We thank the reviewer for their helpful and encouraging comments, and we address their various concerns below. Referee comments are in black, with our response below highlighted in red in each case.

The manuscript “Examining the Role of Environmental Memory in the Predictability of Carbon and Water Fluxes Across Australian Ecosystems” by Cranko Page and coauthors deals with the crucial topic of ecological memory and how this affects biosphere-atmosphere net CO2 fluxes and latent heat.

The manuscript builds on Liu et al., 2018, but it’s incremental because of the use of different models (i.e., k-mean clustering) and the use of terrestrial biosphere models.

The manuscript is well written. The analysis is well done and robust. The results are fascinating and well discussed. I think this will article provide an exciting contribution, and it will be of great interest for the Biogeoscience journal readership. It's an excellent manuscript.

I found that only one aspect should be improved: a deeper discussion on how the community can modify models to describe the ecological memory better. I found this aspect a bit weak and can be improved.

For the rest, I believe that the manuscript is well done, enjoyable, and scientifically very robust.

Thank you for your positive summary of our manuscript.

I have a few comments I suggest the authors should address:

Line 4: I suggest including what is meant for structural lags: “…structural lags (i.e., …)”

We agree with this comment, especially as we previously provided examples of extreme events earlier in the sentence. As such, we have amended this to read “structural lags (such as delays between rainfall and peak plant water content, or between a precipitation deficit and down-regulation of productivity)”.

Line 5: I suggest substituting models with “terrestrial biosphere models” to be specific

Agreed. This has been added.

Line 10: substitute latent hear with lambda

This has been changed as suggested.

Line 45: In my opinion, this statement critically depends on the timing of the precipitation with respect to the phenological stage. Please consider rephrasing to account for this comment

We agree with the Reviewer, particularly for deciduous grasses and trees; however, for evergreen species, the phenological stage has less relevance. Our original text was deliberately general and so we have opted not to change it.

Line 51: I agree on the vegetation type, but it would be better to be more specific on which aspects and differences between vegetation types can be critical confounding factors (different allocation strategies, height, etc). This would set the ground for the discussion.

We agree that "vegetation type" was vague and have changed this sentence to highlight a few of the specific reasons for vegetation type affecting the timescales of response:

"Confounding factors, such as differing vegetation characteristics (including the proportion of woody vegetation, rooting depth and varying allocation strategies), prevailing climate, interacting processes, and prior extreme events, all influence the magnitude and timescale of these lags and their impact on ecosystem fluxes."

Lines 63: "models fail to capture the impact of water stored in reservoirs with longer response times to climate". I think the authors should clarify why models do not describe well the climate impact: missing water table depth, poor description of soil layers and the root profile, etc.

We have added a sentence here to clarify some of the reasons for the failure of models to capture this impact:

"Such model failures may in part be due to incorrect rooting depths, poor soil profile characterisation, or a lack of representation of these long-term storage pools such as groundwater or wetlands."

Table 1: I suggest adding the vegetation type

Thank you for this suggestion. We have added the ecoregion classification for each site to the table. This provides further support to our grouping of sites into two different vegetation groups. The caption for Table 1 has been updated with the following sentence to address the new column:

"World Ecoregion is the biome classification of each site which is based on climatic regime and ecological structure, among other criteria (Olson et al., 2001; Beringer et al., 2016)."

Line 103: If I am not wrong latent heat and net ecosystem exchange are already defined

This is true, and there were further instances where these acronyms were either defined again or used prior to being defined. As such, the below changes have been made:

Line 42: "NEE" has been replaced with "net ecosystem exchange (NEE)" as the first definition

Line 77: "Net ecosystem exchange (NEE)" has been changed to "NEE"

Line 103: Full names replaced with acronyms

Line 105: Specify already here the time scale used (i.e., daily data). This information is reported a few paragraphs later. I think it will help the clarity. Also, the assumption of a certain degree of independence between NEE and latent heat would not be valid at hourly data during the daytime, when photosynthesis dominates the signal of NEE, so better to clarify that you are talking about daily data.

Line 103 has become "In this analysis we used observations of daily NEE and λE fluxes"

Line 115: I see the importance of using the long-term and consistent NDVI data. The authors use NDVI as a phenological and structural proxy. I suggest at least discussing the use of radar data rather than only optical data. For instance, I invite the authors to test Sentinel-1 data rather than MODIS NDVI for a few selected sites or give a perspective beyond the NDVI.

NDVI was used as it represents a widely used proxy for phenology. In our study, the NDVI data allows us to identify the site's sensitivity to each climate driver based on site greenness.

The reviewer makes an interesting suggesting to use Sentinel-1 data; however, as this was launched in 2013, it does not cover our full site record. Moreover, it is worth noting that satellite-based products that work well in other ecosystems/environments often have weaker skill in Australian environments (e.g. Leuning et al., 2005). By contrast, the NDVI product is well tested and understood (e.g. Rifai et al., 2021), which helps statistical interpretation (confidence) of our model results.

We have plans to extend our approach and factoring in multiple remote sensing datasets, so we may revisit the suggested Sentinel data as part of this work, as assessing new satellite data can be an entire piece of research.

Would you please specify which temporal resolution (8-day composite?)? I think it might be a piece of important information for the study of lag effects.

The NDVI data used is daily values calculated from a 16-day period of observations. The full algorithm for the calculation of this composite is available at the cited source. On line 117 we added the following sentence to clarify that we are taking the daily data available from the cited 16-day composite dataset:

"This dataset is calculated at a daily timestep based on a 16-day composite of observations."

The methodology is sounding. I suggest calculating linear relaxed precipitation instead of using the 15-days rainfall average. Linear relaxed precipitation can be calculated with a backward moving weighted average with 15 days width and weights that linearly decrease along with the window. The result is a pseudo soil moisture time series that might serve the purpose of the analysis better than the 15-days rainfall average.

We thank the reviewer for this suggestion – it's indeed a good idea. However, such a change is not necessary, since a mechanism to achieve this temporal weighting already exists within the Stochastic Antecedent Modelling framework. The weights in our 14-day antecedent sums are inferred from the data. If a linear decrease in weights within this window is suggested by the data, then this would be seen in the results. This is indeed the case for some sites – see for instance the short-term precipitation timescale at AU-Cpr in Figure S5 where the logarithmic-like shape of the plot indicates decreasing weight with increasing lag. We also note that some sites do not follow the same timescales, which implies that the response is not universally linear and therefore may not accurately be captured by the reviewer's proposed method.

Finally, the authors assume a Laplace distribution of the error at the daily time scale. There are contrasting reports in the literature, and some suggest that even at half-hourly time-scale, the Laplace distribution is the result of the the superimposition of two gaussian distributions of daytime and nighttime error (e.g. Lasslop et al., 2008), and at daily time scale the error is likely to be gaussian. I don't think this would impact the results, but I think it should be clarified to report all the positions in the literature.

This is a very pertinent point, although we do note that the Laplace distribution still performs well for these fluxes as per Lasslop et al. (2008). We have clarified that differing results are present in the literature by amending line 130 to the below:

"While the exact distribution of flux errors is site-dependent and can vary between super-imposed Gaussian distributions (Lasslop et al., 2008) or Student's t-distribution (Weber et al., 2018), daily NEE is assumed to be Laplace-distributed (Richardson et al., 2006) with mean μNEE and variance σ^2 in line with Liu et al. (2019)"

Figure 2 and results section on the sensitivity: The sensitivities should have units if I understood the method. Would you please add them everywhere?

Thank you for this correction. The relevant units have been added where required (including Figure S3 in the supplement).

Line 195: Something missing at the end of the sentence?

We believe the sentences in this paragraph to be complete.

Line 303: "No clear relationship between lagged responses and the prevailing vegetation at the sites.."

It would be interesting to check that relationship with quantitative vegetation characteristics reported for many OzFlux sites rather than a general statement on the prevailing vegetation. Did the author verify if the average canopy height, C4 fraction, fractional tree cover, rooting depth if known, etc., are possible controlling factors of the lagged response?

Our statement was unclear and has been amended: "No clear relationship between lagged response timescales and the prevalence of woody vegetation at the sites…".

The reviewer notes a range of potential controlling factors, but data on these is not easily obtained (e.g., C4 fraction, which is typically model-derived; rooting depth, which is rarely well resolved across sites). One of our hopes is that our findings provide fresh inquiries into trying to understand site behaviours, which will include collection of information like that suggested by the reviewer.

Line 350-351: I agree, but I think other potential causes are: 1) lags between respiration and photosynthesis due to transport or relocation of CO2 and assimilated (Mencuccini, M. and Hölttä 2010), as well as 2) potential rapid dynamic modulation of allocation due to phenology or response to stress.

We accept these suggestions for other potential causes and have amended line 351 to include the following:

“For instance, this dependence on prior day flux may be affected by seasonal leaf area, delays between photosynthesis and respiration (Mencuccini and Hölttä, 2010), or potential site disturbances. It could also be representing environmental drivers that we have not explicitly accounted for, lagged allocation, or more unique impacts such as insect infestation.”

Line 360: please check the reference

This has been corrected.

Line 416: typo in 1:94 kPa

This now reads “1.94 kPa” as intended.

Figure 3) and 4). There are double labels for some panels. I find it confusing. I suggest to change it.

We appreciate the concern for the clarity of these labels but believe them to be necessary. The plots need to be split by both environmental driver and vegetation group, as otherwise the data become too hard to parse from the plots. We use the Latin alphabet to identify the driver and Roman numerals to identify the vegetation group, together with an explanation in the captions. Both of these remain consistent across Figure 3 and 4 and so we believe the figures are unambiguous.

Discussion on the terrestrial biosphere models: as mentioned above, I think there is a need to describe better processes that can be improved in models to improve the predictability of the fluxes

We thank the reviewer for this comment and agree that, in general, it is an important knowledge gap to better describe the responsible processes. However, to capture environmental memory within TBMs, two separate tasks are required: a) identify and quantify the timescales and b) link these to a process or series of modelled processes. We believe that this study falls within the scope of task a) and adds to the growing body of recent work that is highlighting the timescales at which terrestrial ecosystems respond to the environment (for example, see Bastos et al., 2020; Ciais et al., 2005; Feldman et al., 2020; Liu et al., 2018; Ogle et al., 2015). These results have identified and quantified some of the critical timescales of response, although our understanding of environmental memory is still incomplete - especially when linking these timescales to processes. Identifying the responsible process (or series of) will require model-hypothesis testing against data. While the available flux data could provide a good starting point for this, experimental work that captures step changes in climate or recovery from extremes are likely needed. This is an exciting direction for future research and our future projects may target this more directly.

We have added the below at line 467:

“Our results add to a growing body of research (for example, see Bastos et al., 2020; Ciais et al., 2005; Feldman et al., 2020; Liu et al., 2018; Ogle et al., 2015) that identifies an important role of ecosystem “memory” in the terrestrial fluxes. This first step, including the characterisation of the timescale of influence, the processes affected (e.g., LE vs NEE, etc), the controlling environmental driver and site-to-site variability, is critical to improving TBMs. It is widely acknowledged that capturing legacy processes in TBMs is important (e.g., acclimation, recovery from climate extremes, link between carbon uptake and growth, canopy defoliation, etc.), but to develop the theory, we first need a strong evidence base against which we can probe model predictions. The challenge now is to link the statistical findings to mechanisms and then demonstrate that capturing these processes in models leads to improvements in site predictions. This second step will require applications of our approach (or similar) to both field and targeted experimental data, with progress likely to be made by linking directly to model-hypothesis testing (e.g. Katul et al., 2001; Mahecha et al., 2010)."

Bibliography

Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., Wigneron, J.P., Weber, U., Reichstein, M., Fu, Z., Anthoni, P., Arneth, A., Haverd, V., Jain, A.K., Joetzjer, E., Knauer, J., Lienert, S., Loughran, T., McGuire, P.C., Tian, H., Viovy, N., Zaehle, S., 2020. Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity. Sci. Adv. 6, eaba2724. https://doi.org/10.1126/sciadv.aba2724

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Vesala, T., Valentini, R., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533. https://doi.org/10.1038/nature03972

Feldman, A.F., Short Gianotti, D.J., Konings, A.G., Gentine, P., Entekhabi, D., 2020. Patterns of plant rehydration and growth following pulses of soil moisture availability. Biogeosciences Discussions 1–24. https://doi.org/10.5194/bg-2020-380

Leuning, R., Cleugh, H.A., Zegelin, S.J., Hughes, D., 2005. Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. Agricultural and Forest Meteorology 129, 151–173. https://doi.org/10.1016/j.agrformet.2004.12.004

Liu, L., Zhang, Y., Wu, S., Li, S., Qin, D., 2018. Water memory effects and their impacts on global vegetation productivity and resilience. Scientific Reports 8, 2962. https://doi.org/10.1038/s41598-018-21339-4

Ogle, K., Barber, J.J., Barron-Gafford, G.A., Bentley, L.P., Young, J.M., Huxman, T.E., Loik, M.E., Tissue, D.T., 2015. Quantifying ecological memory in plant and ecosystem processes. Ecol Lett 18, 221–235. https://doi.org/10.1111/ele.12399

Rifai, S.W., De Kauwe, M.G., Ukkola, A.M., Cernusak, L.A., Meir, P., Medlyn, B.E., Pitman, A.J., 2021. Thirty-eight years of CO2 fertilization have outpaced growing aridity to drive greening of Australian woody ecosystems. Biogeosciences Discussions 1–41. https://doi.org/10.5194/bg-2021-218