Response to Reviewer 2

In this study, Mu et al. evaluate the contribution of groundwater (GW) to vegetation water availability during heatwave and drought events in SE Australia. To do this, they implement a GW scheme in the CABLE land-surface model, and perform factorial simulations constrained by LAI to separate the contribution of GW to evapotranspiration and canopy temperature, which they compare with remote-sensing data.

The manuscript is concisely and clearly written, well structured and appropriately referenced and provides an important contribution to the land-surface modelling community. I find that two aspects could be improved:

> We thank the reviewer for the positive summary of our work and we address the reviewer's concerns below. Comments are shown in black, with our response below in blue in each case.

(i) After reading the title one would expect a greater focus on the impacts of GW on vegetation functional aspects (assimilation, stomatal conductance, transpiration, growth…), while the manuscript focuses mostly on hydrometeorology. Transpiration differences between the two simulations are only shown in Fig. 2, for 2019, but they could have been included in Fig. 3 and S6, to complement the discussion about the functional aspects. From the model simulations, one could additionally include, assimilation rates, stomatal conductance, NPP, etc.

> We thank the reviewer for these suggestions about other aspects of vegetation function not shown in the manuscript.
>
> We originally deployed this manuscript with the two considerations. First, as CABLE is run using prescribed LAI for these simulations, growth is not predictive (only gas exchange is predicted). This is a common approach used in a number of LSMs and is extremely helpful in these types of experimental setups as it isolates the change (i.e. directly to GW) without growth feedbacks. Second, our interest is on how GW sustains vegetation function via transpiration, which in turn may cool the boundary layer, when run in a coupled model environment. The carbon uptake plays no role in this link between the surface energy balance and the atmospheric feedback, hence we did not focus on it originally.
>
> Nevertheless, the reviewer's suggestions led us to reconsider these issues and we therefore will add spatial maps of GPP during the two droughts and the GPP difference between GW and FD. We will discuss the GPP result and explain why we do not focus on this in the rest of the manuscript.
>
> We thank the reviewer for the suggestion on transpiration plots. As requested, we will add transpiration plots to Figure 3 and a supplementary figure (see plots below).

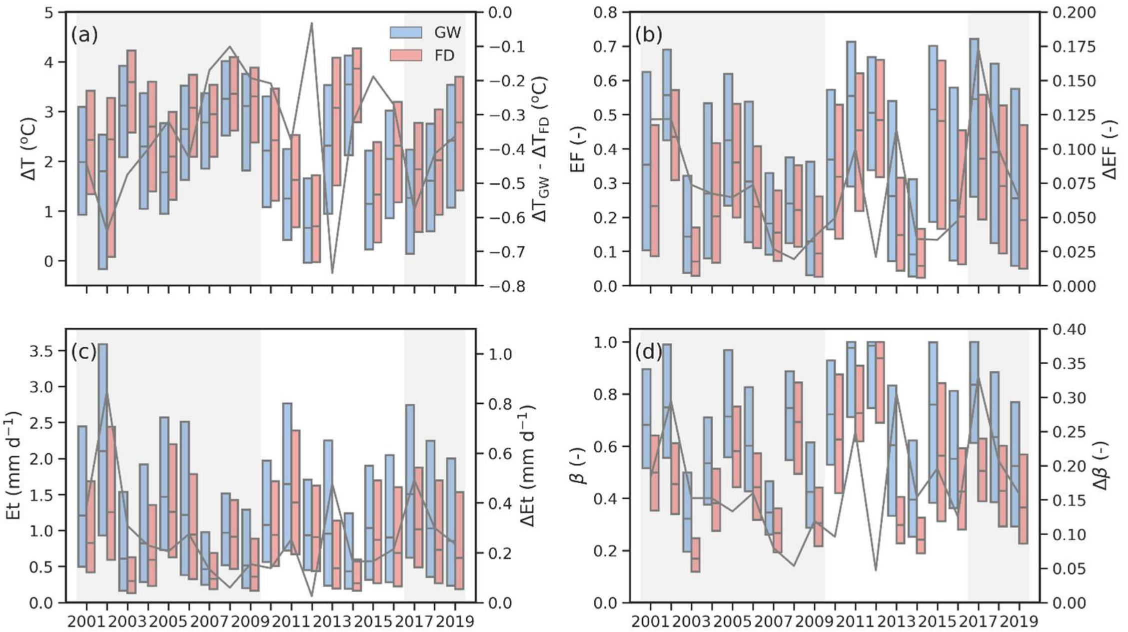
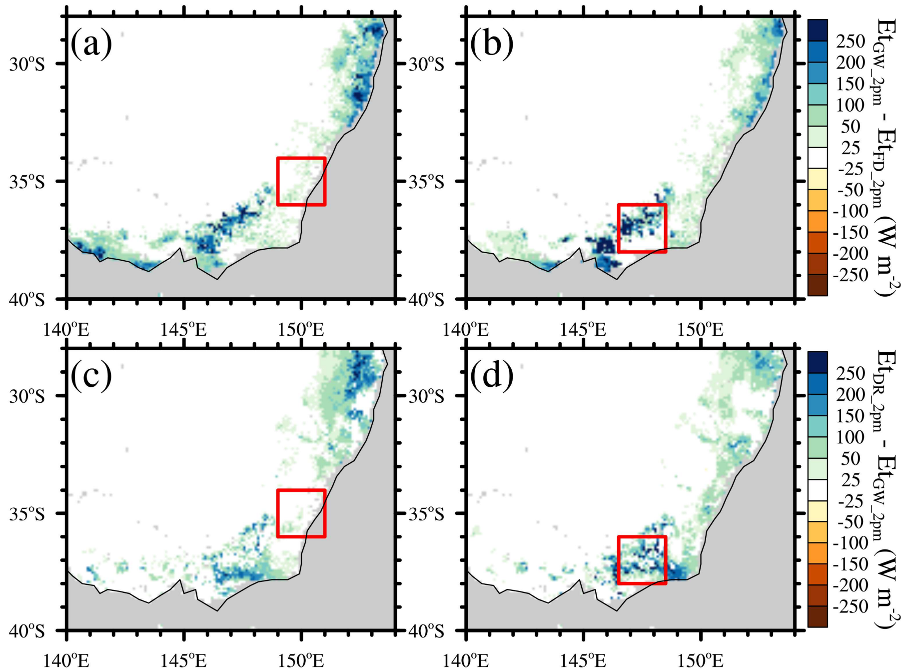


Figure 3. Groundwater-induced differences in (a) T_{canopy}-T_{air} (ΔT), (b) evaporative fraction (EF), (c) transpiration (Et), and (d) water stress factor (β) during 2000-2019 summer heatwaves over forested areas. The left y-axis is the scale for boxes. The blue boxes refer to the GW experiment and the red boxes to FD. For each box, the middle line is the median, the upper border is the 75th percentile, and the lower border is 25th percentile. The right y-axis is the scale for the grey lines which display the difference in the medians (GW-FD). The shadings highlight the two drought periods.



Supplementary Figure: The difference of transpiration (Et) at 2pm between (a)-(b) GW and FD (EtGW_2pm-EtFD_2pm), and between (c)-(d) DR and GW (EtDR_2pm-EtGW_2pm). The left column is for 15th and the right is for 25th Jan 2019.

We will also add appropriate text:

"However, the strength of the cooling effect decreases as the droughts extends and the transpiration difference (ΔEt, mm d^{-1}) diminishes quickly (Figure 3c) because the vegetation becomes increasingly water-stressed which consequently limits transpiration (Figure 3d)",

"While reductions of 5ºC are clearly limited in spatial extent, the overall pattern of cooling is quite widespread, where is coincident with the groundwater-induced Et increase (indicated by the new supplementary figure), implying an extensive reduction in heat stress along shallow WTD coastal regions during heatwaves",

and "By enabling access to moisture in the deeper soil, the LSM simulates further cooling by 0.5–5°C across the forests associated with an Et increase of 25–250 W m^{-2} (in the new supplementary figure)".

Moreover, the DR experiment is one of the most exciting aspects of the study since it highlights the relevance of the interactions between hydrology and physiology, but it is briefly discussed and shown only in Fig. 6. I find that a deeper analysis of the DR experiment and an additional figure on the impacts of the heatwaves for GW, FD and DR would increase the relevance of the study and better support the discussion around improvements to LSMs.

We are happy that the reviewer was interested in our DR simulations. We agree that this is a potentially important aspect of this work but note that the 2019 simulations were simply a proof-of-concept sensitivity experiment. Future work that allows roots to tap into the GW directly in LSMs, or that optimally set rooting depth by historical water availability, is a potential future avenue for research in LSMs (as noted in the discussion 4.3). However, as we lack the observations to meaningfully set root distributions across S.E. Australia, we do not plan to add additional DR plots (with the exception of a revision to Fig 6, see below). We will discuss this in the appropriate text:

"Given we lack the detailed observations to set root distributions across S.E. Australia, we undertake the DR experiment as a simple sensitivity study. We only run this experiment during January 2019, when the record-breaking heatwaves compound with the severe recent drought".

(ii) one of the key conclusions mentioned in the abstract and sections 4 and 5 is that GW helps sustaining higher transpiration rates in the first 1-2 years of multi-year droughts. However, most figures of the paper, and specifically the one showing transpiration differences, refer to the 2019 event. The figures showing differences between GW and FD for the full period do not separate specifically T from ET.

We thank the reviewer for this question. This paper aims to cover two aspects. First, it shows and explains the average behaviour of groundwater-induced transpiration on canopy cooling during heat extremes, in the context of two major droughts. Second, we focus on the short-term feedback (days) during the 2019 summer heatwaves occurred at the end year of the recent multi-year drought. These answer our main questions from the perspective of two time scales. Figures 1-5 relate directly to the average behaviour across the two major droughts. The rest of the Figures concentrate on the Jan 2019 heatwave events during the recent drought. However, we acknowledge that we did not clearly indicate how GW allows transpiration to be sustained during the 1-2 years of the drought events in the context of these figures in the results. We will add the text below to the manuscript:

"However, the strength of the cooling effect decreases as the droughts extends and the transpiration difference (ΔEt, mm d^{-1}) diminishes quickly (Figure 3c) because the vegetation becomes increasingly water-stressed which consequently limits transpiration (Figure 3d). For all variables (ΔT, EF, Et and β), the difference between GW and FD is greatest during the wetter periods (e.g. 2013) and in the first 1-2 years of the multi-year drought (2001-2002 for the Millennium Drought or 2017-2018 for the recent drought). After drought onset, the FD and GW simulations converge as depleting soil moisture reservoirs reduce the impact of groundwater on canopy cooling and evaporative fluxes."

To address the comment on separating transpiration from total ET, we will add Figure 3c and the extra supplementary figure (see plots above) to substantiate our statement that groundwater sustains extra transpiration cooling canopy temperature.

Moreover, even though slightly bigger differences between GW and FD are seen in 2002 (1 year following drought onset) and 2017 (drought start), strong differences are found also in non-drought years, e.g. 2013 (Fig. 3 and S6). I do not think that the results, as currently shown, can support strong conclusions about the duration of the effect of GW on HW effects.

We think that our results do support conclusions about the duration of the effect of GW on water fluxes and canopy cooling (i.e. predominantly in the early years of drought). To better emphasise this point we will highlight the drought periods with shading in both Figures 1 and 3. However, the reviewer correctly points out that larger differences are seen during both non-drought periods and drought onset. We will make this clearer in the relevant result section:

"FD underestimates the magnitude of monthly TWSA variance (standard deviation, SD = 37.18 mm) compared to GRACE (47.74 mm) or GW (47.67 mm), particularly during the wetter periods (2000, 2011-2016) and the first ~2 years of the droughts (2001-2, 2017-8) (Figure 1a). For all variables (ΔT, EF, Et and β), the difference between GW and FD is greatest during the wetter periods (e.g. 2013) and in the first 1-2 years of the multi-year drought (2001-2002 for the Millennium Drought or 2017-2018 for the recent drought)."

However, Figure 1a and Figure 3 (grey lines) show that as the droughts progress, the two experiments tend to converge in their canopy cooling and water fluxes and the impact of GW diminishes as soil water becomes increasingly limiting, supporting our conclusion of a larger GW impact during drought onset. We will make this final point about the difference in canopy temperature in the manuscript to further emphasise the point:

"As the drought lengthens in time, the depletion of moisture gradually reduces this effect, from an average reduction of 0.52°C of the first 3 years to 0.16°C of the last 3 years in Millennium Drought (Figure 3a)",

"However, the strength of the cooling effect decreases as the droughts extends and the transpiration difference (ΔEt, mm d^{-1}) diminishes quickly (Figure 3c) because the vegetation becomes increasingly water-stressed which consequently limits transpiration (Figure 3d). For all variables (ΔT, EF, Et and β), the difference between GW and FD is greatest during the wetter periods (e.g. 2013) and in the first 1-2 years of the multi-year drought (2001-2002 for the Millennium Drought or 2017-2018 for the recent drought). After drought onset, the FD and GW simulations converge as depleting soil moisture reservoirs reduce the impact of groundwater on canopy cooling and evaporative fluxes",

and "We found that the influence of groundwater was the most important during the wetter periods and the first ~ two years of a multi-year drought (~2001-2002 and 2017-2018; Figure 1 and 3)".

Other comments:

L100: I think dimensional analysis gives the units of Fsoil as m3.m-3/s, or 1/s (to match the other two terms), can you confirm?

Thank you for spotting this, we will correct the units to 1/s.

L235: "much closer": indeed, but still very far.

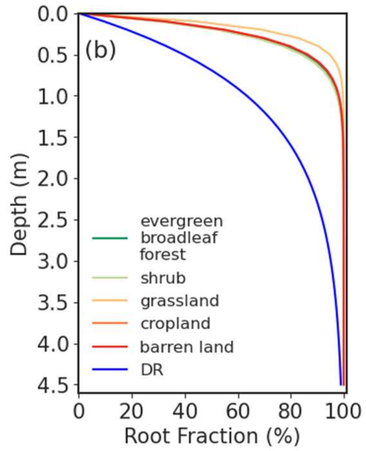
Agreed, we will alter the text as,

"GW increases the evaporation relative to FD such that the accumulated P-E decreases from about 786 mm to 455 mm during the Millennium drought, which is within the range of DOLCE (460 mm) and GLEAM (97 mm) estimates",

and "Although the total land evaporation products display some differences, the GW simulations are closer overall to the DOLCE and GLEAM estimates".

L250: can be complemented by a map of root length in CABLE.

Thank you for this suggestion. We will add the below plot as the panel b to Figure S3 to show the root distribution among the simulated PFTs.



The root fraction (%) above a given depth (m)

.

L312-317: how much is this threshold dependent on model structure and parameterization? And how does it compare with the same results for the DR experiment?

The threshold of ~6 m does likely arise from the model assumption of a 4.6 m soil bucket which also sets the maximum rooting depth (roots are confined to the soil layers and do not extend to the GW aquifer in CABLE). The 4.6m depth comes from observational evidence of most roots being situated within the top 4.6m (Canadell et al. 1996). When the water table is below this depth, the water fluxes largely become uncoupled between the soil column and groundwater, leading to a diminished impact of GW below the ~6m threshold. The DR experiment uses the same soil depth assumption (it only differs from the GW simulation in having a larger fraction of roots situated in deep soil layers) and as such would display a similar threshold. As we only have DR outputs for January 2019, we did not explicitly quantify this.

However, clearly the threshold is CABLE-specific and we will acknowledge this in the manuscript:

"However, the absolute value of the threshold is likely CABLE-specific and associated with CABLE's assumption of a 4.6 m soil depth, which also sets the maximum rooting depth (roots can only extend to the bottom of the soil bucket and cannot access to the groundwater aquifer in CABLE). The CABLE soil depth comes from observational evidence of most roots being situated within the top 4.6 m (Canadell et al. 1996). Since the model assumes no roots in the groundwater aquifer, when the water table is below this depth, the water fluxes become largely uncoupled between the soil column and the groundwater aquifer, leading to a negligible impact of GW below ~6 m depth."

L325-331: very hard to compare panels a-b with c-d. Can you use a consistent mask?

If we used the same mask we would lose a lot of information from the model maps, which we feel is worth keeping. Instead, we have provided the masked plots in Figure S6 a-d and g-h, which compares the model simulated and MODIS-based ΔT.

L339-341: the label of Fig. 6e,f indicates “GW-FD”. Can you check? I would find it important to show DR in more figures, as discussed above.

Thank you for the suggestion. We originally only provided “GW-FD” in Figure 6 but agree with the reviewer that the DR simulation is also of interest. As such we will show the difference “DR-GW” in Figure 6 g-h. The original panels g-h will be moved to a separate Figure 7.

L343: how can one compare panels a,b with g,h in Figure 6?

The panels g-h show the diurnal evolution of ΔT for the regions shown by the red boxes in panels a-f. As the maps only show behaviour at 2pm (when the afternoon MODIS overpass occurs), the line plots were created to show typical diurnal cycles. They allow the comparison of diurnal cycles across the three experiments (GW, FD and DR) in the context of the two available day-time MODIS overpasses.

L375-376: can you support this by separating results per WTD bins rather than simple visual inspection?

Thank you for this suggestion. We will add the metrics on the left bottom corner in each panel of Figure 4.

L376-377: Where can we see the time-dependence of this response?

We thank the reviewer to spot this point. We will clarify in the relative text:

“However, the strength of the cooling effect decreases as the droughts extends and the transpiration difference (ΔEt, mm d^{-1}) diminishes quickly (Figure 3c) because the vegetation becomes increasingly water-stressed which consequently limits transpiration (Figure 3d). For all variables (ΔT, EF, Et and β), the difference between GW and FD is greatest during the wetter periods (e.g. 2013) and in the first 1-2 years of the multi-year drought (2001-2002 for the Millennium Drought or 2017-2018 for the recent drought). After drought onset, the FD and GW simulations converge as depleting soil moisture reservoirs reduce the impact of groundwater on canopy cooling and evaporative fluxes”,

and “Importantly, the role played by groundwater diminishes as the drought lengthens beyond two years (Figure 3)”.

L408: specify what feedback is meant here

We mean the feedback from changes in the land surface fluxes on the boundary layer. We will change the text as "The lack of groundwater in many LSMs suggests a lack of this moderating process and consequently a risk of overestimating the positive feedback on the boundary layer in coupled climate simulations".

L448-449: rather than making a general statement, the authors could analyze variables related with the physiological responses (assimilation, stomatal conductance, WUE, NPP) to show that (if) GW matters.

As we explained above, we will also show GPP in the supplementary but concentrate on transpiration as it is the key variable for canopy temperatures and potential boundary layer feedbacks during heatwaves, which are the key research questions in this paper.

L486: what can we see a figure supporting this conclusion?

As we explained above, it comes from Figure 1 and Figure 3. We will highlight this finding in the results as per previous responses.

L487: the cooling effect is shown only for 2019, correct?

We saw a cooling effect during all 2001-2019 heatwave events, as well as the January 2019 heatwave. To make the point clearer, we will adjust the sentence as

"This cooled the forest canopy on average by 0.03–0.76 ^{o}C in heatwaves during 2001-2019 and as much as 5 ^{o}C in regions of shallow water table depths in the heatwave in January 2019, helping to moderate the heat stress on vegetation during heatwaves".

L490: this is not strongly supported by results (comments above)

Thanks. As per our replies to the previous comments, we will better explain this finding in the result section.

Reference:

Canadell, J., Jackson, R. B., Ehleringer, J. R., Mooney, H. A., Sala, O. E., and Schulze, E. D.: Maximum rooting depth of vegetation types at the global scale, Oecologia, 108(4), 583–595, 1996.