In this manuscript the authors analyse the effects of the Millennium Drought with the aim of identifying possible process explanations for the observed persistent changes in the hydrological response in many catchments in Australia.

The experiment was logically designed and very systematically implemented. In a concerted and probably unique effort, the authors formulated an exhaustive suite of potential processes hypotheses. These hypotheses have then been rigorously confronted and scrutinized with observations. The extremely well documented analysis together with the very robust data support and the detailed and critical interpretation thereof make this manuscript an excellent example of good and relevant science. I commend the authors for this effort.

The manuscript is also very clearly structured and well-written. Overall, I do only have two major observations/comments the authors may want to consider:

(1) Although I highly appreciate the detailed explanations of the individual process hypotheses, some of the hypotheses could strongly benefit from a more precise terminology and/or clearer description. This will make it easier for the reader to appreciate and understand the actual differences between different hypotheses (see below in the list of detailed comments).

(2) Some of the hypotheses could benefit from a stronger and wider connection to literature, in particular outside Australia by providing more references to related studies (please note: below I have added a few suggestions. However, these include for my convenience and to save time, quite some work from our group. Please do *not* feel obliged to cite these papers - other research groups may have published material that fits better).

<u>Detailed comments:</u>

P.2, l.50: not sure that "recently" is the most suitable term here. Literature dedicated to the topic has been around for a while. For example, Destouni et al. (2013), Jaramillo and Destouni (2014) or van der Velde et al. (2014) were published almost a decade ago.

P.2, l.52: not sure China qualifies as a "continent". Perhaps worth to also include the recent analysis by Roodari et al. (2021) in Central Asia here.

P.3, l.71-75: there are quite some ongoing initiatives to address this issues and to find work-arounds. Some recent examples include Speich et al. (2020) and Bouaziz et al. (2022).

P.3, l.81: Roodari et al. (2021) in Central Asia

P.5, l.149-151, Figure 1: Excellent approach and description!

P.6, 171-172: Please define how “shift” was defined here. How much of a change is necessary to be considered as “shift”?

P.6, l.177: “persisting within a low runoff state” sounds awkward, given that runoff is a *change of a state* (i.e. dS/dt!). I know what you intend to say but please try to rephrase.

P.6, l.179-180: storages (or states – interception, groundwater storage) are lumped with changes of storages (dS/dt, i.e fluxes – precipitation, ET, recharge). Please avoid that as they have fundamentally different functions. In addition, it is not clear what “ET” stands for here, potential evaporation or actual evaporation? I cannot fully follow the reasoning: increases in actual evaporation are in many cases the direct consequence of increased interception (e.g. on canopies). Thus, they come hand in hand. Please clarify.

P.7, Figure 2: the reader can only assume that in panel (a) the bar chart indicates precipitation and the cross/lines stream flow. Please explicitly describe that in the caption. For panels (c) and (d) please avoid using red and green shades in the same figure: ~15% of your readers will be red-green colour blind.

P.8, l.195: is this so? Then these studies may be based on a somewhat incomplete definition of “drought”. In any case, please provide supporting reference for this statement or remove it.

P.8, l.200-202: this is a repetition of P.5, L.135-139 and Figure 1. Can be omitted.

P.8, l.218ff: although the explanation of the “two water world” (TWW) hypothesis is correct here, it does not at all support your argument here. TWW describes actual water ages and the related transit and residence time distributions that are largely controlled by the physical transport *velocities* of individual water molecules. Here, seasonal or annual water budgets are considered. These are instead controlled by *response times* which are regulated by the propagation of pressure waves at given *celerities* (see McDonnell and Beven, 2014; and Figure 2a in Hrachowitz et al., 2016). The reference to the TWW is therefore unsuitable and actually incorrect here. Instead what you describe here (“…water can move to the stream only after soil pores have been replenished”) is the functioning and role of water storage following the concepts of Field Capacity and Permanent Wilting Point. Please remove any reference to TWW here.

P.11, l.240 (also Table 1): not sure how HPE02 is different to HPE02. Both are effectively the result of higher water deficits (i.e. lower water content below Field capacity) due to more pronounced dry periods.

In both cases, the water deficit is not overcome – water content in soils remains below field capacity and is therefore held against gravity instead of being released directly (or via groundwater) to the stream. Please clarify the difference between HPE01 and HPE02. In addition, please specify more precisely what is meant by "initial losses". Do you mean water stored in soils and eventually released as transpiration or (to a minor degree) soil evaporation? Where else could this water be lost to?

P.12, l.270: "activation of long-term…processes" is very vague. Please try to be more specific. How does the reader have to imagine that? Is it really a binary phenomenon (active vs. deactivated?) or is it a process that gradually and proportionally becomes more relevant and visible the drier the system becomes?

P.12, l.279: please avoid absolute terms such as "veracity". The best we can do in large-scale hydrology is to test and evaluate our hypothesis.

P.12, l.293ff: perhaps good to refer to Jaramillo et al. (2018), who provide a good illustration of the counteracting effects of fertilization.

P.13, l.304: it is not quite clear to me (1) why HPE07 falls under "vegetation conditions" and not under "meteorological dynamics", (2) what the difference between HPE07 and HPE04 is, as both radiation and temperature are major controls on evaporative demand (HPE04, i.e. potential evaporation) and (3) why in the one sentence description the focus of HPE07 is radiation and temperature, while in the text just above, the turbulence differences are the actual core of this HPE. Please clarify.

P.13, l.322: agreed, but they will not only intercept water on their foliage and thus allow evaporation, but they will also continue to transpire water. Please try to be more specific here.

P.13, l.324-324: perhaps explicitly refer to C4 grass here.

P.14, l.331: why only in south-east Australia? I would suspect this to be a general phenomenon.

P.14, l.336: not sure what is meant by "assuming some other mechanism for lower stream flow,…". Please clarify and rephrase.

P.14, l.344: see above

P.15, l.362: see above

P.15, l.365ff: Not clear what the actual processes involved here are meant to be. In my understanding the idea is that upwelling groundwater, persistent over several decades dissolved salts from deeper parts of the soil and moved it to near-surface layers. If this is so, I do not understand how lowering the groundwater tables would reduce salinity over a relatively short period: solute movement in the subsurface, in particular during dry conditions (i.e. droughts!) is characterized by very slow transport velocities and long-term legacy effects (e.g. Basu et al., 2010; Hrachowitz et al., 2015). Instead, solutes moved into near-surface layers during wet conditions will frequently undergo evapo-concentration effects (e.g. Hrachowitz et al., 2015), thus temporarily making conditions for plant growth even more unfavourable.

P.15, l.376: the use of the term “interception” here and elsewhere in the manuscript is ambiguous. Please note that “interception” has a very specific meaning in hydrological literature (e.g. Miralles et al., 2020; Savenije, 2004). What is specifically meant here? A process that retains water on the canopy, foliage or near-surface soil layers to supply water for the “evaporation” process or is it used in a more general way to also include “interception” in in the root zone to supply root-water uptake for transpiration? If it is the latter, I strongly recommend to rephrase to avoid misunderstandings.

P.15, 377ff: HPE15 is described only in very broad and vague terms. I agree, that the actual processes here may be unknown. However, in such a case, I am not sure if a meaningful hypothesis can be formulated, because a meaningful hypothesis always needs to be testable, otherwise it cannot be qualified as hypothesis. Please try to more explicitly specify this hypothesis or remove it, as it may be indistinguishable from most of the other vegetation-related HPEs here.

P.16, l.392: not sure if “activated” is the most suitable term here (see one of the comments above)

P.16, l.403: what is “diffuse discharge”?

P.16, l.411: please specify “discharge areas”

P.16, l.414: “interception by transpiration”? Please see comment above.

P.18, l.435: Hulsman et al. (2021a) similarly found supporting evidence for the importance of upland groundwater sustaining alluvial evaporation/transpiration in a large scale study in the Zambezi basin. Perhaps nice to include as reference.

P.18, l.444-446: explicitly mention the role of evaporation/transpiration here to be more specific.

P.18, l.449: potentially very relevant and ofter overlooked process. Bouaziz et al. (2018), Condon et al. (2020), Hulsman et al. (2021b) but also the authors themselves (Fowler et al., 2020) provide different recent perspectives on the potential importance of this process. Frisbee et al. (2012) also provide an excellent synthesis and illustration (Figure 3 therein!). Would be good to add at least some of these references here to provide a stronger context and background for the reader.

P.19, l.461: "[…] some systems […] that are more extensive […] have longer response times […]". Really? I am not sure this reasoning is generally valid. In larger systems the average flow distance in the subsurface domain, the controlling factor on response times at time scales > one month, is in most environments very similar to those of smaller, headwater systems. In other words, no matter if you are anywhere upstream or downstream in the landscape, the distance to the next river will not be that different, due to the fractal, scale-invariant nature of river networks (e.g. Rodriguez-Iturbe and Rinaldo, 2001). Please provide a reference or remove.

P.19, l.485ff: this HPE needs more explanation. It is not clear why the vadose zone should indeed be drier. The reasoning here, as far as I understand, is that declining GW tables result in deeper vadose zones. Ok. In these parts of the subsurface, all the water that cannot be held against gravity (water above Field Capacity) will released and "follow" the falling GW table. The remainder, i.e. soil water content at Field Capacity, will largely be held against gravity. This water can only be released by evaporation or plant water uptake for transpiration. Assuming that in many locations the GW-table is below the root zone, plant water uptake drops out as potential process to remove water. However, soil water at depth below 20-30cm can also not be evaporated at very high rates (e.g. Brutsaert, 2014). The deeper, the less relevant soil evaporation will be due to the limited diffusive gas/vapour exchange with the surface (there is no wind in the soil pores for turbulent exchange!). Soil deeper below the root zone will thus frequently be close to Field Capacity, as the water cannot be released with gravity only and very limited evaporation. It would be great if you could provide a more detailed description of your hypothesis that soils in a deeper vadose zone can be drier (i.e. below Field Capacity – because if they remain at or above field capacity, they will be hydraulically and hydrologically irrelevant. In that case, all the water that enters this zone from above will be again released as it cannot be held, i.e. $dS/dt \sim 0$ over time scales larger than a few days).

P.19, l.487-488: I suspect you mean that the "infiltration capacity" declines with increasing wetness as described by Darcy-Richards. In contrast, "hydraulic conductivity" typically increases with increasing wetness!

P.20, l.495: that cracked soils result in less runoff is of course not impossible. However, also the opposite, the importance of cracks as preferential flow pathways, is frequently observed and documented (e.g. Zehe at al., 2013). Please adjust the hypothesis accordingly.

P.20, l.497-498: “Streamflow was lower […] because […] higher infiltration […]”. This does not quite add up for me. Water that infiltrates surely does not disappear. Was the assumption that most of it will be held in the soil and evaporated/transpired instead of recharging the GW?

P.21, l.529: perhaps better to replace “interception” by something like “retention and subsequent evaporation/transpiration”

P.23, l.584: should this read as “10^5”?

P.23, l.600: please also explicitly mention the three plausible hypotheses here and not only in the table.

P.24, HPE08: “[…] it is doubtful whether modest historical CO_2 increases could have caused larger changes […]”. Without any further data support, more detailed reasoning and/or references this remains largely speculation and cannot be used as hypothesis test.

P.25, HPE13: in the light of my comments above, the reasoning here (low water table) is not very convincing.

P.26, HPE19: see above. A deeper vadose zone will only allow further water retention if water from pores is being extracted by soil evaporation and or transpiration. Otherwise the zone will, on average, remain close to Field Capacity and act as a hydraulically and hydrologically passive part of the soil. In other words, it will cause some delay in the water percolating through this zone, but it will not provide additional “storage”, i.e. on time scales of more than a few days dS/dt~0.

P.33, 684ff: see above. Also, the term “interception” is not suitable here

<u>References:</u>

Basu, N. B., Destouni, G., Jawitz, J. W., Thompson, S. E., Loukinova, N. V., Darracq, A., ... & Rao, P. S. C. (2010). Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. Geophysical Research Letters, 37(23).

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M. (2018). Redressing the balance: quantifying net intercatchment groundwater flows. Hydrology and Earth System Sciences, 22(12), 6415-6434.

Bouaziz, L. J., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., ... & Hrachowitz, M. (2022). Ecosystem adaptation to climate change: the sensitivity of hydrological predictions to time-dynamic model parameters. Hydrology and Earth System Sciences, 26(5), 1295-1318.

Brutsaert, W. (2014). Daily evaporation from drying soil: Universal parameterization with similarity. Water Resources Research, 50(4), 3206-3215.

Condon, L. E., Markovich, K. H., Kelleher, C. A., McDonnell, J. J., Ferguson, G., & McIntosh, J. C. (2020). Where is the bottom of a watershed?. Water Resources Research, 56(3), e2019WR026010.

Destouni, G., Jaramillo, F., & Prieto, C. (2013). Hydroclimatic shifts driven by human water use for food and energy production. Nature climate change, 3(3), 213-217.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., ... & Western, A. (2020). Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections. Water Resources Research, 56(5), e2019WR025286.

Frisbee, M. D., Phillips, F. M., Weissmann, G. S., Brooks, P. D., Wilson, J. L., Campbell, A. R., & Liu, F. (2012). Unraveling the mysteries of the large watershed black box: Implications for the streamflow response to climate and landscape perturbations. Geophysical Research Letters, 39(1).

Hrachowitz, M., Fovet, O., Ruiz, L., & Savenije, H. H. (2015). Transit time distributions, legacy contamination and variability in biogeochemical 1/fα scaling: how are hydrological response dynamics linked to water quality at the catchment scale?. Hydrological Processes, 29(25), 5241-5256.

Hrachowitz, M., Benettin, P., Van Breukelen, B. M., Fovet, O., Howden, N. J., Ruiz, L., ... & Wade, A. J. (2016). Transit times—The link between hydrology and water quality at the catchment scale. Wiley Interdisciplinary Reviews: Water, 3(5), 629-657.

Hulsman, P., Savenije, H. H., & Hrachowitz, M. (2021a). Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement. Hydrology and Earth System Sciences, 25(2), 957-982.

Hulsman, P., Hrachowitz, M., & Savenije, H. H. (2021b). Improving the representation of long-term storage variations with conceptual hydrological models in data-scarce regions. Water Resources Research, 57(4), e2020WR028837.

Jaramillo, F., & Destouni, G. (2014). Developing water change spectra and distinguishing change drivers worldwide. Geophysical Research Letters, 41(23), 8377-8386.

Jaramillo, F., Cory, N., Arheimer, B., Laudon, H., Van Der Velde, Y., Hasper, T. B., ... & Uddling, J. (2018). Dominant effect of increasing forest biomass on evapotranspiration: interpretations of movement in Budyko space. Hydrology and Earth System Sciences, 22(1), 567-580.

McDonnell, J. J., & Beven, K. (2014). Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. Water Resources Research, 50(6), 5342-5350.

Miralles, D. G., Brutsaert, W., Dolman, A. J., & Gash, J. H. (2020). On the use of the term "evapotranspiration". Water Resources Research, 56(11), e2020WR028055.

Roodari, A., Hrachowitz, M., Hassanpour, F., & Yaghoobzadeh, M. (2021). Signatures of human intervention–or not? Downstream intensification of hydrological drought along a large Central Asian river: the individual roles of climate variability and land use change. Hydrology and Earth System Sciences, 25(4), 1943-1967.

Savenije, H. H. (2004). The importance of interception and why we should delete the term evapotranspiration from our vocabulary. Hydrological processes, 18(8), 1507-1511.

Speich, M. J., Zappa, M., Scherstjanoi, M., & Lischke, H. (2020). FORests and HYdrology under Climate Change in Switzerland v1. 0: a spatially distributed model combining hydrology and forest dynamics. Geoscientific Model Development, 13(2), 537-564.

Van der Velde, Y., Vercauteren, N., Jaramillo, F., Dekker, S. C., Destouni, G., & Lyon, S. W. (2014). Exploring hydroclimatic change disparity via the Budyko framework. Hydrological Processes, 28(13), 4110-4118.