Evaluating a land surface model at a water-limited site: implications for land surface contributions to droughts and heatwaves

Response to Reviewer 2

We thank the reviewer for the positive comments and constructive suggestions. We have addressed the various concerns below. Our responses to reviewer comments are highlighted in blue below each reviewer comment.

Mu et al., evaluate the performance of the Community Atmosphere-Biosphere Land Exchange (CABLE) land surface model for a water-limited measurement site in southeastern Australia. The stand-alone model performance is assessed by comparing the simulation results to soil moisture and evapotranspiration measurements. By changing specific model configurations, the general model bias is tried to be reduced. In this context, one focus of the study is on heatwaves and droughts. Results show that a meaningful improvement of the model performance can only be achieved if both quantities, soil moisture and evapotranspiration, are considered for model validation.

Recommendation: The study is within the scope of HESS and addresses a relevant and interesting topic for the modelling community. The manuscript is well structured and comprehensibly written. Nevertheless, there are some issues which should be addressed before publication.

Comments:

1) The study highlights the large uncertainties related to the simulation of evapotranspiration and soil moisture. Thus, averaged land use specific parameters used in LSMs can deviate considerably from the actual hydrological characteristics at measurement sites. Large differences between simulation results and observations are the consequence. In order to improve the model performance, therefore, model configurations have to be adapted. This issue is clearly and comprehensibly demonstrated in the manuscript. But due to such site-specific changes the adapted model can only be applied at the location for which it is tuned and the model results are not transferable to other situations (or would the authors say that the results are transferable? If yes, please discuss it). Therefore, it is difficult to state lessons learned from this study beyond its specific application on southeastern Australia.

We agree that our study is site-specific and that some of the insights gained are only applicable at this scale. Nevertheless, we attempted to frame our discussion (as the reviewer notes below) in a general sense, such that findings would be of interest to other model groups. Similarly, whilst it is speculative to state how widespread the biases we identified are, it has been demonstrated that CABLE displays similar biases in other water-limited ecosystems as well as mesic sites (Haverd et al., 2016, Decker et al., 2017; Ukkola et al., 2016b). Furthermore, similar biases have also been identified in evaluations of other state-of-the-art LSMs over multiple sites (De Kauwe et al., 2017; Powell et al., 2013; Ukkola et al., 2016a).

We have added more details to our original discussion on the common problems in LSMs and how the CABLE processes explored here offer lessons for other LSMS:

[line 514-515]: "*LSMs commonly overestimate soil evaporation especially under a sparse canopy or* *over bare land (De Kauwe et al., 2017; Swenson and Lawrence, 2014), suggesting this is a key model weakness.*"

[line 538-540]: "*When we initialised from a drier starting position (Watr), the simulated soil moisture profile matched the observed better, with implications for other models using similar groundwater schemes (e.g. CLM4.5, Noah-MP, JULES and LEAFHYDRO).*"

[line 552-554]: "*LSMs typically define a fixed number of soil layers globally, anywhere up to 20 layers. Most LSMs assume constant parameters across the entire soil profile, either using an experimental look-up table based on soil classification or estimating parameters from empirical pedotransfer functions.*"

[line 588-590]: "*Studies commonly highlight the functions used to limit photosynthesis and stomatal conductance with water stress as a key weakness among models. The lack of theory in this space (Medlyn et al., 2016) has led to models employing a range of functions encompassing different shapes and sensitivities that are not constrained by data.*"

[line 598-599]: "*the linear θ-based function used in Ctl (common among models, e.g. SDGVM, Orchidee-CN and JULES)*"

We also explain more clearly how our study has relevance beyond SE Australia:

[line 495-496]: "*Whilst our analysis is site specific, the issues indicated here have been reported to lead to systematic biases in LSMs across multiple sites (Ukkola et al., 2016a; Trugman et al., 2018)*"

[line 503-505]: "*Since our study attempts to articulate the common issues in the simulated dry conditions in LSMs, we anticipate our findings would be applicable in many water-limited conditions, but equally, more mesic systems too.*"

At a few places in the discussion section the authors try to derive general conclusions, which could be beneficial also for modelling groups in other regions and with other models (e.g. implications for incorporated groundwater schemes, suitability of satellite-derived soil moisture estimates for model calibration), but this discussion should be more detailed. For instance, are there any processes to which special attention should be paid in LSM developments, or can you derive minimum requirements (e.g. spatial resolution) for external model data (e.g. soil texture), etc., or are such statements not possible for the chosen model setup? I recommend to address this in a separate sub-section in the discussion.

Thanks for the suggestion. We have opted to add detail to the existing sections to avoid repetition and now provide implications for other LSMs at the end of each sub-section and added more details to substantiate our recommendations:

[line 526-533]: "*However, a number of studies using alternative process-based schemes have been shown to improve individual model simulations (Haverd and Cuntz, 2010; Lehmann et al., 2018; Or and Lehmann, 2019). For example, Swenson and Lawrence (2014) introduced a dry surface layer-based soil evaporation resistance into CLM to depict water diffusion from dry soil, reducing biases in evapotranspiration and total water storage relative to FLUXNET-MTE and GRACE datasets. Based* *on a pore-scale model (Haghighi and Or, 2015), Decker et al. (2017) added the resistances of capillary-viscous and boundary layer to CABLE soil evaporation scheme and lowered the positive Es bias in springtime and improved seasonality of evapotranspiration. Hence, a focussed intercomparison of competing approaches against data originating from different ecosystems would be a valuable area of future work.*"

[line 540-550]: "*First, our results imply that LSMs that incorporate groundwater schemes need to be careful about aquifer initialisation because this strongly affects soil moisture dynamics. Second, there is no obvious solution to this initialisation and spin-up problem because drainage into the aquifer is a slow process, and it may take hundreds of years to reach a realistic equilibrium state. For global simulations, this suggests the need to a priori initialise the starting aquifer state and to assess against satellite-based products like GRACE (Döll et al., 2014; Niu et al., 2007) or implement off-line spin-up using meteorological forcing consistent with the subsequent simulations. However, while spin-up with observations is attractive, when the resulting states are incorporated into a coupled global model, inconsistencies are inevitable. Third, CABLE currently assumes an identical spin-up approach for the aquifer as the soil moisture, iterating until state changes between sequences of years are smaller than some threshold. LSMs that employ similar iteration approaches (Gilbert et al., 2017) are likely to encounter similar problems as CABLE because the rate of drainage into the aquifer is very slow, leading to negligible changes between iterations and thus satisfying the criteria for equilibrium.*"

[line 561-567]: "*The development in pedotransfer functions via machine learning or multi-model ensemble provides new avenues to reduce errors from parameters (Zhang and Schaap, 2017; Dai et al., 2019). High-resolution global soil datasets (e.g. SoilGrids, Hengl et al., 2017) covering multiple soil layers up to 2m depth offer opportunities to improve LSM simulations of soil moisture by incorporating depth-varying soil parameters. It is noteworthy that these global datasets of soil hydraulic parameters (Montzka et al, 2017; Zhang et al., 2019) have existed for several years but have not been widely used. Furthermore, at the EucFACE site, the observed soil texture information enabled the separation of parameter uncertainties from biases in process representations and model structural errors, a valuable step in better constraining LSM simulations.*"

[line 611-617]: "*Alternatives to the β functions have emerged to fill the theoretical gap, including plant hydraulic (Christoffersen et al., 2016; Xu et al., 2016) and stomatal optimality approaches (Sperry et al., 2017) but are yet to be widely adopted in LSMs (but see Eller et al., 2020; De Kauwe et al., 2020; Kennedy et al., 2019; Sabot et al., 2020). Replacing the empirical soil water stress factor by these plant physiology schemes reduces model arbitrariness associated with the representation of soil water stress and reduces the simulated biases in transpiration either over water deficit regions or areas with obvious dry seasons (Bonan et al., 2014; De Kauwe et al., 2020; Sabot et al., 2020). We can envision a wider application of these processes-based models will offer a chance to improve water stress representation in more LSMs.*"

2) please discuss the uncertainties in the observations in more detail. I suppose that especially for the "indirect" or "derived" observations of Etr und Es, uncertainties are quite large and thus affect the assessment of the model performance.

We have added further details about the uncertainties of soil evaporation and volumetric water content and now also discuss transpiration measurements:

[line 520-523]: *"Notably, soil evaporation was not directly measured at the site, but instead derived from the change in observed soil moisture over the top 5 cm, while ignoring days following rain (when the soil evaporative flux would likely be largest) (Gimeno et al., 2018a). As such, it contains soil moisture changes due to transpiration from a seasonal grass understorey but ignores evaporation below the top 5 cm, complicating model evaluation."*

[line 578-584]: *"On the other hand, the neutron probe measurements of soil moisture used for calibration also involve uncertainties (Gimeno et al. 2018a). The soil moisture estimates were derived by fitting two distinctive linear relationships between soil volumetric water content and raw neutron probe counts (see Figure S6) for clay (below 3m) and non-clay soil (above 3m). As a result, the observation error would be greatest in layers where the soil type differs from the assumed soil type at that depth. However, the fitted relationships were robust, since clay soils largely dominated the deeper profile (below 3 m depth) and sand soils mostly dominated shallow profile (above 3m depth)."*

[line 115-118]: "*Etr estimates are derived from tree sapflow velocities (3-4 trees per experimental ring) using the heat pulse compensation technique (Gimeno et al., 2018a). Sapflow velocity is translated to Etr by multiplying the sapwood area estimated from basal area inside each ring and a correlation between sapwood and basal areas based on 35 trees adjacent to the experimental rings.*"

[line 120-121]: "*To represent variability in Etr and Es across rings, we show the mean and the uncertainty within ring estimates in all figures."*

3) the control run exhibits an overestimated Es in conjunction with a soil moisture wet bias. Because of that, I was quite surprised about the first step to improve the model performance by increasing the resistance for soil evaporation Sres. Of course, such an increase in Sres results in a reduced Es, but must, at the same time, inevitably cause an intensified wet bias. Therefore, it would have been more intuitive to first increase the vertical drainage (as it is later done in the Watr experiment) to reduce the available water amount for evaporation in the upper soil. Is there any reason for the chosen sequence of experiments? I suppose that especially for the chosen "layering" approach, the order of the experiments is essential.

We agree with the reviewer that the order of experiments may influence our understanding of these parameterisations. However, our sequence was decided under the following considerations. We have stated our justification of the experiment order in ln 240-250 (also see our response to Reviewer 1): *"We choose to first resolve a soil evaporation bias as it affects ET partitioning; however, its importance is limited to the top soil layers and particularly to the period following rain. We then modified the initial water table depth as this fundamentally affects the root-zone soil moisture state. Next, we explored assumptions related to soil column discretisation and parameters, and further optimised key hydraulic parameters to improve overall soil moisture biases. We choose to explore parameter assumptions at this point in the experimental set up as the previous experiments aimed to resolve* *existing biases that affected the overall soil moisture availability. Finally, we explored alternative soil moisture stress functions as the last step because this factor integrates, and arises from, the soil moisture state. The experimental order allowed us to probe model biases in a systematic way but it is important to note that there is no perfect experimental order and alternative permutations would lead to subtly different interpretation of results. In fact, this is what commonly happens in model evaluation that explore a single factor (e.g. the soil moisture stress factor)."*

We agree with the reviewer that the order of experiments is important and may affect interpretation; however, we contend that there is no "perfect" experimental order and would argue there is merit to our chosen order. One could select an alternative order, but without any obvious reasoning, and the possible permutations are vast. We can assure the reviewer that we discussed the merits of these choices at length. We do note that we examine the impact of increasing the vertical drainage first, but it did not lead to differences that affected our conclusions.

4) In this study, the influence of non-hydrological factors on evapotranspiration (e.g. temperature, aerodynamic characteristics of the surface) is neglected. For instance, how good are the surface temperatures (soil and vegetation surface) simulated in CABLE? Are there any surface or soil temperature measurements which can be used for validation?

Thanks for the suggestion. We agree that non-hydrological factors are also significant for evapotranspiration but several of these factors were constrained in our experiments using observations. We have added details into the future direction text (see below).

We note that we did not evaluate our modelled canopy temperature against site measurements. We think that such a comparison would only be relevant had we used a vertically layered canopy model (e.g. https://bg.copernicus.org/articles/17/265/2020/), set up to accurately depict ring-to-ring tree variability, which would affect canopy-scale light interception/shading (and so evapotranspiration). Direct comparison of data to a big-leaf (two-leaf) model would have been unlikely to have added further constraints due to the importance of the position within the canopy of the measurements. Besides, we use half-hourly meteorology observations (e.g. radiation, air temperature and humidity) to force the model which ensures the simulated soil surface and canopy temperatures cannot diverge far from the observed state.

We have added:

[line 657-661]: "*While the focus of our study has been primarily on the parameterisations of hydrology and sub-surface processes, we did aim to minimise the uncertainties from the non-hydrological factors by using site characteristics, such as the aerodynamic conductance (determined by setting the canopy height) and photosynthesis parameters (e.g. maximum carboxylation rate and maximum rate of electron transport). However, due to the significance of these non-hydrological factors on evapotranspiration (e.g. Breil et al., 2020), further evaluation should be considered in future studies."*

References:

Decker, M., Or, D., Pitman, A. and Ukkola, A.: New turbulent resistance parameterization for soil evaporation based on a pore-scale model: Impact on surface fluxes in CABLE, J. Adv. Model. Earth Syst., 9(1), 220–238, doi:10.1002/2016MS000832, 2017.

Haverd, V., Cuntz, M., Nieradzik, L. P. and Harman, I. N.: Improved representations of coupled soil–

canopy processes in the CABLE land surface model (Subversion revision 3432), Geosci. Model Dev., 9(9), 3111–3122, doi:10.5194/gmd-9-3111-2016, 2016.

De Kauwe, M. G., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., Harper, A. B., Hickler, T., Jain, A. K., Luo, Y., Lu, X., Luus, K., Parton, W. J., Shu, S., Wang, Y. P., Werner, C., Xia, J., Pendall, E., Morgan, J. A., Ryan, E. M., Carrillo, Y., Dijkstra, F. A., Zelikova, T. J. and Norby, R. J.: Challenging terrestrial biosphere models with data from the long-term multifactor Prairie Heating and CO_2 Enrichment experiment, Glob. Chang. Biol., 23(9), 3623–3645, doi:10.1111/gcb.13643, 2017.

Powell, T. L., Galbraith, D. R., Christoffersen, B. O., Harper, A., Imbuzeiro, H. M. A., Rowland, L., Almeida, S., Brando, P. M., da Costa, A. C. L., Costa, M. H., Levine, N. M., Malhi, Y., Saleska, S. R., Sotta, E., Williams, M., Meir, P. and Moorcroft, P. R.: Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought, New Phytol., 200(2), 350–365, doi:10.1111/nph.12390, 2013.

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., Decker, M. and Haughton, N.: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts, Environ. Res. Lett., 11(10), 104012, doi:10.1088/1748-9326/11/10/104012, 2016a.

Ukkola, A. M., Pitman, A. J., Decker, M., De Kauwe, M. G., Abramowitz, G., Kala, J. and Wang, Y. P.: Modelling evapotranspiration during precipitation deficits: identifying critical processes in a land surface model, Hydrol. Earth Syst. Sci., 20(6), 2403–2419, doi:10.5194/hess-20-2403-2016, 2016b.