimilation, respiration, etc.).”

Whilst we concede that this does not explore the problem in detail, lacking predictions (and observations) of grass leaf gas-exchange, it is only possible to infer the problems of modelling grasses. We believe the representation of grasses in modern day TBMs is not sufficient to be representative of savanna ecosystems and our results flag this as a priority for future development, which we have stated in our conclusions.

Regarding the dynamic representation of LAI, we don't believe the models are failing due to lacking this process. Rather by prescribing LAI as an input, the model becomes limited to a more rigid representation of the ecosystem; i.e. it becomes difficult to explore specific vegetation feedbacks (e.g. tree:grass ratio) as a result of changing climate. For example, model simulations that perturb seasonal rainfall would have less effect in a model that prescribes LAI than in one that predicts it. Consequently, DGVMs (which predict LAI) have a greater potential value in modelling savanna ecosystems. However, as has been shown in this study, the DGVM used here performed no better, if not worse, than the other models. As we point out in the text, this is likely a result of the model not truly being dynamic; i.e. allocation is performed at an annual time-step, rather than on a sub-monthly basis. Without performing further modelling experiments that investigate the sensitivity of model predictions to LAI, it is difficult to provide any further reasoning than what we have provided. However, we have furnished the text in Section 4.3 of the Discussion to make these points clearer. Again, we caution that the representation of LAI is only one of many reasons of why the models are failing for these savanna sites, and flag this as an opportunity for future model development.

6. [Specific Comments]

6a Comment: something the authors don't mention, which is critical to studies of savanna, are the limitations of the 'traditional' eddy covariance (EC) observational suite in these environments. There are a lot of moving parts in these ecosystems, and a single observation of net flux does not provide the detailed partitioning that we really need to understand with respect to relative contribution of trees, grass, and soil. I'd like to see the authors comment on the ability of above-canopy EC measurements to constrain simulations of spatiotemporally heterogeneous system such as savannas? Yes, there are a number of papers that describe site-level analyses; has a consensus or climatological pattern emerged? What does this mean for our ability to simulate these systems and evaluate our models?

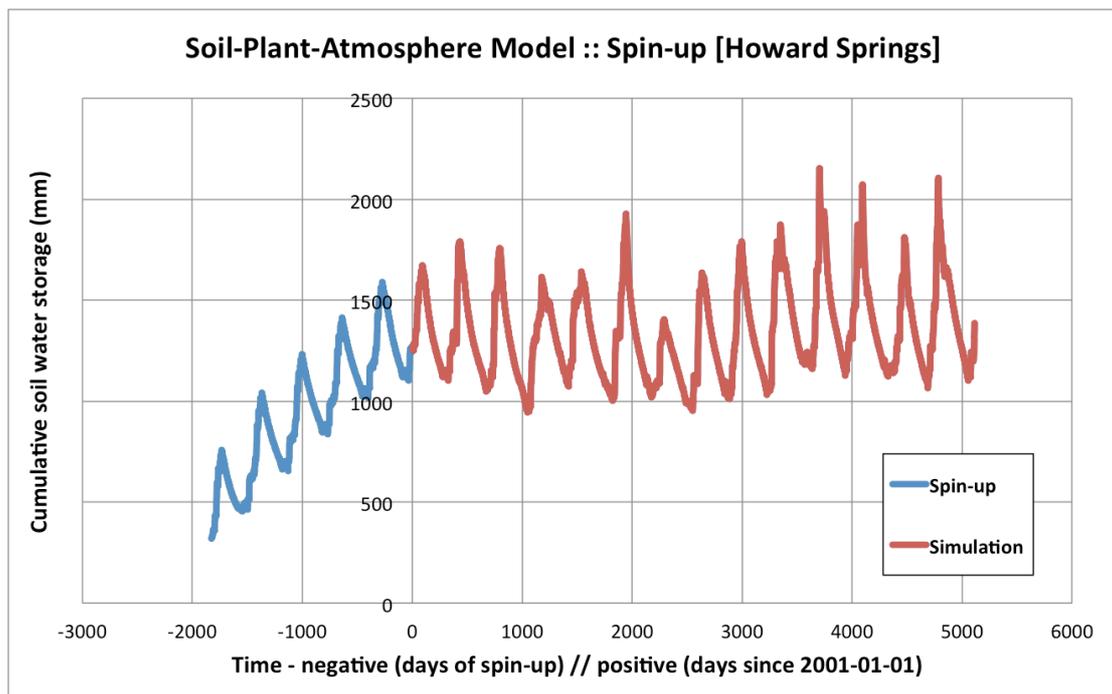
Response: The open canopy of the savanna should ideally be beneficial towards adequate mixing of gases, and numerous studies by Beringer et al. (2003, 2007, 2011), Hutley et al. (2000, 2001, 2011) have shown EC measurements for these savanna sites to be robust and have an energy balance closure better than most other global tower sites. Beringer et al (2007) have demonstrated that the carbon budget derived from EC can be mostly (all but 10%) accounted for in changes in carbon pools. The datasets provided by these flux tower sites have meteorological information that is directly measured at the same temporal and spatial scales as the flux measurements, such that they are highly valuable for model evaluation and benchmarking. However, we are aware that the co-dominance of trees and grasses in savannas, makes model evaluation and benchmarking difficult, as it is limited to bulk land-surface fluxes only. However, we point to recent work by Moore et al. (this issue), which provides the critical understorey flux dataset, will allow models to be truly evaluated in terms of how they are simulating the soil, tree and grass energy-exchange. While such datasets are still being developed, we believe the benchmarking system used here is still critical in providing a way of testing how well a model can simulate a savanna and be a foundation for future work.

6b Comment: Are the u-star/respiration models used to partition observed net carbon flux into respiration and GPP robust in savannas?

Response: We refer to Beringer et al. (2007), which shows that the respiration model estimates agree well with the inventory-based carbon balance results from Chen et al. (2003) – both studies conducted at Howards Springs.

6c Comment: Was 5 years enough time to spinup the soil? In my experience it can take 10 years or more. Was spinup tested?

Response: For the savanna sites used in this study, there was sufficient wet season rainfall for the state variables (soil moisture content and temperature) to reach equilibrium within a 5-year spin-up cycle. This of course depends on the initial conditions for soil moisture and temperature as set by the models, and these were high anyway and available in the NATT datasets. It should be noted that because the models were not attempting to simulate growth, a sensible enough water balance would be sufficient for the models to reasonably predict the fluxes. To alleviate any doubt however, we demonstrate in the figure below an example of the SPA model as run at Howard Springs. Here five years was sufficient enough for soil water storage to reach a stationary point where the initial soil water content was set at $0.05 \text{ m}^3 \text{ m}^{-3}$ (much lower than what was set in the study simulations).



6d Comment: As of January 2016, Moore et al (Contribution of trees and grasses to an Australian tropical savanna) does not appear to have been published. Can it still be cited?

Response: The Moore et al. paper has been published in discussion as part of this journal's OzFlux special issue and the citation refers to the DOI. We are happy to reference this differently and follow the discretion of the editor.

6e Comment: Is one study (Hutley 2000) enough to define tree/grass ET contributions across all savannas? Or is that description valid only at Howard Springs?

Response: This description is only valid for Howard Spring. Our aim here was not to use this example to represent ET partitioning for all savannas, but provide an example of ground-based observations that verify our inferred findings of models incorrectly determining the components of bulk land-surface LE.

6f Comment: Page 19017, line 12; Insert 'of' between 'adjustments stomatal'

Response: Done.

6g Comment: Page 19018, line 8: '(lacking observed grass LE)' is redundant

Response: Removed.

6h Comment: Page 19021, line 16: 'are therefore do not mechanistic' needs a rewrite

Response: Done.

References:

- Beringer, J., Hutley, L. B., Tapper, N. J., Coutts, a., Kerley, a. and O'Grady, a. P.: Fire impacts on surface heat, moisture and carbon fluxes from a tropical savanna in northern Australia, *Int. J. Wildl. Fire*, 12, 333, doi:10.1071/WF03023, 2003.
- Beringer, J., Hutley, L. B., Tapper, N. J. and Cernusak, L. a.: Savanna fires and their impact on net ecosystem productivity in North Australia, *Glob. Chang. Biol.*, 13(5), 990–1004, doi:10.1111/j.1365-2486.2007.01334.x, 2007.
- Beringer, J., Hutley, L. B., Hacker, J. M., Neiningner, B. and Paw U, K. T.: Patterns and processes of carbon, water and energy cycles across northern Australian landscapes: From point to region, *Agric. For. Meteorol.*, 151(11), 1409–1416, doi:10.1016/j.agrformet.2011.05.003, 2011.
- Bowman, D. M. J. S. and Prior, L. D.: Why do evergreen trees dominate the Australian seasonal tropics?, *Aust. J. Bot.*, 53(5), 379–399, doi:10.1071/BT05022, 2005.
- Chen, X., Hutley, L. B. and Eamus, D.: Carbon balance of a tropical savanna of northern Australia., *Oecologia*, 137(3), 405–16, doi:10.1007/s00442-003-1358-5, 2003.
- Donohue, R. J., Mcvicar, T. R. and Roderick, M. L.: Climate-related trends in Australian vegetation cover as inferred from satellite observations, 1981-2006, *Glob. Chang. Biol.*, 15, 1025–1039, doi:10.1111/j.1365-2486.2008.01746.x, 2009.
- Hutley, L. B., O'Grady, A. P. O. and Eamus, D.: Evapotranspiration from Eucalypt open-forest savanna of Northern Australia, *Funct. Ecol.*, 14, 183–194, 2000.
- Hutley, L. B., Grady, A. P. O. and Eamus, D.: Monsoonal influences on evapotranspiration of savanna vegetation of northern Australia, *Oecologia*, 126(3), 434–443, doi:10.1007/s004420000539, 2001.
- Hutley, L. B., Beringer, J., Isaac, P. R., Hacker, J. M. and Cernusak, L. A.: A sub-continental scale living laboratory: Spatial patterns of savanna vegetation over a rainfall gradient in northern Australia, *Agric. For. Meteorol.*, 151(11), 1417–1428, doi:10.1016/j.agrformet.2011.03.002, 2011.
- Ma, X., Huete, A., Yu, Q., Coupe, N. R., Davies, K., Broich, M., Ratana, P., Beringer, J., Hutley, L. B., Cleverly, J., Boulain, N. and Eamus, D.: Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect, *Remote*

Sens. Environ., 139, 97–115, doi:10.1016/j.rse.2013.07.030, 2013.

Moore, C. E., Beringer, J., Evans, B., Hutley, L. B., McHugh, I. and Tapper, N. J.: The contribution of trees and grasses to productivity of an Australian tropical savanna, *Biogeosciences*, this issue.

Murphy, B. P. and Bowman, D. M. J. S.: Seasonal water availability predicts the relative abundance of C-3 and C-4 grasses in Australia, *Glob. Ecol. Biogeogr.*, 16(2), 160–169, doi:DOI 10.1111/j.1466-8238.2006.00285.x, 2007.

Sea, W. B., Choler, P., Beringer, J., Weinmann, R. a., Hutley, L. B. and Leuning, R.: Documenting improvement in leaf area index estimates from MODIS using hemispherical photos for Australian savannas, *Agric. For. Meteorol.*, 151(11), 1453–1461, doi:10.1016/j.agrformet.2010.12.006, 2011.

Smith, B., Prentice, I. C. and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space, *Glob. Ecol. Biogeogr.*, 10(6), 621–637, doi:10.1046/j.1466-822X.2001.t01-1-00256.x, 2001.

Still, C. J., Berry, J. a., Collatz, G. J. and DeFries, R. S.: Global distribution of C 3 and C 4 vegetation: Carbon cycle implications, *Global Biogeochem. Cycles*, 17(1), 6–1–6–14, doi:10.1029/2001GB001807, 2003.

Williams, R. J., Duff, G. A., Bowman, D. M. J. S. and Cook, G. D.: Variation in the composition and structure of tropical savannas as a function of rainfall and soil texture along a large-scale climatic gradient in the Northern Territory, Australia, *J. Biogeogr.*, 23(6), 747–756, doi:10.1111/j.1365-2699.1996.tb00036.x, 1996.