

Author responses to reviewer and editor comments for manuscript submission BG-2015-585

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 and made the appropriate changes where necessary in our manuscript – attached to the end of this document as a marked-up version. 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, line 212 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 response to Comment 1a]. 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 and mention in the Conclusions).

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, line 601) of the Discussion, we have added the following the 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?

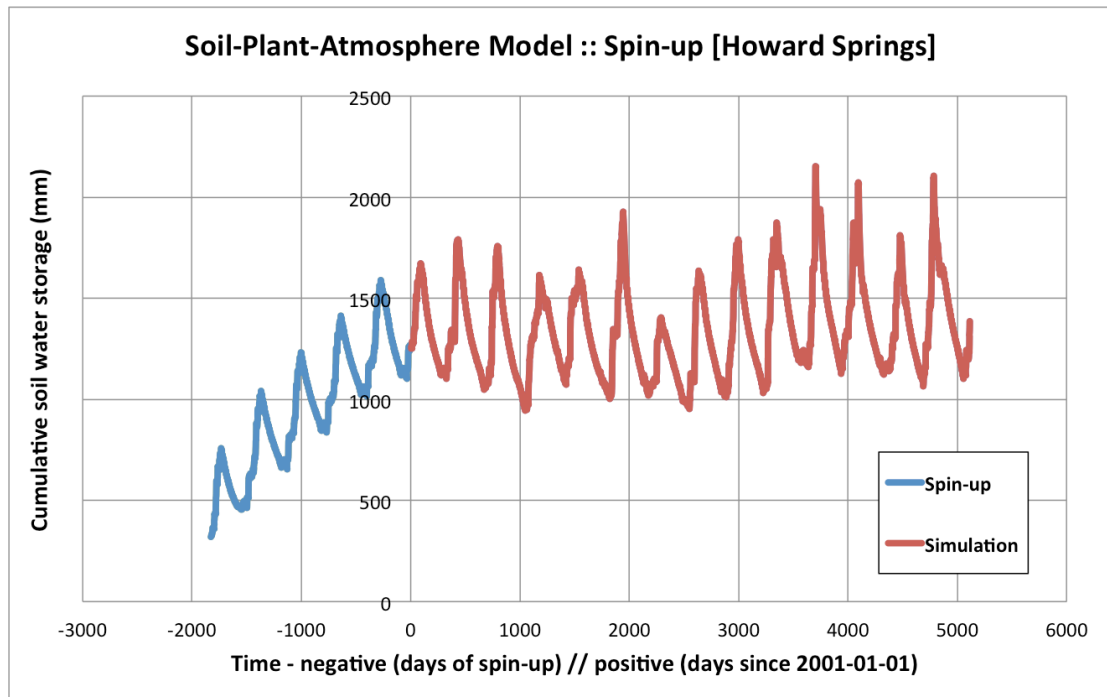
Response: The open canopy of the savanna should ideally be beneficial towards adequate mixing of gases, and numerous studies by Beringer et al. (2003, 2007, 2011), Hutley et al. (2000, 2001, 2011) have shown EC measurements for these savanna sites to be robust and have an energy balance closure better 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 understory 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 as a discussion paper 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., Neining, 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*, doi:10.5194/bgd-12-19307-2015, 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.

1 **A model inter-comparison study to examine limiting factors in**
2 **modelling Australian tropical savannas**

3

4 **Authors:** Rhys Whitley¹, Jason Beringer², Lindsay Hutley³, Gabriel Abramowitz⁴, Martin
5 G. De Kauwe¹, Remko Duursma⁵, Bradley Evans⁶, Vanessa Haverd⁷, Longhui Li⁸,
6 Youngryel Ryu⁹, Benjamin Smith¹⁰, Ying-Ping Wang¹¹, Mathew Williams¹², Qiang Yu⁷

7

8 **Institutions:**

9 ¹Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109,
10 Australia

11 ²School of Earth and Environment, University of Western Australia, Crawley, WA 6009,
12 Australia

13 ³School of Environment, Charles Darwin University, Casuarina, NT 0810, Australia

14 ⁴Climate Change Research Centre, University of New South Wales, Kensington, NSW
15 2033, Australia

16 ⁵Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, New
17 South Wales 2751, Australia

18 ⁶Faculty of Agriculture and Environment, University of Sydney, Eveleigh, NSW 2015,
19 Australia

20 ⁷CSIRO Ocean and Atmosphere, Canberra 2601, Australia

21 ⁸School of Life Sciences, University of Technology Sydney, Ultimo, NSW 2007, Australia

22 ⁹Department of Landscape Architecture and Rural Systems Engineering, Seoul National
23 University, Seoul, South Korea

24 ¹⁰Department of Physical Geography and Ecosystem Science, Lund University, Lund,
25 Sweden

26 ¹¹CSIRO Ocean and Atmosphere, Aspendale, Victoria 3195, Australia

27 ¹²School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

28

29 **Corresponding author:**

30 Name: Rhys Whitley

31 Email: rhys.whitley@mq.edu.au

32 Address:

33 Department of Biological Sciences

34 Macquarie University

35 North Ryde, NSW

36 2109, Australia

37

38

39 **Abstract:**

40 Savanna ecosystems are one of the most dominant and complex terrestrial biomes that
41 derives from a distinct vegetative surface comprised of co-dominant tree and grass
42 populations. While these two vegetation types co-exist functionally, demographically
43 they are not static, but are dynamically changing in response to environmental forces
44 such as annual fire events and rainfall variability. Modelling savanna environments with
45 the current generation of terrestrial biosphere models (TBMs) has presented many
46 problems, particularly describing fire frequency and intensity, phenology, leaf
47 biochemistry of C₃ and C₄ photosynthesis vegetation, and root water uptake. In order to
48 better understand why TBMs perform so poorly in savannas, we conducted a model
49 inter-comparison of 6 TBMs and assessed their performance at simulating latent energy
50 (LE) and gross primary productivity (GPP) for five savanna sites along a rainfall gradient
51 in northern Australia. Performance in predicting LE and GPP was measured using an
52 empirical benchmarking system, which ranks models by their ability to utilise
53 meteorological driving information to predict the fluxes. On average, the TBMs
54 performed as well as a multi-linear regression of the fluxes against solar radiation,
55 temperature and vapour pressure deficit, but were outperformed by a more complicated
56 nonlinear response model that also included the leaf area index (LAI). This identified
57 that the TBMs are not fully utilising their input information effectively in determining
58 savanna LE and GPP, and highlights that savanna dynamics cannot be calibrated into
59 models and that there are problems in underlying model processes. We identified key
60 weaknesses in a model's ability to simulate savanna fluxes and their seasonal variation,
61 related to the representation of vegetation by the models and root water uptake. We
62 underline these weaknesses in terms of three critical areas for development. First,
63 prescribed tree-rooting depths must be deep enough, enabling the extraction of deep
64 soil water stores to maintain photosynthesis and transpiration during the dry season.
65 Second, models must treat grasses as a co-dominant interface for water and carbon
66 exchange, rather than a secondary one to trees. Third, models need a dynamic
67 representation of LAI that encompasses the dynamic phenology of savanna vegetation
68 and its response to rainfall interannual variability. We believe this study is the first to
69 assess how well TBMs simulate savanna ecosystems, and that these results will be used
70 to improve the representation of savannas ecosystems in future global climate model
71 studies.

72

73 **Introduction**

74 Savanna ecosystems are a diverse and important biome that play a significant role in
75 global land-surface processes (van der Werf et al., 2008). Globally, they occupy regions
76 around the wet-dry tropical to sub-tropical equatorial zone, covering approximately 15
77 to 20% of the terrestrial surface and contribute ~30% to global net primary production
78 (Grace et al., 2006; Lehmann et al., 2014). Savannas are water-limited ecosystems where
79 rainfall is often seasonal or monsoonal, and have a spatial extent that can cover an area
80 with annual rainfall in the range of 500 to 2000 mm (Bond, 2008; Kanniah et al., 2010;
81 Sankaran et al., 2005). The variability in the amount and timing of annual rainfall,
82 coupled with local topo-edaphic properties, and the frequency and intensity of seasonal
83 fires strongly influences the structure and function of savanna vegetation (Beringer et
84 al., 2007; Kanniah et al., 2010; Ma et al., 2013; Sankaran et al., 2005). Savannas are
85 characterised by a multi-layer stratum of vegetation, where an open and discontinuous
86 canopy overstorey is seasonally dominated by understorey grasses (Scholes and Archer,
87 1997). These tree and grass layers are distinctly and functionally different, fixing carbon
88 using different photosynthetic pathways, C₃ and C₄ photosynthesis respectively (Bond,
89 2008; Scholes and Archer, 1997; Williams et al., 1996b). The canopy overstorey can be
90 either evergreen or deciduous (depending on the evolutionary history), while the grass
91 understorey is annual: active only in the wet season and senescing at the end of this
92 period (Williams et al., 1996b). Consequently, water, carbon and nutrient cycling in
93 savannas is largely determined from the balance and co-existence of these two life forms
94 (Lehmann et al., 2009; Sankaran et al., 2005).

95 Given the complex nature of savannas, modelling the land surface exchange and
96 vegetation dynamics for this biome is challenging for terrestrial biosphere models
97 (TBMs). Here we define TBMs to broadly encompass stand, land-surface, and dynamic
98 global vegetation models (Pitman, 2003). Most land surface schemes that feed into
99 larger earth system models use simplistic representations of vegetation, and these will
100 have difficulty describing the complex structure of savannas ecosystems. Such issues
101 may be: simplistic assumptions in relation to rooting depth and inadequate responses to
102 drought (De Kauwe et al., 2015; Li et al., 2012); ignoring the multilayered nature of
103 savannas and the differing structural (including radiation), functional (including
104 different plant functional types) and phenological differences (Whitley et al., 2011); and
105 in some cases neglecting the C₄ photosynthetic pathway entirely (Parton et al., 1983;
106 Schymanski et al., 2007) It is therefore critical that TBMs meet the challenges that

Rhys Whitley 2/2/16 11:22 AM

Deleted: ,

108 savanna dynamics present if water and carbon exchange are to be correctly simulated in
109 response to global change.

110 Despite these issues, there have been significant advances in modelling savanna
111 dynamics in recent years, and this has been focused on integrating important features
112 specific to savanna ecosystems, namely frequent fire and tree-grass competitive
113 interactions, processes that shape savanna structure and function (Haverd et al., 2015;
114 Higgins and Scheiter, 2012; Scheiter and Higgins, 2007; Scheiter et al., 2014; Simioni et
115 al., 2003). Nevertheless, little work has been undertaken to critically evaluate the
116 performance and processes of TBMs when used to capture water and carbon cycling in
117 savannas, most notably in west Africa (Simioni et al., 2000) and Australia (Schymanski
118 et al., 2007, 2008, 2009; Whitley et al., 2011). Many global ecosystem models moreover
119 use broad plant functional types (PFTs) with single parameter values to describe whole
120 biomes (Pitman, 2003), making them unable to represent changing vegetation structure
121 (tree:grass ratio) in the continuum of grassland to woodland savanna. Approaches have
122 been developed that can account for savanna dynamics, such as using mixed tiles,
123 whereby trees and grasses are simulated as separate surfaces that are then aggregated
124 together (Kowalczyk et al., 2006). However, this approach fails to capture the
125 competition between trees and grasses for light, water and nutrient resources.

126 In this study, we take 6 TBMs of distinctly different conceptual frameworks, and assess
127 their ability at simulating savanna water and carbon exchange along the North
128 Australian Tropical Transect (NATT) that is defined by a strong rainfall gradient.
129 Australian tropical savannas can be considered largely intact compared to South
130 American and African savannas, and provide a 'living laboratory' to understand the links
131 between vegetation structure and function and how it responds to environmental
132 change (Hutley et al., 2011). We challenge the models by evaluating them along the
133 rainfall gradient, which extends over a broad biogeographical extent and strong
134 interannual variability in climate (Koch et al., 1995). The aim of this study is to highlight
135 critical processes that may be missing in current TBMs and are required to adequately
136 simulate savanna ecosystems. Specifically, we examine whether a TBM's structural
137 framework, such as the representation of the understorey grasses (C₄ photosynthesis),
138 tree rooting depth, and description of phenology (prescribed vs. dynamic) can
139 adequately replicate observed carbon and water fluxes. To achieve this we measure the
140 performance of each TBM by comparing its predictions to a set of empirical benchmarks
141 that describe *a priori* expected levels of model performance. We identify regions of low
142 performance among sites and seasons, to diagnose under what climate conditions

Rhys Whitley 2/2/16 11:26 AM

Deleted: with

Rhys Whitley 2/2/16 11:26 AM

Deleted: :

145 reduced model performance occurs. We then infer what processes (present or missing)
146 may be the cause for reduced performance when applied to savannas ecosystems. Our
147 intention is that these results can be used to flag high priorities for future development
148 by the terrestrial biosphere modelling community.

149

150 2. Methodology

151 2.1 Observational data

152 The North Australian Tropical Transect (NATT) is a sub-continental rainfall gradient in
153 the wet-dry tropical climate zone of Northern Australia, which encompasses a distance
154 of approximately 1000 km over a latitudinal range of -12 to -23 °S and a decline in mean
155 annual precipitation (MAP) from 1700 mm to 300 mm (Hutley et al., 2011). It is one of
156 three savanna transects established in the mid 1990's, forming part of the International
157 Geosphere Biosphere Program (IGBP) along with the *SAvannas* in the Long Term (SALT)
158 transect in West Africa and the Kalahari Transect (KT) in Southern Africa (Koch et al.,
159 1995). Soils range from sand dominated red Kandosols to black, cracking clay soils that
160 are more extensive in the southern end of the NATT that are limiting to woody plant
161 growth (Hutley et al., 2011; Williams et al., 1996b). Kandosols are ancient and
162 weathered, such that they have been leached of nutrients by the large monsoonal
163 rainfall (McKenzie et al., 2004). Close to the northern coastline, vegetation is comprised
164 primarily of evergreen *Eucalyptus* and *Corymbia* tree species that overly an understorey
165 of *C₄ Sorghum* and *Heteropogon* spp. grasses. Inland, tree biomass, leaf area index (LAI)
166 and cover tends to decline and by -18 °S savanna vegetation transitions to less dense
167 *Acacia* woodlands, shrublands and grasslands *that are* dominated by *Astrebla* grass
168 species (Hutley et al., 2011). Fires occur regularly in these environments, increasing in
169 frequency with higher rainfall (MAP > 1000 mm), *and are* fuelled by the accumulation of
170 understorey *C₄* grasses that cure in the dry season (Beringer et al., 2014; Russell-Smith
171 and Edwards, 2006).

172 The five flux tower sites along the NATT used in this study are outlined in Table 1, *and*
173 describes stand soil and vegetation characteristics, as well as a summary of local
174 meteorology (Hutley et al., 2011). These sites represent a sampling of savanna
175 environments covering a wide range of MAP and a much smaller range of mean annual
176 temperature (MAT) (Fig. 1). At each site, an eddy covariance system was used to
177 measure the ecosystem-atmosphere exchange of radiation, heat, water and CO₂. Quality

Rhys Whitley 2/2/16 11:31 AM

Deleted: which

179 assurance and control (QA/QC) and corrections on the fluxes were carried out on the 30
180 minute dataset using the OzFlux QC/QA protocol (v2.8.5), developed by the OzFlux
181 community under creative commons licensing (www.ozflux.org.au) (see Eamus et al.,
182 2013). Missing or rejected data were gap-filled using the DINGO (Dynamic INtegrated
183 Gap filling and partitioning for Ozflux) system (see Moore *et al.* this issue). Gross
184 primary productivity (GPP) *was* not observed but determined from the difference
185 between measured net ecosystem exchange (NEE) and modelled ecosystem respiration
186 (Re). Values of Re were determined by assuming nocturnal NEE equals Re under the
187 conditions for sufficient turbulent transport. Values that meet these requirements are
188 then used to make daytime predictions of Re, using an artificial neural network (ANN),
189 with soil moisture and temperature, air temperature, and the normalised difference
190 vegetation index (NDVI) used as predictors. Additionally, the effect of fire on the water
191 and carbon fluxes are quantified and incorporated into the datasets accounting for the
192 nonlinear response in productivity (becoming a carbon source) during the post-fire
193 recovery period (Beringer et al., 2007). Because the TBMs used here do not attempt to
194 simulate stochastic fire events (and other disturbance regimes), these post-fire recovery
195 periods were removed when determining the benchmarks and model performance as
196 described below.

197 Finally, we use the definitions for water and carbon exchange as outlined by Chapin et
198 al. (2006), whereby the sub-daily rate of GPP is expressed in $\mu\text{mol m}^{-2} \text{s}^{-1}$ and uses a
199 negative sign (-) to denote the removal of CO_2 from the atmosphere. Similarly, LE is
200 expressed in terms of energy as W m^{-2} and uses a positive sign to denote the addition of
201 H_2O to the atmosphere.

202

203 2.2 Terrestrial biosphere models

204 The 6 TBMs used in this study cover a wide spectrum of characteristics of operation,
205 scale and function, and include differences in operational time-step (30min vs. daily),
206 scope of simulated processes (soil hydrology, static or dynamic vegetation, multi-layer
207 or big leaf description of the canopy) and intended operational use (coupled to [earth](#)
208 [system models](#), offline prediction, driven by remote sensing products). These
209 characteristics along with what we define as a model 'functional class' are given in Table
210 2 and are defined as follows. Stand models (SMs) give detailed multi-layer descriptions
211 of canopy and soil processes for a particular point, operating at a sub-daily time-step

Rhys Whitley 2/2/16 11:35 AM

Deleted: at point

213 (Soil-Plant-Atmosphere model: SPA, and MAESPA). Land-Surface models (LSMs) operate
214 at the same temporal resolution as SMs, but adopt a simpler representation of canopy
215 processes, allowing them to be applied spatially (Community Atmosphere Biosphere
216 Land Exchange model; CABLE, and BIOS2; a modified version of CABLE). Dynamic Global
217 Vegetation Models (DGVMs) simulate water, carbon much like the other models, but
218 simulate dynamic rather than static vegetation that changes in response to climate and
219 disturbance (Lund-Potsdam-Jena General Ecosystem Simulator; LPJGUESS). Lastly,
220 Remote Sensing models (RSMs) are driven by remotely sensed atmospheric products,
221 and infer water-stress of vegetation through changes in fractional cover rather than
222 detailed soil hydrological processes (Breathing Earth System Simulator; BESS). Some of
223 the TBMs share similar structural frameworks in parts: for example, both SPA and
224 MAESPA use similar below-ground soil hydrology and root-water uptake schemes, while
225 BIOS2 is fine-spatial-resolution (0.05 degree), offline modelling environment for
226 Australia, in which predictions of CABLE (with alternate parameterisations of drought
227 response and soil hydrology) are constrained by multiple observation types (see Haverd
228 *et al.* 2013). Although [these similarities](#) reduces the number of truly, functionally,
229 independent models used in the experiment, the presence of such overlap can be useful
230 in identifying if particular frameworks are the cause for model success or failure.

231

232 2.3 Experimental protocol

233 All TBMs were parameterised for each of the five savanna sites using standardised
234 information on vegetation and soil profile characteristics (Table 1). For TBMs that
235 required them, parameter values pertaining to leaf biochemistry, such as maximum
236 Rubisco activity (V_{cmax}) and leaf nitrogen content per leaf area (N_{area}), were assigned
237 from Cernusak *et al.* (2011), who undertook a physiological measurement campaign
238 during the SPECIAL program (Beringer *et al.* 2011). Parameters relating to soil sand and
239 clay content were taken from the Australian Soil Classification (Isbell, 2002), while root
240 profile information was sourced from Chen *et al.* (2003) and Eamus *et al.* (2002). [Each](#)
241 [TBM was setup to describe a C₃ evergreen overstorey with an underlying C₄ grass](#)
242 [understorey, and conforms well with the characteristics of savannas in Northern](#)
243 [Australia](#) (Bowman and Prior, 2005). [All TBMs \(excluding LPJGUESS\) prescribed LAI as](#)
244 [an input, to characterise the phenology of vegetation at each site. In these cases LAI was](#)
245 [determined from](#) MODIS derived approximations that [were](#) well matched to ground-
246 based estimations of LAI at the SPECIAL sites (Sea *et al.*, 2011). [The fraction of C₃ to C₄](#)

Rhys Whitley 1/2/16 3:56 PM

Deleted: -

Rhys Whitley 2/2/16 11:36 AM

Deleted: this

249 [vegetation was handled differently by each model and was determined for each as](#)
250 [follows. For](#) MAESPA and SPA, the models allowed for time-varying tree and grass
251 fractions to be assigned [as direct inputs](#), and these time-varying fractions were
252 determined using the method of Donohue et al. (2009). [BIOS2 similarly used the same](#)
253 [method to extract time-varying fractions, while CABLE used a static fraction that did not](#)
254 [change. The BESS model derived the C₃:C₄ fraction from the C₃ and C₄ distribution map of](#)
255 [Still et al. \(2003\), while for LPJGUESS this fraction is a prognostic determination](#)
256 [resulting from the competition between trees and grasses \(see Smith et al., 2001\).](#) Model
257 simulations were driven using observations of solar radiation, air temperature, relative
258 humidity (or vapour pressure deficit; VPD), rainfall, atmospheric CO₂ concentration and
259 LAI (if prescribed), and included a spin-up period of 5 years to allow internal states,
260 such as the soil water balance and soil temperature to reach equilibrium. The exception
261 to the above was the BIOS2 model, which was run using gridded meteorological inputs
262 and had its model parameters optimised through a model-data fusion process (see
263 Haverd et al., 2013).

264 Simulations for each savanna site covered a period of 2 to 10 years depending on the
265 availability of data from each flux site (Table 1) and results were standardised to the
266 ALMA (Assistance for Land-surface Modelling Activities) convention. Model predictions
267 of LE and GPP were then evaluated against local observations at each site from the eddy
268 covariance datasets and benchmarked following the methodology proposed by the PALS
269 Land Surface Model Benchmarking Evaluation Project (PLUMBER) (Abramowitz, 2012;
270 Best et al., 2015) as described below.

271

272 **2.4 Empirical benchmarking**

273 The paradigm for model assessment set out in the Protocol for the Analysis of Land-
274 surface models (PALS) (Abramowitz, 2012) suggests that model assessment is more
275 meaningful when *a priori* expectations of performance in any given metric can be
276 defined. Such benchmarks can be created using simple empirical models, built on
277 statistical relationships between the fluxes and drivers, and establish the degree to
278 which models utilise the information available in their driving data about the fluxes they
279 aim to predict. Additionally, these empirical models are simple in the sense that they are
280 purely instantaneous response to time-varying meteorological forcing and contain no
281 internal states or expression of ecophysiological processes. This is in comparison to

282 TBMs that are complex, having some 20+ soil and vegetation parameters, internal states,
283 partitioning of light, as well as soil and vegetation, carbon and nitrogen pools
284 (Abramowitz et al., 2008).

285 We created a set of 3 empirical models of increasing complexity following the procedure
286 of Abramowitz (2012), which we compared with the TBMs. The first benchmark (emp1)
287 is simply a linear relationship between a turbulent flux (LE or GPP) and downward
288 short-wave radiation (R_s). The second benchmark (emp2) is slightly more complex, and
289 is a multi-linear regression between a flux and R_s , air temperature (T_a), and vapour
290 pressure deficit (VPD). Finally, the third benchmark (emp3) is the most complex and is a
291 nonlinear regression of the fluxes against R_s , T_a , VPD and LAI, determined from an ANN.
292 This benchmark is constructed using a self-organising linear output map that clusters
293 the four covariates into 10^2 distinct nodes and performs a multi-linear regression
294 between the fluxes and the 4 covariates at each node, resulting in a nonlinear (piece-
295 wise linear) response to the meteorological forcing data (Abramowitz et al., 2008; Hsu,
296 2002). In a departure from Abramowitz (2012), we include LAI as an additional
297 covariate, as the seasonal variance of savanna water and carbon exchange is strongly
298 coupled to the phenology of the grasses and to the deciduous and semi-deciduous
299 woody species (Moore et al., *this issue*). [The seasonal behaviour of the empirical](#)
300 [benchmark drivers along the NATT can be referred to in the supplementary information.](#)
301 Empirical benchmarks are created for each of the five flux sites using non gap-filled data,
302 and are parameterised *out-of-sample*, such that they use data from all sites except the
303 one in question. For example, the Howard Springs empirical benchmark models would
304 use information from Adelaide River, Daly Uncleared, Dry River and Sturt Plains to
305 establish their parameter values, but would exclude Howard Springs itself. Constructing
306 the benchmarks out-of-sample results in what is effectively a generalised response to an
307 independent dataset. Once the empirical models were calibrated for each site,
308 benchmarks were then created for [both fluxes](#) using the same meteorological forcing
309 used to run the TBMs.

310 Finally, we assess ecosystem model performance in terms of a ranking system, following
311 the PLUMBER methodology of Best et al. (2015). The performance of each individual
312 ecosystem model in predicting both LE and GPP at each site was determined using four
313 statistical metrics that describe the mean and variability of a model compared to the
314 observations. These metrics included the correlation coefficient (r), standard deviation
315 (sd), normalised mean error (NME), and mean bias error (MBE) (see Table B1).
316 Similarly, the same metrics were determined for each of [the](#) 3 benchmarks at each

317 savanna site. Each TBM was then ranked against the benchmarks (independently of the
318 other models) for each of the metrics listed above., where the ranking is between 1 and
319 4 (1 model + 3 benchmarks), and the best performing model for a given metric is ranked
320 as 1. An average ranking is then determined across all metrics for each TBM, and all
321 benchmarks to give a final ranking of performance for each savanna site. The ranks
322 denote the number of metrics being met by the models and are not a measure of the
323 smallest absolute error. In determining the average ranks, the metrics were evaluated at
324 the daily time scale, as this was the lowest temporal resolution common amongst the
325 TBMs. Additionally, days where either driver or flux had been gap-filled were removed
326 in the determination of these metrics. Herewith we use the term performance to relate
327 to how well the TBMs compare to the benchmarks as expressed by the ranks.

328

329 3. Results

330 3.1 Model predictions

331 Figure 2 shows the daily time-course of LE and GPP from the flux tower, models, and
332 benchmarks at each of the five savanna sites. Models, benchmarks and observations are
333 represented as a smoothed time-series (7-day running mean) and have been aggregated
334 into an ensemble year to express the typical seasonality of savanna water and carbon
335 exchange. Visually, the TBMs showed varying levels of performance across the rainfall
336 gradient. None of the models showed a clear consistency in simulating either flux and
337 responded differently to the meteorological drivers across sites. Additionally, some of
338 the models, such as CABLE and LPJ-GUESS, showed difficulty in simulating the
339 seasonality of the fluxes across the transect, particularly GPP. Differences among model
340 simulated LE and GPP were larger in the wet season than the dry season. However,
341 modelled LE and GPP appeared to co-vary quite strongly; overall both fluxes were
342 underestimated across sites by most models. Simulations by SPA and MAESPA were the
343 exception to this, broadly capturing tower GPP, despite consistently underestimating LE
344 across sites.

345 Figure 3 shows the probability density functions (PDFs) for the wet (Nov – Apr) and dry
346 season (May – Oct) fluxes at each savanna site. Tower and model PDFs were determined
347 by binning each flux into the respective seasons and using kernel density estimation
348 (Bashtannyk and Hyndman, 2001) to determine smoothed distributions. The shape and
349 spread of the distributions highlight possible biases in the models (over- or

Rhys Whitley 2/2/16 11:49 AM

Deleted: T

Rhys Whitley 2/2/16 11:49 AM

Deleted: ,

Rhys Whitley 2/2/16 11:49 AM

Deleted: where

Rhys Whitley 2/2/16 11:49 AM

Deleted: ecosystem mo

Rhys Whitley 2/2/16 11:49 AM

Deleted: de

Rhys Whitley 2/2/16 11:49 AM

Deleted: l

Rhys Whitley 2/2/16 11:50 AM

Deleted: at the

Rhys Whitley 2/2/16 11:50 AM

Deleted: in question

Rhys Whitley 2/2/16 11:51 AM

Deleted: To

Rhys Whitley 2/2/16 11:51 AM

Deleted: e

Rhys Whitley 2/2/16 11:51 AM

Deleted: models

Rhys Whitley 1/2/16 3:56 PM

Deleted: -

Rhys Whitley 2/2/16 11:54 AM

Deleted: ,

363 underestimating the tower fluxes), as well as their ability to capture the spread of
364 values. The PDFs for the tower fluxes tended to shift to low values and became narrower
365 as annual rainfall declined, and this was most prominent in the dry season. By contrast,
366 the PDFs of the model simulations did not replicate this trend, being mostly stationary
367 (i.e. density of values do not shift) across sites, especially for wet season GPP. For
368 example, the LPJGUESS and MAESPA models showed distributions with higher densities
369 at low LE ($20 - 40 \text{ W m}^{-2}$) across all sites and seasons, whereas during the wet season
370 almost all models except MAESPA and SPA had higher densities of GPP at low values ($3 -$
371 $4 \mu\text{mol m}^{-2} \text{ s}^{-1}$). In models that did replicate the observed spread and shape of the tower
372 fluxes (CABLE, LPJGUESS), PDFs were biased towards lower values particularly for LE,
373 where there was little overlap. A consistent pattern emerged among the models,
374 suggesting that drier conditions resulted in PDFs with higher densities at low values for
375 both fluxes.

376 The benchmarks showed that there was enough information in R_s (emp1) to predict the
377 magnitude and daily time-course of the tower fluxes (data not shown). However,
378 additional information contained in T_a and VPD (emp2) was needed to capture the
379 water and carbon flux seasonality. Importantly additional phenological information
380 (provided as LAI in emp3) was required in order to fully capture the seasonality of the
381 savannas fluxes and provide predictions with the lowest absolute error. This indicates
382 that in order for models to achieve the best possible performance simulating savanna
383 ecosystems they will require correct information on phenology.

384

385 **3.2 Residual analysis**

386 Figure 4 shows how model structure (internal processes) is affecting predictions of
387 savanna fluxes across the rainfall gradient. To do this we examine the standardised
388 model residuals from each TBM where the scale is expressed in terms of standard
389 deviations. Figure 4 shows the LE and GPP residuals at each savanna site represented in
390 two ways: (i) against time, and (ii) against the flux prediction itself. Plotting the
391 residuals against time provides an effective way of examining how a model responds to
392 progressive changes in the environment, while plotting the residuals against the model
393 prediction gives a measure of a model's bias (Medlyn et al., 2005). A linear regression
394 has been applied to the residuals versus flux prediction scatter plots in order to better
395 visualise the degree of bias.

Rhys Whitley 1/2/16 3:56 PM

Deleted: -

397 Model residuals for both LE and GPP illustrated a larger tendency towards
398 underestimation in the wet than in the dry season. The LE residuals showed a larger
399 between-day variance than the GPP residuals in the wet season. A possible explanation
400 for this is that LE is the sum of multiple evaporation components: vegetation
401 transpiration, soil evaporation, **and** wet canopy **evaporation**, i.e. 3 sources for potential
402 error. However, the error and variance of the LE residuals reduced into the dry season,
403 and as the site itself became drier. It is likely that this reduced error and variance in LE
404 was a combination of the increasing contribution of woody transpiration to bulk land-
405 surface LE as the grass transpiration and soil evaporation declined (via senescence and
406 low surface soil moisture respectively). Despite this reduction in error, there was still an
407 underestimation of dry season LE that was larger at wetter sites, suggesting that
408 processes affecting simulated plant hydrology (e.g. root-water uptake, plant
409 conductance, rooting depth) are inadequate for some of the TBMs (e.g. LPJGUESS,
410 CABLE). The GPP residuals broadly showed a similar response over time, with the
411 exception that the between-day variance was lower and did not change between
412 seasons. Again, moving into the dry season where the grasses become inactive, the
413 residuals go to zero. This suggests that tree transpiration and photosynthesis are in
414 general being adequately modelled, albeit with some underestimation. A possible
415 explanation for this is too shallow tree rooting depths at the wetter sites. The large error
416 in the wet season that is visible across the transect points to a structural issue with how
417 the grasses are modelled.

418 All models show different biases in their predictions (indicated by the slope), and these
419 biases in almost all cases change with site and are not consistent among the models. The
420 biasing for the CABLE and BIOS2 models is similar and positive for LE, yet divergent for
421 GPP. For the LPJGUESS, MAESPA and SPA models, there is significant negative bias in
422 predicted LE, while for the GPP predictions the bias is smaller. The BESS model shows
423 both negative and positive biasing, but is small and somewhat consistent among sites.

424

425 **3.3 Model performance**

426 Figure 5 shows a comparison of individual TBM performance ordered by site from
427 wettest (Howard Springs) to driest (Sturt Plains) and in terms of their annual, wet and
428 dry season predictions for each flux. Despite differences in model complexity (Table 1),
429 the TBMs showed a similar performance across sites and seasons. For almost all sites,

430 the TBMs outperformed the emp1 benchmark for annual flux predictions (Fig. 5a).
431 However, there were some exceptions to this, and good performance in one flux did not
432 necessarily result in good performance in the other. For example, MAESPA was unable
433 to beat the emp1 benchmark for LE at sites where MAP > 1000 mm, but performed
434 better than the emp2 benchmark for GPP. In general, there was a slight pattern of
435 increased model performance as annual rainfall declined, though with a degree of site-
436 to-site variability in the rankings for some of the TBMs.

437 In order to examine how seasonal changes affect model performance, we additionally
438 determined the metrics and rankings for the wet and dry season periods (Fig. 5b-c).
439 Seasonal differences were immediately obvious. Model performance for wet season LE
440 and GPP was low to moderate, and the majority of the TBMs showed a performance that
441 ranged between the emp1 and emp2 benchmarks. In contrast, there were noticeable
442 improvements to dry season model performance amongst the TBMs. For dry season LE,
443 half the models (BIOS2, BESS, and SPA) were able to consistently outperform the emp2
444 benchmark, and come close to meeting the same number of metrics as the emp3
445 benchmark particularly at the drier sites. In comparison, predicted dry season GPP saw
446 a larger enhancement in model performance, with TBMS more frequently outperforming
447 the emp2 benchmark and even some outperforming the emp3 benchmark (LPJGUESS,
448 BESS, and SPA at the Daly Uncleared site). The exception to all this was the CABLE
449 model, which showed surprisingly little loss or gain in performance despite the season.
450 The results give an indication that as a whole, input information was better utilised by
451 each TBM at drier sites and in the dry season, suggesting that there are problems in wet
452 season processes.

453

454 4. Discussion

455 The NATT, which covers a marked rainfall gradient, presents a natural 'living laboratory'
456 with which a models ability to simulate fluxes in savanna ecosystems may be assessed.
457 Our results have highlighted that there is a clear failure of the models to adequately
458 perform at predicting wet season dynamics, as compared to the dry season, and
459 suggests that modelled processes relating to the C₄ grass understorey are insufficient.
460 This highlights a key weakness of this group of TBMs, which likely extends to other
461 models outside of this study. The inability of these TBMs to capture wet season
462 dynamics is highlighted by the benchmarking, where the performance for many of the

Rhys Whitley 1/2/16 4:01 PM

Deleted: MAESPA

464 models was at best equivalent to that of a multi-linear regression against R_s , T_a and VPD
465 (emp2) and in some cases no better than a linear regression against R_s (emp1). Given
466 that this subset of TBMs are sophisticated process-based models that represent our best
467 understanding of land-surface, atmospheric exchange processes, we would expect them
468 to perform as well as a neural network prediction (emp3). Consequently there is an
469 evident underutilisation of the driving information (i.e. a failure to describe the
470 underlying relationships in the data) impeding the performance of these models when
471 predicting savanna fluxes. However, there were instances where some of the TBMs were
472 able to reach similar levels of performance with the emp3 benchmark, and strongly
473 suggests that each of these models is capable of replicating savanna dynamics under
474 certain conditions (e.g. during the dry season).

475
476 Our results suggest that errors among models are likely to be systematic, rather than
477 related to calibration of existing parameters. For example, BIOS2 had previously
478 optimised model parameters for Australian vegetation (see Haverd et al.2013), but was
479 still unable to out-perform the emp3 benchmark in most cases, although performed
480 better than an un-calibrated CABLE, to which it is functionally similar. Similarly,
481 MAESPA and SPA, which used considerable site characteristic information to
482 parameterise their simulations, did not significantly outperform un-calibrated models
483 (e.g. CABLE). Additionally, despite these models using the same leaf, root and soil
484 parameterisations, both SPA and MAESPA displayed markedly different performances in
485 predicting LE. Consequently, improving how models represent key processes that drive
486 savanna dynamics is critical to improving model performance across this ecosystem.

487
488 There is certainly enough information in the time-varying model inputs to be able to
489 adequately simulate wet and dry season dynamics, as is evidenced by the benchmarks.
490 We therefore consider the implications of our results, and present possible reasons
491 below for why this group of TBMs is failing to capture water and carbon exchange along
492 the NATT, and make suggestions as to how this could be improved.

493

494 **4.1 Water access and tree rooting depth**

495 During the late dry season surface soil moisture in the sandy soils declines to less than
496 3% volumetric water content, with an equivalent matric potential of 3 to 4 MPa (Prior et
497 al., 1997). During this seasonal phase, the grass understorey becomes inactive and LE
498 can be considered as equivalent to tree transpiration, such that it is the only active

499 component during this period (O'Grady et al., 1999). Using this equivalence, one can
500 infer the relative effect that rooting depth has on LE during this period. Previous studies
501 have shown that for these savanna sites along the NATT, tree transpiration is
502 maintained throughout the dry season by deep root systems that access deep soil-water
503 stores, which in turn are recharged over the wet season (Eamus et al., 2000; Hutley et
504 al., 2001; Kelley et al., 2007; O'Grady et al., 1999). In order for models to perform well
505 they will need to set adequate rooting depths and distributions, along with root water
506 uptake process, to enable a model response to such seasonal variation. Examining
507 performance across the models, we can infer this to be a key deficiency. As expected,
508 TBMs that prescribed shallow rooting depths (e.g. LPIGUESS) did not simulate this
509 process well, and underestimated dry season LE at 3 of the 5 savanna sites by up to 30
510 to 40%. The two sites at Adelaide River and Sturt Plains were an exception to this with
511 the TBMs displaying a low residual error, which is likely to be a consequence of heavier
512 textured soils and trees at these sites having shallow root profiles. At Adelaide River
513 shallow root profiles are a consequence of shallow, heavier textured soils, however dry
514 season transpiration is sustained due to the presence of saturated yellow hydrosol soils.
515 Sturt Plains is a grassland (end member of the savanna continuum) where C₄ grasses
516 dominate and no trees are present such that transpiration is close to zero in the dry
517 season. The few small shrubs that have established have shallow root profiles that have
518 adapted to isolated rainfall events driven by convective storms (Eamus et al., 2001;
519 Hutley et al., 2001, 2011). Consequently, the TBMs would be expected to perform better
520 at these sites, as water and carbon exchange will be modulated by the soil-water status
521 of the sub-surface soil layers. For the other sites, models which assumed a root depth > 5
522 m (BIOS2, SPA and MAESPA), showed the most consistent performance in predicting dry
523 season LE, and we suggest for seasonally water-limited ecosystems, such as savanna,
524 that deeper soil water access is critical. Our results highlight the need for data with
525 which to derive more mechanistic approaches to setting rooting depth, such as that of
526 Schymanski et al. (2009).

527 Interestingly, a low residual error for LE in the dry season, did not translate as good
528 performance in the overall model ranking. This suggests that other processes along the
529 soil-vegetation-atmosphere continuum need to be considered to improve simulated
530 woody transpiration. Such processes may include root-water uptake (distribution of
531 roots and how water is extracted), and the effect of water stress and increased
532 atmospheric demand at the leaf-level (adjustment of stomatal conductance due to

Rhys Whitley 1/2/16 3:56 PM

Deleted: -

Rhys Whitley 1/2/16 3:52 PM

Deleted: s

535 changes in leaf water potential). More detailed model experiments that examine how
536 each TBM simulates these processes would help identify how they can be improved.

537 An exception to the above is the BESS model, which forgoes simulating belowground
538 processes of soil hydrology and root-water uptake entirely. Rather, this model assumes
539 that the effects of soil-moisture stress on water and carbon exchange is expressed
540 through changes in LAI (and by extension V_{cmax}), which acts as a proxy for changes in soil
541 moisture content (Ryu et al., 2011). The fact that BESS performed moderately well along
542 the NATT, coupled with the fact that tree transpiration continues through the dry season
543 suggests that there may be enough active green material for remote sensing proxies of
544 water-stress to generally work rather well for savanna ecosystems. It is notable that
545 BESS overestimated both GPP and ET in dry season at the driest site, Sturt Plains (Fig
546 2e), implying that greenness detected by satellite remote sensing might not capture
547 carbon and water dynamics well in such a dry site.

548

549 4.2 Savanna wet season dynamics

550 The relative performance of the TBMs at predicting LE was much poorer in the wet
551 season compared to the dry season. The reason for this difference is that wet season LE
552 is the sum of woody and herbaceous transpiration (E_{veg}) as well as soil and wet-surface
553 evaporation (E_{soil}); in contrast dry season LE is predominantly woody transpiration as
554 described previously. During the wet season, up to 75% of total LE arises from
555 understorey herbaceous transpiration and soil evaporation (Eamus et al., 2001; Hutley
556 et al., 2000; Moore et al., *this issue*) and of this fraction the C_4 understorey contributes a
557 significant daily amount (Hutley et al., 2000). In the absence of observations of
558 understorey LE it can be difficult to determine whether grass transpiration is being
559 simulated correctly. However, separating out the components of wet season LE into soil
560 and vegetation can help identify which of these components are causes for error.

561

562 Separating the outputs of simulated E_{veg} and E_{soil} from each TBM (excluding BESS which
563 did not determine these as outputs during the study) shows that simulated wet season
564 E_{veg} was particularly low for a lot of the models, despite high LAI and non-limiting soil-
565 water conditions (Figure 6). A previous study at Howard Springs by Hutley et al. (2000)
566 observed that during the wet season, the grass understorey could transpire $\sim 2.8 \text{ mm d}^{-1}$,
567 while the tree canopy transpired only 0.9 mm d^{-1} ($E_{veg} = 3.7 \text{ mm d}^{-1}$). Of the 6 TBMs at

Rhys Whitley 1/2/16 3:53 PM

Deleted: grass

569 | Howard Springs, only CABLE and SPA were able to predict an E_{veg} close to this level,
570 | while the other models predicted values closer to tree transpiration (i.e. an under-
571 | estimate). This pattern is similar for other NATT sites, where predicted wet season E_{veg}
572 | remained low and was dominated by E_{soil} at the southern end of the NATT. An
573 | underestimation of wet season LE could be due to underestimated E_{soil} in some of the
574 | models. Conversely, CABLE and BIOS2 predicted a higher E_{soil} than the other models, and
575 | this could be a reason for their higher LE performance during the wet season. Although
576 | E_{soil} has been reported to reach as high as 2.8 mm d^{-1} at Howard Springs (Hutley et al.,
577 | 2000), predicted E_{soil} by these models may still be overestimated, given that vegetation
578 | cover during this period is at a seasonal peak (limiting energy available at the soil
579 | surface) and transpiration is only limited by available energy not water (Hutley et al.,
580 | 2000; Ma et al., 2013; Schymanski et al., 2009; Whitley et al., 2011). Given the limited
581 | data for E_{soil} along the NATT, it is difficult to determine how large E_{soil} should be.
582 | However, the ratios displayed by the TBMs appear to be reasonable though, with
583 | vegetation acting as the predominant pathway for surface water flux.

584 |
585 | Grass transpiration is thus clearly being under-represented by most of TBMs, and
586 | reasons for this could be due to multiple factors that we discuss below. The evolution of
587 | C_4 grasses to fix carbon under low light, low CO_2 concentrations and high temperatures
588 | has resulted in a gas-exchange process that is highly water-use efficient (von
589 | Caemmerer and Furbank, 1999). Consequently, this life form is abundant in tropical,
590 | water-limited ecosystems, where it can contribute to more than 50% of total LAI (2.0 to
591 | 2.5), particularly at high rainfall sites (Sea et al., 2011). The annual strategy of the C_4
592 | grasses at these sites is to indiscriminately expend all available resources to maximise
593 | productivity during the monsoon period, for growth and to increase leaf area. This
594 | therefore allows grass transpiration to exceed tree transpiration during the peak wet
595 | season as evergreen trees will be more conservative in their water-use, allowing them to
596 | remain active in the dry season (Eamus et al., 2001; Hutley et al., 2000; Scholes and
597 | Archer, 1997). Following this logic, our results suggest that the TBMs are either: i)
598 | incorrectly ascribing leaf area to the understory (i.e. the C_4 fractional cover is too low),
599 | ii) incorrectly describing the C_4 leaf-gas exchange physiology, iii) incorrectly describing
600 | the understory micro climatic environment (R_s , T_a , VPD), or iv) a combination of these
601 | causes. Furthermore, it should be noted that the TBMs used in this study are not truly
602 | modelling grasses, but approximating them. Grasses are effectively simulated as 'stem-
603 | less' trees, and the distinction between the two life forms is reliant on different
604 | parameter sets (e.g. V_{cmax} , height, etc.) and slight modifications of the same process (e.g.

605 [rate of assimilation, respiration, etc.](#)). While our results and the tower data do not allow
606 us to directly determine how C₄ grasses may be misrepresented in these TBMs, they
607 clearly indicate that future development and evaluation should be focused on these
608 issues. Eddy covariance studies of understory savanna vegetation as conducted by
609 Moore et al. (*this issue*) will be critical to this process.
610

611 **4.3 Savanna phenology**

612 The results from this study have shown that to simulate savanna fluxes, TBMs must be
613 able to simulate the dynamics of savanna phenology, expressed by LAI. This was
614 highlighted by the empirical benchmarks, where the results showed that while R_s , T_a and
615 VPD were important drivers, LAI was required to capture the seasonality and magnitude
616 of the fluxes to achieve good performance. LAI integrates the observed structural
617 changes of the savanna as annual rainfall declines with reduced woody stem density;
618 driving water and carbon exchange as a result (Kanniah et al., 2010; Ma et al., 2013; Sea
619 et al., 2011). If LAI is prescribed in a model, it is important that leaf area is partitioned
620 correctly between the trees and grass layers to describe their respective phenology. This
621 partitioning is important, as the C₄ grass understory explains most of the seasonal
622 variation in LAI, and is a consequence of an annual phenology that exhibits rapid growth
623 at the onset of the wet season and senescence at the onset of the dry (Williams et al.,
624 1996b). By contrast the evergreen eucalypt canopy shows modest reductions in canopy
625 leaf area during the dry season, especially as mean annual rainfall declines (Bowman
626 and Prior, 2005; Kelley et al., 2007). The strong seasonal dynamics of the grasses result
627 in large changes in LAI, with levels varying between 0.7 and 2.5 at high rainfall sites (Sea
628 et al., 2011). The phenological strategy of the C₄ grasses also changes with rainfall
629 interannual variability, with the onset of the greening period becoming progressively
630 delayed as sites become drier, to become eventually rain-pulse driven as the monsoonal
631 influence weakens (Ma et al., 2013).
632

633 With the exception of LPJGUESS, all models prescribed LAI as an input driver.
634 Prescribing LAI can be problematic depending on the time-scale and how it is
635 partitioned between trees and grass layers. At large time-steps (months) it will fail to
636 capture the rapidly changing dynamics of vegetation during the transition periods, and
637 this is particularly true for the onset of the wet season (Sep-Nov) especially at drier sites
638 that are subject to larger interannual rainfall variability (Hutley et al., 2011).
639 Additionally, as the sites become drier the tree:grass ratio will become smaller and this

Rhys Whitley 29/2/16 11:28 AM

Deleted: -

641 dynamic can be difficult to predict, although methods do exist (see Donohue et al. 2009).
642 From the results, we infer that TBMs that prescribe LAI and allow for a dynamic
643 representation of tree and grass ratios are better able to capture the changing dynamics
644 of the savanna system. This is a possible explanation for the better performance of [the](#)
645 [BIOS2](#), MAESPA and SPA models in simulating GPP as [these](#) models dynamically
646 partition leaf area between trees and grasses at the sub-monthly time-scale, rather than
647 [using](#) a bulk value. However, there are limitations to using prescribed LAI,
648 predominantly in that it describes a stable system, of which savannas are typically not;
649 having a large sensitivity to changes in climate; particularly rainfall variability and
650 disturbance (Sankaran et al., 2005). DGVMs that consider dynamic vegetation and use a
651 prognostic LAI can simulate the feedback between the climate and the relative cover of
652 trees and grasses, which shapes the savanna continuum. This feedback allows the
653 simulated savanna structure to potentially shift to alternate states (e.g. grassland or
654 forest) in response to changes in annual rainfall and fire severity (Scheiter and Higgins,
655 2007, 2009). While LPJGUESS was the only TBM to use a prognostic LAI in our study, it
656 achieved only moderate performance, and this may be due to how carbon is allocated
657 from the pool on an annual time step, such that it is not as dynamic as it could be.
658 However, its capability to simulate the feedback between climate and LAI is critical for
659 simulating how savanna dynamics may change from year to year. There may also be
660 issues with how phenology is simulated, particularly as it is determined from empirical
661 formulations, which are: i) not specifically developed for savanna environments and ii)
662 calculated before the growing season begins. Such formulations are therefore not
663 mechanistic, [and do not respond](#) to actual season dynamics (e.g. limiting soil water), but
664 are empirically determined (Richardson et al., 2013).

665

666 **5. Conclusions**

667

668 This study set out to assess how well a set of functionally different, state-of-the-art
669 TBMs perform at predicting the bulk exchanges of carbon and water over savanna land
670 surfaces. Our model inter-comparison has identified key weaknesses in the assumptions
671 of biosphere-atmosphere processes, which do not hold for savanna environments. Our
672 benchmarking has identified low model performance by TBMs is likely a result of
673 incorrect assumptions related to: i) deep soil water access, ii) a systematic under-
674 estimation of the contribution of the grass understorey in the wet season, and iii) the
675 use of static phenology to represent dynamic vegetation. Our results showed that these

676 assumptions, as they currently exist in TBMs, are not wholly supported by 'observations'
677 of savanna water and carbon exchange and need to be addressed if more reliable
678 projections are to be made on how savannas respond to environmental change. Despite
679 this, our benchmarking has shown that all TBMs could potentially operate well for
680 savanna ecosystems, provided that the above issues are developed on. We suggest that
681 further work investigates how particular processes in the models may be affecting
682 overall predicted water and carbon fluxes, and may include testing variable rooting
683 depths, alternate root-water uptake schemes and how these might affect leaf-level
684 outputs (e.g. stomatal conductance, leaf water potential) among TBMs, and different
685 phenology schemes.-The issues highlighted here also have scope beyond savanna
686 environments, and are relevant to other water-limited ecosystems. The results from this
687 study provide a foundation for improving how savanna ecosystem dynamics are
688 simulated.

689

690 **Acknowledgements**

691

692 This study was conducted as part of the 'Australian Savanna Landscapes: Past, Present
693 and Future' project funded by the Australian Research Council (DP130101566). The
694 support, collection and utilization of data were provided by the OzFlux network
695 (www.ozflux.org.au) and Terrestrial Ecosystem Research Network (TERN)
696 (www.tern.org.au), and funded by the ARC (DP0344744, DP0772981 and
697 DP130101566). PALS was partly funded by the TERN ecosystem Modelling and Scaling
698 infrAstructure (eMAST) facility under the National Collaborative Research
699 Infrastructure Strategy (NCRIS) 2013-2014 budget initiative of the Australian
700 Government Department of Industry. Rhys Whitley was supported through the ARC
701 Discovery Grant (DP130101566). Jason Beringer is funded under an ARC FT
702 (FT110100602). Vanessa Haverd's contribution was supported by the Australian
703 Climate Change Science Program. We acknowledge the support of the Australian
704 Research Council Centre of Excellence for Climate System Science (CE110001028).

705

706 | **References:**

- 707 Abramowitz, G.: Towards a public, standardized, diagnostic benchmarking system for
708 land surface models, *Geosci. Model Dev.*, 5(3), 819–827, doi:10.5194/gmd-5-819-2012,
709 2012.
- 710 Abramowitz, G., Leuning, R., Clark, M. and Pitman, A.: Evaluating the Performance of
711 Land Surface Models, *J. Clim.*, 21(21), 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.
- 712 Ball, J. T., Woodrow, I. E. and Berry, J. A.: A model predicting stomatal conductance and
713 its contribution to the control of photosynthesis under different environmental
714 conditions., in *Progress in Photosynthesis Research*, pp. 221–224, Martinus-Nijhoff
715 Publishers, Dordrecht, the Netherlands., 1987.
- 716 Bashtannyk, D. M. and Hyndman, R. J.: Bandwidth selection for kernel conditional
717 density estimation, *Comput. Stat. Data Anal.*, 36(3), 279–298, doi:10.1016/S0167-
718 9473(00)00046-3, 2001.
- 719 Beringer, J., Hutley, L. B., Tapper, N. J. and Cernusak, L. a.: Savanna fires and their impact
720 on net ecosystem productivity in North Australia, *Glob. Chang. Biol.*, 13(5), 990–1004,
721 doi:10.1111/j.1365-2486.2007.01334.x, 2007.
- 722 Beringer, J., Hutley, L. B., Abramson, D., Arndt, S. K., Briggs, P., Bristow, M., Canadell, J. G.,
723 Cernusak, L. a, Eamus, D., Evans, B. J., Fest, B., Goergen, K., Grover, S. P., Hacker, J.,
724 Haverd, V., Kanniah, K., Livesley, S. J., Lynch, A., Maier, S., Moore, C., Raupach, M., Russell-
725 Smith, J., Scheiter, S., Tapper, N. J. and Uotila, P.: Fire in Australian Savannas: from leaf to
726 landscape., *Glob. Chang. Biol.*, 11, 6641, doi:10.1111/gcb.12686, 2014.
- 727 Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M.,
728 Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J.,
729 Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello Jr, J. A., Stevens, L. and Vuichard, N.:
730 The plumbing of land surface models: benchmarking model performance, *J.*
731 *Hydrometeorol.*, 16, 1425–1442, 2015.
- 732 Bond, W. J.: What Limits Trees in C4 Grasslands and Savannas?, *Annu. Rev. Ecol. Evol.*
733 *Syst.*, 39(1), 641–659, doi:10.1146/annurev.ecolsys.39.110707.173411, 2008.
- 734 Bowman, D. M. J. S. and Prior, L. D.: Why do evergreen trees dominate the Australian
735 seasonal tropics?, *Aust. J. Bot.*, 53(5), 379–399, doi:10.1071/BT05022, 2005.
- 736 von Caemmerer, S. and Furbank, R. T.: Modeling C4 Photosynthesis, in *C4 Plant Biology*,
737 edited by R. F. Sage and R. K. Monson, pp. 173–211, Academic Press, Toronto., 1999.

738 Cernusak, L. A., Hutley, L. B., Beringer, J., Holtum, J. A. M. and Turner, B. L.:
739 Photosynthetic physiology of eucalypts along a sub-continental rainfall gradient in
740 northern Australia, *Agric. For. Meteorol.*, 151(11), 1462–1470,
741 doi:10.1016/j.agrformet.2011.01.006, 2011.

742 Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D.
743 D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., Wirth, C., Aber, J. D., Cole, J. J.,
744 Goulden, M. L., Harden, J. W., Heimann, M., Howarth, R. W., Matson, P. A., McGuire, A. D.,
745 Melillo, J. M., Mooney, H. A., Neff, J. C., Houghton, R. A., Pace, M. L., Ryan, M. G., Running, S.
746 W., Sala, O. E., Schlesinger, W. H. and Schulze, E. D.: Reconciling carbon-cycle concepts,
747 terminology, and methods, *Ecosystems*, 9(7), 1041–1050, doi:10.1007/s10021-005-
748 0105-7, 2006.

749 Chen, X., Hutley, L. B. and Eamus, D.: Carbon balance of a tropical savanna of northern
750 Australia, *Oecologia*, 137(3), 405–16, doi:10.1007/s00442-003-1358-5, 2003.

751 Collatz, G. J., Ball, J. T., Grivet, C. and Berry, J. A.: Physiological and environmental
752 regulation of stomatal conductance, photosynthesis and transpiration: a model that
753 includes a laminar boundary layer, *Agric. For. Meteorol.*, 54, 107–136, 1991.

754 Collatz, G. J., Ribas-Carbo, M. and Berry, J. A.: Coupled photosynthesis-stomatal
755 conductance model for leaves of C4 plants, *Funct. Plant Biol.*, 19(5), 519 – 538, 1992.

756 Donohue, R. J., Mcvicar, T. R. and Roderick, M. L.: Climate-related trends in Australian
757 vegetation cover as inferred from satellite observations, 1981-2006, *Glob. Chang. Biol.*,
758 15, 1025–1039, doi:10.1111/j.1365-2486.2008.01746.x, 2009.

759 Duursma, R. A. and Medlyn, B. E.: MAESPA: A model to study interactions between water
760 limitation, environmental drivers and vegetation function at tree and stand levels, with
761 an example application to [CO₂] ?? drought interactions, *Geosci. Model Dev.*, 5(4), 919–
762 940, doi:10.5194/gmd-5-919-2012, 2012.

763 Eamus, D., O’Grady, A. P. O. and Hutley, L. B.: Dry season conditions determine wet
764 season water use in the wet-tropical savannas of northern Australia, *Tree Physiol.*,
765 20(18), 1219–1226, 2000.

766 Eamus, D., Hutley, L. B. and O’Grady, A. P. O.: Daily and seasonal patterns of carbon and
767 water fluxes above a north Australian savanna, *Tree Physiol.*, 21(12-13), 977–88, 2001.

768 Eamus, D., Chen, X., Kelley, G. and Hutley, L. B.: Root biomass and root fractal analyses of
769 an open Eucalyptus forest in a savanna of north Australia, *Aust. J. Bot.*, 50, 31–41, 2002.

770 Eamus, D., Cleverly, J., Boulain, N., Grant, N., Faux, R. and Villalobos-Vega, R.: Carbon and
771 water fluxes in an arid-zone Acacia savanna woodland: An analyses of seasonal patterns
772 and responses to rainfall events, *Agric. For. Meteorol.*, 182-183, 225–238,
773 doi:10.1016/j.agrformet.2013.04.020, 2013.

774 Farquhar, G. D., von Caemmerer, S. and Berry, J. A.: A Biochemical Model of
775 Photosynthetic CO₂ Assimilation in Leaves of C₃ species, *Planta*, 149, 78–90, 1980.

776 Grace, J., Jose, J. S., Meir, P., Miranda, H. S. and Montes, R. A.: Productivity and carbon
777 fluxes of tropical savannas, *J. Biogeogr.*, 33(3), 387–400, doi:10.1111/j.1365-
778 2699.2005.01448.x, 2006.

779 Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Pickett-Heaps, C.,
780 Roxburgh, S. H., Van Gorsel, E., Viscarra Rossel, R. a. and Wang, Z.: Multiple observation
781 types reduce uncertainty in Australia’s terrestrial carbon and water cycles,
782 *Biogeosciences*, 10(3), 2011–2040, doi:10.5194/bg-10-2011-2013, 2013.

783 Haverd, V., Smith, B., Raupach, M. R., Briggs, P. R., Nieradzik, L., Beringer, J., Hutley, L. B.,
784 Trudinger, C. M. and Cleverly, J. R.: Coupling carbon allocation with leaf and root
785 phenology predicts tree-grass partitioning along a savanna rainfall gradient. *Vanessa*,
786 *Biogeosciences*, 2015.

787 Haxeltine, A. and Prentice, I. C.: A general model for the light-use efficiency of primary
788 production, *Funct. Ecol.*, 10(5), 551–561, doi:10.2307/2390165, 1996.

789 Higgins, S. I. and Scheiter, S.: Atmospheric CO₂ forces abrupt vegetation shifts locally,
790 but not globally., *Nature*, 488, 209–212, doi:10.1038/nature11238, 2012.

791 Hsu, K.: Self-organizing linear output map (SOLO): An artificial neural network suitable
792 for hydrologic modeling and analysis, *Water Resour. Res.*, 38(12), 1–17,
793 doi:10.1029/2001WR000795, 2002.

794 Huntingford, C. and Monteith, J. L.: The behaviour of a mixed-layer model of the
795 convective boundary layer coupled to a big leaf model of surface energy partitioning,
796 *Boundary-Layer Meteorol.*, 88(1), 87–101, doi:10.1023/A:1001110819090, 1998.

797 Hutchinson, M. F. and Xu, T.: ANUCLIM v6.1, Fenner School of Environment and Society,
798 Australian National University, Canberra, ACT., 2010.

799 Hutley, L. B., O’Grady, A. P. O. and Eamus, D.: Evapotranspiration from Eucalypt open-
800 forest savanna of Northern Australia, *Funct. Ecol.*, 14, 183–194, 2000.

801 Hutley, L. B., Grady, A. P. O. and Eamus, D.: Monsoonal influences on evapotranspiration

802 of savanna vegetation of northern Australia, *Oecologia*, 126(3), 434–443,
803 doi:10.1007/s004420000539, 2001.

804 Hutley, L. B., Beringer, J., Isaac, P. R., Hacker, J. M. and Cernusak, L. A.: A sub-continental
805 scale living laboratory: Spatial patterns of savanna vegetation over a rainfall gradient in
806 northern Australia, *Agric. For. Meteorol.*, 151(11), 1417–1428,
807 doi:10.1016/j.agrformet.2011.03.002, 2011.

808 Isbell, R.: *The Australian Soil Classification, Revised Ed.*, CSIRO Publishing, Collingwood,
809 Victoria., 2002.

810 Kanniah, K. D., Beringer, J. and Hutley, L. B.: The comparative role of key environmental
811 factors in determining savanna productivity and carbon fluxes: A review, with special
812 reference to northern Australia, *Prog. Phys. Geogr.*, 34(4), 459–490,
813 doi:10.1177/0309133310364933, 2010.

814 De Kauwe, M. G., Zhou, S.-X., Medlyn, B. E., Pitman, A. J., Wang, Y. P., Duursma, R. A. and
815 Prentice, I. C.: Do land surface models need to include differential plant species
816 responses to drought? Examining model predictions across a latitudinal gradient in
817 Europe, *Biogeosciences Discuss.*, 12(15), 12349–12393, doi:10.5194/bgd-12-12349-
818 2015, 2015.

819 Kelley, G., O’Grady, A. P., Hutley, L. B. and Eamus, D.: A comparison of tree water use in
820 two contiguous vegetation communities of the seasonally dry tropics of northern
821 Australia: The importance of site water budget to tree hydraulics, *Aust. J. Bot.*, 55(7),
822 700–708, doi:10.1071/BT07021, 2007.

823 Koch, G. W., Vitousek, P. M., Steffen, W. L. and Walker, B. H.: Terrestrial transects for
824 global change research, *Vegetation*, 121(1-2), 53–65, doi:10.1007/BF00044672, 1995.

825 Kowalczyk, E. A., Wang, Y. P. and Law, R. M.: The CSIRO Atmosphere Biosphere Land
826 Exchange (CABLE) model for use in climate models and as an offline model, Aspendale,
827 Victoria., 2006.

828 Lehmann, C. E. R., Prior, L. D. and Bowman, D. M. J. S.: Decadal dynamics of tree cover in
829 an Australian tropical savanna, , 601–612, doi:10.1111/j.1442-9993.2009.01964.x,
830 2009.

831 Lehmann, C. E. R., Anderson, T. M., Sankaran, M., Higgins, S. I., Archibald, S., Hoffmann, W.
832 A., Hanan, N. P., Williams, R. J., Fensham, R. J., Felfili, J., Hutley, L. B., Ratnam, J., San Jose,
833 J., Montes, R., Franklin, D., Russell-Smith, J., Ryan, C. M., Durigan, G., Hiernaux, P., Haidar,
834 R., Bowman, D. M. J. S. and Bond, W. J.: Savanna vegetation-fire-climate relationships

835 differ among continents., *Science* (80-.), 343, 548–552, doi:10.1126/science.1247355,
836 2014.

837 Leuning, R.: A critical appraisal of a combined stomatal-photosynthesis model for C3
838 plants, *Plant, Cell Environ.*, 18(4), 339–355, 1995.

839 Li, L., Wang, Y. P., Yu, Q., Pak, B., Eamus, D., Yan, J., Van Gorsel, E. and Baker, I. T.:
840 Improving the responses of the Australian community land surface model (CABLE) to
841 seasonal drought, *J. Geophys. Res. Biogeosciences*, 117(4), 1–16,
842 doi:10.1029/2012JG002038, 2012.

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

847 McKenzie, N. N., Jacquier, D., Isbell, R. and Brown, K.: *Australian soils and landscapes : an
848 illustrated compendium*, CSIRO Publishing, Collingwood, Victoria., 2004.

849 Medlyn, B. E., Robinson, A. P., Clement, R. and McMurtrie, R. E.: On the validation of
850 models of forest CO2 exchange using eddy covariance data: some perils and pitfalls, *Tree
851 Physiol.*, 25, 839–857, 2005.

852 Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M.,
853 Crous, K. Y., De Angelis, P., Freeman, M. and Wingate, L.: Reconciling the optimal and
854 empirical approaches to modelling stomatal conductance, *Glob. Chang. Biol.*, 17(6),
855 2134–2144, doi:10.1111/j.1365-2486.2010.02375.x, 2011.

856 Moore, C. E., Beringer, J., Evans, B., Hutley, L. B., McHugh, I. and Tapper, N. J.: The
857 contribution of trees and grasses to productivity of an Australian tropical savanna,
858 *Biogeosciences*, doi:10.5194/bgd-12-19307-2015, this issue.

859 O’Grady, A. P. O., Eamus, D. and Hutley, L. B.: Transpiration increases during the dry
860 season: patterns of tree water use in eucalypt open-forests of northern Australia., *Tree
861 Physiol.*, 19(9), 591–597, 1999.

862 Parton, W. J., Anderson, D. W., Cole, C. V and Stewart, J. W. B.: Simulation of soil organic
863 matter formation and mineralization in semiarid agroecosystems, in *Nutrient Cycling In
864 Agricultural Ecosystems*, vol. 23, pp. 533–550., 1983.

865 Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for
866 climate models, *Int. J. Climatol.*, 23, 479–510, doi:10.1002/joc.893, 2003.

867 Prior, L. D., Eamus, D. and Duff, G. A.: Seasonal Trends in Carbon Assimilation, Stomatal
868 Conductance, Pre-dawn Leaf Water Potential and Growth in *Terminalia ferdinandiana*, a
869 Deciduous Tree of Northern Australian Savannas, *Aust. J. Bot.*, 45, 53–69, 1997.

870 Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O. and Toomey, M.:
871 Climate change, phenology, and phenological control of vegetation feedbacks to the
872 climate system, *Agric. For. Meteorol.*, 169, 156–173,
873 doi:10.1016/j.agrformet.2012.09.012, 2013.

874 Russell-Smith, J. and Edwards, A. C.: Seasonality and fire severity in savanna landscapes
875 of monsoonal northern Australia, *Int. J. Wildl. Fire*, 15(4), 541–550,
876 doi:10.1071/WF05111, 2006.

877 Ryu, Y., Baldocchi, D. D., Kobayashi, H., van Ingen, C., Li, J., Black, T. A., Beringer, J., van
878 Gorsel, E., Knohl, A., Law, B. E. and Rouspard, O.: Integration of MODIS land and
879 atmosphere products with a coupled-process model to estimate gross primary
880 productivity and evapotranspiration from 1 km to global scales, *Global Biogeochem.*
881 *Cycles*, 25(GB4017), doi:10.1029/2011GB004053, 2011.

882 Ryu, Y., Baldocchi, D. D., Black, T. A., Detto, M., Law, B. E., Leuning, R., Miyata, A.,
883 Reichstein, M., Vargas, R., Ammann, C., Beringer, J., Flanagan, L. B., Gu, L., Hutley, L. B.,
884 Kim, J., McCaughey, H., Moors, E. J., Rambal, S. and Vesala, T.: On the temporal upscaling
885 of evapotranspiration from instantaneous remote sensing measurements to 8-day mean
886 daily-sums, *Agric. For. Meteorol.*, 152(1), 212–222,
887 doi:10.1016/j.agrformet.2011.09.010, 2012.

888 Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., Gignoux,
889 J., Higgins, S. I., Le Roux, X., Ludwig, F., Ardo, J., Banyikwa, F., Bronn, A., Bucini, G., Caylor,
890 K. K., Coughenour, M. B., Diouf, A., Ekaya, W., Feral, C. J., February, E. C., Frost, P. G. H.,
891 Hiernaux, P., Hrabar, H., Metzger, K. L., Prins, H. H. T., Ringrose, S., Sea, W., Tews, J.,
892 Worden, J. and Zambatis, N.: Determinants of woody cover in African savannas., *Nature*,
893 438, 846–849, doi:10.1038/nature04070, 2005.

894 Scheiter, S. and Higgins, S. I.: Partitioning of root and shoot competition and the stability
895 of savannas., *Am. Nat.*, 170(4), 587–601, doi:10.1086/521317, 2007.

896 Scheiter, S. and Higgins, S. I.: Impacts of climate change on the vegetation of Africa: an
897 adaptive dynamic vegetation modelling approach, *Glob. Chang. Biol.*, 15(9), 2224–2246,
898 doi:10.1111/j.1365-2486.2008.01838.x, 2009.

899 Scheiter, S., Higgins, S. I., Beringer, J. and Hutley, L. B.: Climate change and long-term fire

900 management impacts on Australian savanna, *Glob. Chang. Biol.*, Submitted , 1211–1226,
901 doi:10.1111/nph.13130, 2014.

902 Scholes, R. J. and Archer, S. R.: Tree-grass interactions in savannas, *Annu. Rev. Ecol. Syst.*,
903 28, 517–544, doi:10.1146/annurev.ecolsys.28.1.517, 1997.

904 Schymanski, S. J., Roderick, M. L., Sivapalan, M., Hutley, L. B. and Beringer, J.: A test of the
905 optimality approach to modelling canopy properties and CO₂ uptake by natural
906 vegetation., *Plant. Cell Environ.*, 30(12), 1586–98, doi:10.1111/j.1365-
907 3040.2007.01728.x, 2007.

908 Schymanski, S. J., Sivapalan, M., Roderick, M. L., Beringer, J. and Hutley, L. B.: An
909 optimality-based model of the coupled soil moisture and root dynamics, *Hydrol. Earth
910 Syst. Sci.*, 12(3), 913–932, doi:10.5194/hess-12-913-2008, 2008.

911 Schymanski, S. J., Sivapalan, M., Roderick, M. L., Hutley, L. B. and Beringer, J.: An
912 optimality-based model of the dynamic feedbacks between natural vegetation and the
913 water balance, *Water Resour. Res.*, 45(1), doi:10.1029/2008WR006841, 2009.

914 Sea, W. B., Choler, P., Beringer, J., Weinmann, R. a., Hutley, L. B. and Leuning, R.:
915 Documenting improvement in leaf area index estimates from MODIS using
916 hemispherical photos for Australian savannas, *Agric. For. Meteorol.*, 151(11), 1453–
917 1461, doi:10.1016/j.agrformet.2010.12.006, 2011.

918 Simioni, G., Roux, X. Le, Gignoux, J. and Sinoquet, H.: Treegrass: a 3D, process-based
919 model for simulating plant interactions in tree–grass ecosystems, *Ecol. Modell.*, 131, 47–
920 63, 2000.

921 Simioni, G., Gignoux, J. and Le Roux, X.: Tree layer spatial structure can affect savanna
922 production and water budget: Results of a 3-D model, *Ecology*, 84(7), 1879–1894,
923 doi:10.1890/0012-9658(2003)084[1879:TLSSCA]2.0.CO;2, 2003.

924 Sitch, S., Smith, B., Prentice, I. C., Arneeth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S.,
925 Lucht, W., Sykes, M., Thonicke, K. and Venevski, S.: Evaluation of ecosystem dynamics,
926 plant geography and terrestrial carbon cycling in the LPJ dynamic vegetation model,
927 *Glob. Chang. Biol.*, 9, 161–185, 2003.

928 Smith, B., Prentice, I. C. and Sykes, M. T.: Representation of vegetation dynamics in the
929 modelling of terrestrial ecosystems: Comparing two contrasting approaches within
930 European climate space, *Glob. Ecol. Biogeogr.*, 10(6), 621–637, doi:10.1046/j.1466-
931 822X.2001.t01-1-00256.x, 2001.

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

935 Wang, Y. P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M. R., Pak, B., Van
936 Gorsel, E. and Luhar, A.: Diagnosing errors in a land surface model (CABLE) in the time
937 and frequency domains, *J. Geophys. Res. Biogeosciences*, 116, 1–18,
938 doi:10.1029/2010JG001385, 2011.

939 van der Werf, G. R., Randerson, J. T., Giglio, L., Gobron, N. and Dolman, a. J.: Climate
940 controls on the variability of fires in the tropics and subtropics, *Global Biogeochem.*
941 *Cycles*, 22(3), 1–13, doi:10.1029/2007GB003122, 2008.

942 Whitley, R. J., Macinnis-Ng, C. M. O., Hutley, L. B., Beringer, J., Zeppel, M., Williams, M.,
943 Taylor, D. and Eamus, D.: Is productivity of mesic savannas light limited or water
944 limited? Results of a simulation study, *Glob. Chang. Biol.*, 17(10), 3130–3149,
945 doi:10.1111/j.1365-2486.2011.02425.x, 2011.

946 Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R.,
947 Melillo, J. M., Munger, J. W., Fan, S.-M. and Nadelhoffer, K. J.: Modelling the soil-plant-
948 atmosphere continuum in a *Quercus Acer* stand at Harvard Forest: the regulation of
949 stomatal conductance by light, nitrogen and soil/plant hydraulic properties, *Plant, Cell*
950 *Environ.*, 19, 911–927, 1996a.

951 Williams, R. J., Duff, G. A., Bowman, D. M. J. S. and Cook, G. D.: Variation in the
952 composition and structure of tropical savannas as a function of rainfall and soil texture
953 along a large-scale climatic gradient in the Northern Territory, Australia, *J. Biogeogr.*,
954 23(6), 747–756, doi:10.1111/j.1365-2699.1996.tb00036.x, 1996b.

955

	Howard Springs ^a	Adelaide River ^b	Daly Uncleared ^c	Dry River ^d	Sturt Plains ^e	
Years (inclusive)	2001 – 2012	2007 – 2009	2008 – 2012	2008 – 2012	2008 – 2012	
Co-ordinates	12°29'39.12" S	13°04'36.84" S	14°09'33.12" S	15°15'31.62" S	17°09'02.76" S	
	131°09'09" E	131°07'04.08" E	131°23'17.16" E	132°22'14.04" E	133°21'01.14" E	
Elevation (m)	64	90	110	175	250	
^fMeteorology						
Annual Rainfall (mm)	1714	1460	1170	850	535	
Min/Max Daily Temperature (°C)	22.0/33.0	21.8/35.3	20.8/35.0	20.0/34.8	19.0/34.2	
Min/Max Absolute Humidity (g m ⁻³)	11.0/18.5	8.9/17.7	8.6/15.1	7.8/12.3	6.1/9.0	
Min/Max Soil Moisture (m ³ m ⁻³)	0.06/0.1	0.09/0.14	0.03/0.06	0.03/0.05	0.04/0.1	
Soil Temperature (°C)	32.7	35.7	32.8	<i>n.a.</i>	30.2	
Solar Radiation (W m ⁻²)	256.5	258.1	270.6	266.5	269.7	
Bowen Ratio	1.7	3.1	3.2	4.6	15.8	
^fVegetation						
Overstorey species	<i>Eu. Miniata</i>	<i>Eu. tectifera</i>	<i>Te. grandiflora</i>	<i>Eu. tetradonta</i>	<i>n.a.</i>	
	<i>Eu. tetradonta</i>	<i>Pl. careya</i>	<i>Eu. tetradonta</i>	<i>Co. terminalis</i>		
	<i>Er. chlorostachys</i>	<i>Co. latifolia</i>	<i>Co. latifolia</i>	<i>Eu. dichromophloia</i>		
Understorey species	<i>Sorghum</i> spp.	<i>Sorghum</i> spp.	<i>Sorghum</i> spp.	<i>Sorghum intrans</i>	<i>Astrabla</i> spp.	
	<i>He. triticeus</i>	<i>Ch. fallax</i>	<i>He. triticeus</i>	<i>Th. Tiandra</i>		
				<i>Ch. fallax</i>		
Basal Area (m ² ha ⁻¹)	9.7	5.1	8.3	5.4	<i>n.a.</i>	
Canopy Height (m)	18.9	12.5	16.4	12.3	0.2	
LAI (m ² m ⁻²)	1.04 ± 0.07	0.68 ± 0.07	0.80 ± 0.12	0.58 ± 0.11	0.39 ± 0.11	
Total Leaf Nitrogen (g m ⁻³)	1.42 ± 0.20	1.27 ± 0.18	1.35 ± 0.19	1.97 ± 0.15	2.37 ± 0.17	
^gSoil						
Type	Red kandosol	Yellow hydrosol	Red kandosol	Red kandosol	Grey vertosol	
A Horizon	Texture	Sandy loam	Sandy loam	Loam	Clay loam	
	Clay PSD (%)	15	20	20	50	20
	Sand PSD (%)	60	50	40	25	40
	Thickness (m)	0.30	0.30	0.20	0.15	0.20
	Bulk Density (Mg m ⁻³)	1.29	1.60	1.39	1.20	1.39
	Hydraulic Conductivity (mm hr ⁻¹)	9	7	9	3	9
	Field Capacity (mm m ⁻¹)	156	132	147	140	147
B Horizon	Texture	Clay loam	Clay	Clay loam	Clay	Clay loam
	Clay PSD (%)	40	55	35	55	35
	Sand PSD (%)	30	20	30	20	30
	Thickness (m)	1.20	0.60	0.69	1.29	0.69
	Bulk Density (Mg m ⁻³)	1.39	1.70	1.39	1.39	1.39
	Hydraulic Conductivity (mm hr ⁻¹)	8	5	7	2	7
	Field Capacity (mm m ⁻¹)	146	31	146	107	146

Table 1: Summarised dataset information for each of the five savanna sites used in this study. This includes site descriptions pertaining to local meteorology, vegetation and below ground soil characteristics. Where data were not available, the abbreviation *n.a.* is used. Definitions for the species genus mentioned in the table are as follows: *Eucalyptus* (*Eu.*), *Erythrophleum* (*Er.*), *Terminalia* (*Te.*), *Corymbia* (*Co.*), *Planchonia* (*Pl.*), *Buchanania* (*Bu.*), *Themda* (*Th.*), *Hetropogan* (*He.*), and *Chrysopogon* (*Ch.*). Eddy covariance datasets relating to each of the 5 sites here can be download from www.ozflux.org.au and hdl references are given by order of column (Jason Beringer (2013) – ^ahdl: 102.100.100/14228, ^bhdl: 102.100.100/14239, ^chdl: 102.100.100/14229, ^dhdl: 102.100.100/14234, ^ehdl: 102.100.100/14230). Site meteorology is given as 30 year averages with values taken from ^fHutley, et al. (2011). Soil descriptions are taken from the Digital Atlas of Australian Soils (www.asris.csiro.au) ^gIsbell, (2002).

Model Name	SPA	MAESPA	CABLE	BIOS2	BESS	LPJGUSS
Model definition	Soil-Plant-Atmosphere Model	MAESTRA-SPA	Community Atmosphere Biosphere Land-surface Exchange Model	Modified CABLE (CABLE + SLI + CASA-CNP)	Breathing Earth System Simulator	Lund-Potsdam-Jena General Ecosystem Simulator
Version	1.0	1.0	2.0	2.0	1.0	2.1
Reference	Williams et al. (1996a)	Duursma & Medlyn (2012)	Kowalyzck et al. (2006), Wang et al. (2011)	Haverd et al. (2013)	Ryu et al. (2011, 2012)	Smith et al. (2001)
Temporal resolution	30-min	30-min	30-min	Daily (30-min time-steps are generated from daily time-series)	Snap shot with MODIS overpass, then up-scaled to a daily and 8-day time series	Daily
Spatial resolution	Point	Point	0.05° (5 km)	0.05° (5 km)	0.05° (5 km)	Patch (c. 0.1 ha)
Functional class	Stand model	Individual Plant or Stand Model	Land-Surface Model	Land-Surface Model	Remote Sensing Model	Dynamic Global Vegetation Model
Canopy Description						
C₃ Assimilation	Farquhar et al. (1980)	Farquhar et al. (1980)	Farquhar et al. (1980)	Farquhar et al. (1980)	Farquhar et al. (1980)	Collatz et al. (1991)
C₄ Assimilation	Collatz et al. (1992)	Collatz et al. (1992)	Collatz et al. (1992)	Collatz et al. (1992)	Collatz et al. (1992)	Collatz et al. (1992)
Stomatal conductance	Williams et al. (1996a)	Medlyn et al. (2011)	Leuning (1995)	Leuning (1995)	Ball et al. (1987)	Haxeltine & Prentice (1996)
Transpiration	Penman-Monteith calculated at leaf-scale accounting for g_b and limitation of soil-water supply via ψ_l	Penman-Monteith calculated at the leaf scale	Penman-Monteith	Penman-Monteith	Penman-Monteith	Haxeltine & Prentice (1996)
Boundary layer resistance	$f(\text{wind speed, leaf width, air temperature})$	$f(\text{wind speed, leaf width, air temperature and atmospheric pressure})$	$f(\text{wind speed, leaf width, air temperature})$	$f(\text{wind speed, leaf width, air temperature})$	Not Modelled	Huntingford & Monteith (1998)
Aerodynamic resistance	$f(\text{wind speed, canopy height})$	Not calculated unless transpiration is calculated at the canopy scale, in which case g_b above isn't calculated.	$f(\text{wind speed, canopy height})$	$f(\text{wind speed, canopy height})$	$f(\text{wind speed, canopy height})$	Huntingford & Monteith (1998)
Leaf area index	Prescribed (MODIS)	Prescribed (MODIS)	Prescribed (MODIS)	Prescribed (MODIS)	Prescribed (MODIS)	Prognostic (C allocation)
Canopy structure	Canopy + understorey divided into 10 layers	Individual plant crowns, spatially explicit locations and uniform understorey	2 (tree/grass) big leaf (sunlit/shaded)	2 (tree/grass) big leaf (sunlit/shaded)	2 (tree/grass) big leaf (sunlit/shaded)	5-year age/size cohorts for trees, single-layer grass understorey
C₃:C₄ fraction	Dynamic ratio variable with time. Compete for water and light.	Dynamic ratio variable with time. Compete for water and light.	Simulated as independent layers	Dynamic ratio variable with time. Compete for water not light.	Still et al. (2003) Ratio changes 70:30 to 10:90 down transect	Prognostic, determined as the outcome of the competition with trees
Canopy interception	YES	YES	YES	YES	NO	YES
Simulates growth	NO	NO	NO	NO	NO	YES
Soil Profile Description						
Soil profile structure	Profile divided into N layers (prescribed - 20 in this case.)	Profile divided into N layers (prescribed - 20 in this case.)	Profile divided into 6 layers	Profile divided into 12 layers (adjustable)	Not Modelled	2 layers (0-0.5, 0.5-2 m) with 10 cm evaporation sub-layer
Soil hydraulic properties	Function of sand and clay particle size distributions	Function of sand and clay particle size distributions	Prescribed	Australian Soils Resource Information System (ASRIS)	Not Modelled	Sitch et al. (2003)
Soil depth	6.5 m	5.0 m	4.5 m	10.0 m	Not Modelled	2 m
Root depth	6.5 m	5.0 m	4.5 m	0.5 m (grasses), 5.0 m (trees)	Not Modelled	2 m
Root distribution	Prescribed; exponential decay as a function of surface biomass and the total root biomass of the column	Prescribed; exponential decay as a function of surface biomass and the total root biomass of the column	Prescribed; exponential decay	Prescribed; exponential decay	Not Modelled	PFT-specific, trees have deeper roots on average
Soil-water stress modifier	E_t via g_s is increased to meet atmospheric demand while ψ_l remains above a critical threshold	Maximum transpiration rate calculated from hydraulic conductance (soil-to-leaf) sets limit on actual transpiration, OR uses the Tuzet et al. (2003) model of stomatal conductance	Supply/Demand	g_s scaled by a soil moisture limitation function related to extractible water accessible by roots	Assumes LAI and seasonal variation of V_{cmax} reflect soil water stress	Supply/Demand
Hydraulic pathway resistance	$R_{soil} + R_{plant}$	$R_{soil} + R_{plant}$	Not Modelled	Not Modelled	Not Modelled	Not explicit, min(supply, demand) determines sapflow

965
966 **Table 2:** Summary table of the ecosystem models used in the experiment; highlighting differences and similarities in model structure and
967 shared processes. Information is broken down into how each model describes aboveground canopy and belowground soil processes.
968

Statistical Metric	Definition
Correlation coefficient (r)	$\frac{n \sum_{i=1}^n (O_i M_i) - \sum_{i=1}^n O_i \sum_{i=1}^n M_i}{\sqrt{\left(n \sum_{i=1}^n O_i^2 - \left(\sum_{i=1}^n O_i \right)^2 \right) \left(n \sum_{i=1}^n M_i^2 - \left(\sum_{i=1}^n M_i \right)^2 \right)}}$
Standard Deviation (sd)	$\sqrt{1 - \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (M_i - \bar{M})^2}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (O_i - \bar{O})^2}}}$
Normalised mean error (NME)	$\frac{\sum_{i=1}^n M_i - O_i }{\sum_{i=1}^n \bar{O} - O_i }$
Normalised mean bias (MBE)	$\frac{1}{n} \sum_{i=1}^n (M_i - O_i)$

969

970

Table A1: Definition of common metrics used to determine ranks against the empirical benchmarks.

971

The terms M and O stand for model and observations respectively, while n denotes the length of the

972

data, and i is the datum.

973

974 **Figure Captions**

975

976 **Figure 1:** The Northern Territory of Australia and the North Australian Tropical Transect (NATT)
977 showing (a) the flux site locations with an accompanying 30-year (1970 to 2000) expression of the
978 average meteorological conditions for (b) mean annual temperature, and (c) total annual
979 precipitation derived from ANUCLIM v6.1 climate surfaces (Hutchinson and Xu, 2010).

980

981 **Figure 2:** Time-series of daily mean latent heat (LE) flux and gross primary productivity (GPP)
982 depicting an average year for each of the 5 savanna sites using a smoothed, 7-day moving average.
983 The sites are ordered from wettest to driest; (a) Howard Springs, (b) Adelaide River, (c) Daly River,
984 (d) Dry River and (e) Sturt Plains. The joined, black dots are the tower flux time-series, while the
985 grey lines are the performance benchmarks (emp1, emp2, emp3). Predictions of LE and GPP for each
986 of the six terrestrial biosphere models are given by a spectrum of colours described in the legend.

987

988 **Figure 3:** Probability densities (expressed in scientific notation) of daily mean latent heat (LE) flux
989 and gross primary productivity (GPP) at each of the 5 savanna sites, where the distributions for each
990 flux are partitioned into wet and dry seasons. The order of the sites are from wettest to driest; (a)
991 Howard Springs, (b) Adelaide River, (c) Daly River, (d) Dry River and (e) Sturt Plains. The grey
992 region is the tower flux, while the dotted lines are the empirical benchmarks. Predicted LE and GPP
993 probability densities from each of the six process-based models are given by a spectrum of colours
994 described in the legend.

995

996 **Figure 4:** Standardised model residuals for latent energy (LE) and gross primary productivity (GPP)
997 expressed in units of standard deviations (sd) $[(\text{modelled flux} - \text{observed flux})/\text{sd}(\text{observed flux})]$.
998 Residuals are presented for each model: (a) CABLE, (b) BIOS2, (c) LPJGUESS, (d) MAESPA, (d) BESS
999 and (e) SPA, where each flux site is represented by a blue-green-yellow gradient. For both fluxes, the
1000 residuals are plotted against time (ensemble average year) and against the flux prediction (bias).

1001

1002 **Figure 5:** Average rank plot showing the performance of the terrestrial biosphere models for all
1003 sites across the North Australian Tropical Transect (NATT) ordered in terms of annual rainfall as
1004 follows: Howard Springs (HowSpr), Adelaide River (AdrRiv), Daly Uncleared (DalUnc), Dry River
1005 (DryRiv), and Sturt Plains (StuPla). Models are individually ranked against the benchmarks in order
1006 of 1 to 4 (1 model + 3 benchmarks) and express the amount of metrics the models are meeting listed
1007 in Table S1. The rankings are determined individually for latent energy (LE) and gross primary
1008 productivity (GPP). The coloured lines represent each of the 6 models in the study, while the grey

1009 lines represent the empirical benchmarks. The average ranking for each model was determined for
1010 (a) a complete year, (b) the wet season and (c) the dry season.

1011

1012 **Figure 6:** Average year outputs of vegetation transpiration (grass + trees) and soil evaporation, as
1013 well as their percentage contributions to total latent energy (LE) for each of the 6 terrestrial
1014 biosphere models at each of the 5 savanna sites.

1015

1016 **Figure S1:** A smoothed (7-day moving average) representation of the environmental drivers used to
1017 construct the empirical benchmarks at each of the 5 savanna sites, and are shown from wettest to
1018 driest: (a) Howard Springs, (b) Adelaide River, (c) Daly River, (d) Dry River and (e) Sturt Plains. The
1019 time-series represents the seasonality over an average year for mean daily solar radiation (R_s), mean
1020 daily air temperature (T_a), mean daily vapour pressure deficit (VPD) and leaf area index (LAI).

1021