We thank the editor and referees for the encouraging comments and suggestions to improve the manuscript. We believe that the revised manuscript is substantially improved particularly in detailing the technical aspects of the modelling and input data; discussing soil depth and soil data considerations. The improvements made to the entire revised article are summarised below.

- Added more detail about the model parameterisations
- Included information about the calibration decisions and the rationale behind them
- Added a new supplementary figure to detail the changes in input precipitation rates due to bias correction
- Included discussion on the impact of soil depth on model biases
- Improvements to wording and figures to enhance clarity
- Updated color schemes to be be colorblind-safe

We have given a point-by-point response to all the referees' comments and made necessary changes to the manuscript in response to suggestions. We believe that the revised manuscript fully satisfies the referees' concerns.

We provide complete responses to each comment below, independent of the other comments and responses. While this is comprehensive, it requires some duplication of material throughout the document, but we hope that it helps to swiftly and efficiently evaluate exactly how each comment has been addressed. The line numbers correspond to the revised manuscript.

Referee comments are shown in bold. Author responses are shown in plain text.

Referee 1

General Comments:

This is an interesting paper that investigates the importance of lateral transfers of water and its effects on energy (mainly ET) within semiarid and complex terrain locations of Southeastern Australia. While not novel in addition of any extra model physics, it is an important addition to the scientific community that investigates land surface models and the bridge between hydrologic models. That being said, there are a number of comments and concerns that I have had while reading through this text. Specifically, I am concerned about the method that was used to bias correct precipitation inputs and the calibration period (which was only a length of 45 days). I would implore the authors to better support these decisions within the manuscript. Based on this initial draft, I would rate this as Fair on Scientific Significance, Good on scientific quality (mainly needing more justification), and Good on presentation quality and suggest major revisions to address comments below:

We thank the referee for the encouraging comments, and particularly the suggestions to justify the decisions around bias correction of precipitation and the calibration period.

Major Comments (in order of where they are in the text, not in order of importance):

1. **Paragraph beginning on Line 103: Within this paragraph, the authors explain different overland flow and sub surface flows and how these and cannot feedback into soil water and energy fluxes. Please explicitly state what is meant here by lateral transfers (e.g. case 2b), and if the subsurface flow is still being parametrized despite the baseflow package being turnoff due to calibration. This paragraph is critical to understanding the scientific set-up of the study, and thus needs to be crystal clear.**

We thank the referee for the question. The lateral transfer we refer to includes both overland and subsurface flow (despite the baseflow being turned off), and these processes feedback to affect the LSM states in our 'LAT' simulations. We have revised the description of the scientific set-up for clarity as below (pages 4-5, lines 109-122).

"WRF-Hydro has the capability to simulate overland, shallow subsurface, and channel flows. The subsurface flow is simulated on the LSM grid, while overland and channel flows are simulated on the fine grid. The subsurface lateral flux in the saturated portion of the soil column is calculated based on hydraulic gradients using the method documented in Wigmosta et al. (1994) and Wigmosta and Lettenmaier (1999), implemented in the Distributed Hydrology and Soil Vegetation Model (Gochis et al., 2020). Overland and channel flow is calculated using the diffusive wave formulation, using Manning's equation as the resistance formulation (Gochis et al., 2020). In the simulations presented in this paper, overland and shallow sub-surface flows are modelled, and these processes feedback to influence soil moisture and surface fluxes in the LSM. The water that flows into the channel grids is routed downstream to model streamflow and does not feedback to affect the LSM soil moisture. In other words, channel leakages are not represented in the model. The WRF-Hydro modelling system also has the functionality to specify a conceptual representation of baseflow by passing the underground runoff from the LSM directly into the channel network, but this is not used in the simulations presented in this paper. We calibrate model parameters to match streamflow at selected gauges. Based on preliminary calibration results, the baseflow representation is turned off in our simulations. It is worth noting that lateral saturated sub-surface flow is being modelled in our simulations as described above, despite baseflow being turned off."

2. **Please expand, especially on the precipitation, the bias correction used. Is the idea here that you take a monthly accumulated rainfall at each grid cell from ERA5 land and the Australian Gridded Climate Data (AGCD) and scale each month to directly match the Australian Gridded climate data set? How does this effect the hourly precipitation rates? Infiltration rates will be highly sensitive to the hourly rainfall rates, so ensuring this is clearly explained is critical. See "*Sampson AA, Wright DB, Stewart RD, LoBue AC. The role of rainfall temporal and spatial averaging in seasonal simulations of the terrestrial water balance. Hydrological Processes. 2020; 34: 2531–2542*" for evidence showing that at hourly scale, rainfall is driving much of the uncertainty of infiltration, not necessarily the soil parameters (though these are very much still important).**

Yes, the referee is correct about the idea of bias correction used. We have clarified this in the revised manuscript. We thank the referee for raising the point about the changes in precipitation rates and its effect on infiltration. We agree that this is important to clarify this in the paper. In the revised manuscript, we elaborate on the necessity for bias correction and document the changes in precipitation rates induced by the monthly scaling correction as below (page 6-7, lines 170-178).

"The monthly accumulated precipitation and mean temperature from ERA5-Land reanalysis are scaled to match the AGCD observations for the corresponding month at each grid. This correction is performed as AGCD is a high-quality historical data developed by applying topography resolving analyses methods to in situ observations (Jones et al., 2009) and is expected to be more accurate than reanalysis data. The corrected forcing data would thus match the AGCD closely at monthly timescales, while using the sub daily pattern of variation in the ERA5-Land reanalysis. Compared to ERA5-Land, the AGCD precipitation exhibits higher spatial variation primarily over the mountainous areas in the domain (Fig. A2), and these differences can have substantial effect on infiltration rates in land simulations (Sampson et al., 2020). Application of the monthly scaling correction induces localised increases and decreases in hourly precipitation rates from ERA5-Land as shown in Fig. A2(c)."
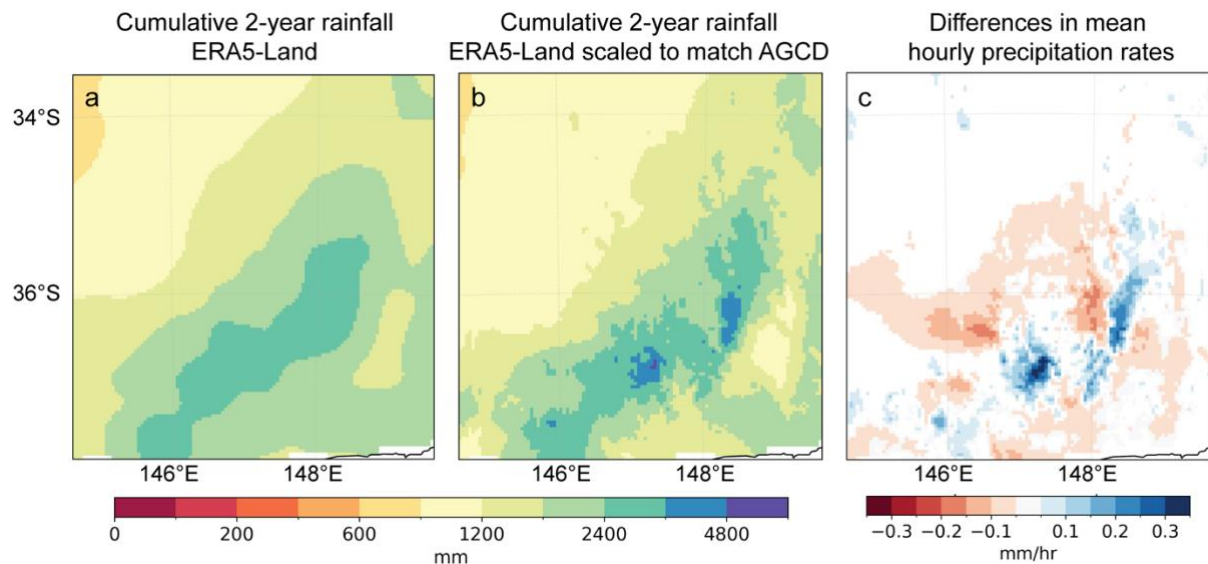
Figure A2: (a) Cumulative 2-year rainfall during 2016-17 from the ERA5-Land dataset, (b) Cumulative 2-year rainfall during 2016-17 from ERA5-Land scaled to match the AGCD dataset, (c) mean differences in hourly precipitation rates in 2016-17. Only grids that show significant differences (at 5% significance level) are shaded in panel (c).

3. **Please provide information on the 45 day period that was used to calibrate the model. Were these high flow days? Were they low flow days? Why was such a small period of time (45 days aggregated by 3 days is 15 data points to calibrate on). More justification is needed. Specifically, why does it make sense here to calibrate to 3 daily flow (assuming accumulated), when the comparisons will be on monthly flow (accumulated as well?) I understand calibration is tricky, and am not advocating for the authors to do more work, but do think that justifying this choice somehow is necessary.**

We thank the referee for raising this important point. Calibration is indeed tricky, and preliminary work was undertaken to arrive at the calibration settings documented in the manuscript. We have included the following information in the revised manuscript to provide justification for the choices (pages 7-8, lines 196-206).

"Calibration of WRF-Hydro is computationally intensive and involves choices that may be aligned to the purpose of the simulations. Here we study the influence of lateral flow on seasonal timescales and hence the main purpose of calibration is to obtain better streamflow outcomes on monthly to seasonal timescales, rather that improved simulations of daily scale streamflow events. The streamflow in the domain primarily occurs in the cool season (May to October), and model simulations using default parameter values exhibit biases during these high flow months (Fig. 2). However, preliminary results showed that event-based calibration to high flow days did not translate to improved monthly flows indicating that it is necessary to use a period at least of the order of a month, which includes both high and low flow days at the four gauges for calibration. As a 45-day period is computationally feasible, and yields reasonable outcomes at monthly to seasonal timescales, this length of time is chosen for calibration. The daily streamflow data is smoothed by aggregating to 3-day flows to dampen the effect of individual high flow days."

4. **Figure 7: Please add a ET Change Relative to CTL label on the y axis. Also please ensure the labels are all correct (CTL1-250 doesn't exist in this study).**

We thank the referee for the comment on Figure 7, and for noting the error in the legend. We have addressed these in the revised version of Figure 7 shown below.
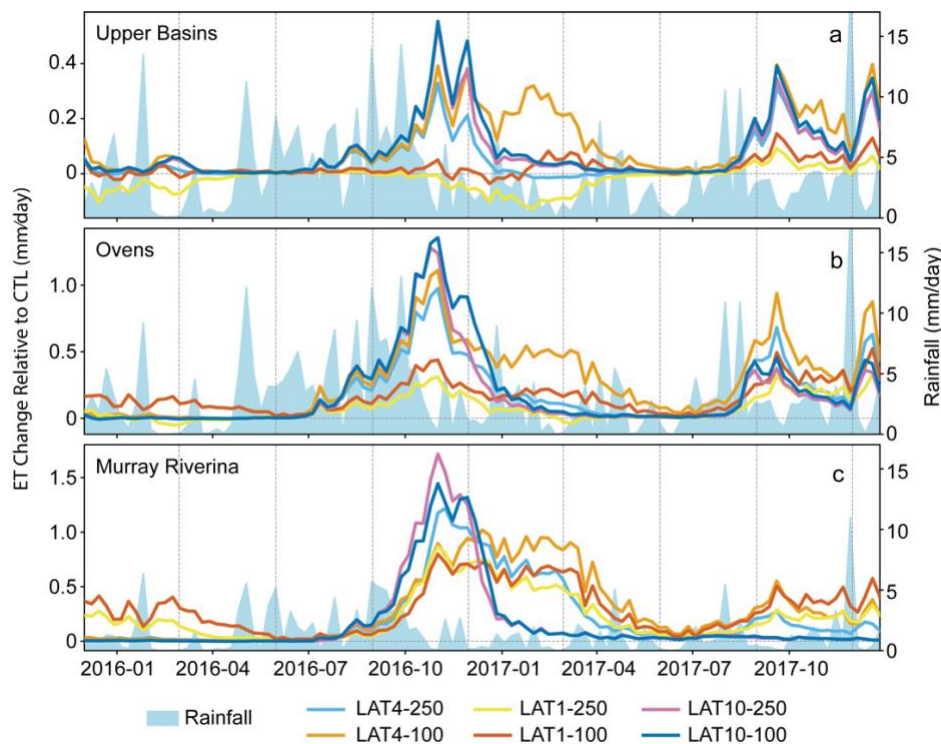
Figure 7: Timeseries of basin average rainfall (filled line plot) and changes in ET (in mm/day; lines) with inclusion of lateral flow in the Upper, Ovens, and Murray Riverina Basins.

Minor comments (in order of where they are in the text, not in order of importance):

5. **Great introduction! I would contend that there could be a nod to some of the work that is being done in the Urban world with lateral transfers (understanding that this is not the scope of this paper, but is an important emerging area where hydrologic processes are just as important and often overlooked in LSMs). I would think a clear location to add would be in the paragraph starting on line 70.**

We thank the referee for the suggestion and have included a mention of the work on lateral flow in urban areas in the introduction (page 3, lines 74 to 79).

"….These studies report that while increases in regional mean soil moisture and evapotranspiration (ET) induced by lateral flow are small, the changes are spatially heterogenous (Chaney et al., 2021; Fersch et al., 2020; Lahmers et al., 2020; Qiu et al., 2024; Zhang et al., 2024). In urban areas, fine scale heterogeneity of impervious areas and open spaces may induce substantial changes in the surface energy balance (Alexander et al., 2024; Reyes et al., 2016). On regional scales, including lateral ground water flow is reported to increase the proportion of ET from transpiration over the United States (Maxwell and Condon, 2016)….."

References added:

Alexander, G. A., Voter, C. B., Wright, D. B., and Loheide II, S. P.: Urban Ecohydrology: Accounting for Sub-Grid Lateral Water and Energy Transfers in a Land Surface Model, Water Resources Research, 60, e2023WR035511, https://doi.org/10.1029/2023WR035511, 2024.

Reyes, B., Maxwell, R. M., and Hogue, T. S.: Impact of lateral flow and spatial scaling on the simulation of semi-arid urban land surfaces in an integrated hydrologic and land surface model, Hydrological Processes, 30, 1192–1207, https://doi.org/10.1002/hyp.10683, 2016.

6. **Figure 1b): please change the outline color of the Upper Basins, Ovens, and Murray Riverina to something that isn't blue. These are currently will be difficult to differentiate given the light blue color used for the terrain height being for much of the lowlands.**

Thanks for the suggestion. We have revised Figure 1 as shown below to incorporate this comment. We have also include channels in the figure as suggested by Referee 2.
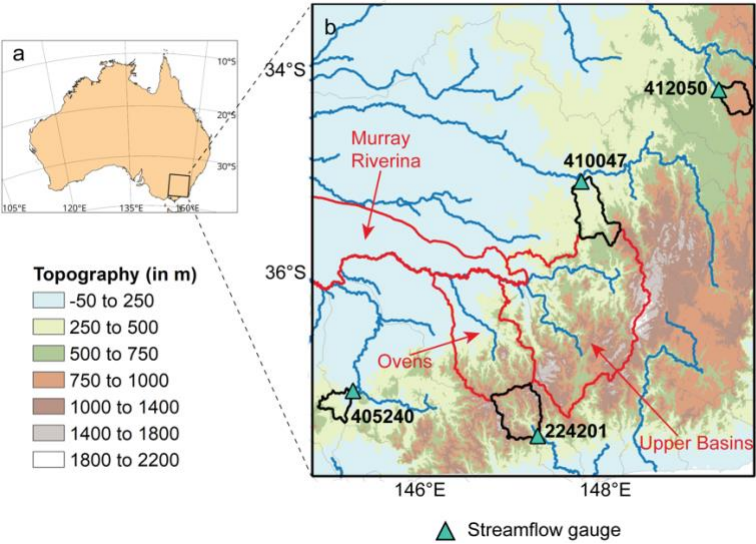

Figure 1: (a) The WRF-Hydro model domain in southeast Australia. (b) Features in the domain. Background shading indicates topography, and the surface water catchments (black outlines) that drain into the streamflow gauges (blue triangles) used for calibration are marked on the map. The basins outlined in red (Upper Basins, Ovens, and Murray Riverina) are used to analyse the influence of lateral flow in basins with varying topographic characteristics. The network of major rivers (blue lines) based on data from Geoscience Australia (https://pid.geoscience.gov.au/dataset/ga/42343) are shown in panel b.

7. **Somewhere within the manuscript or within an appendix, please list the specific choices made for the Noah-MP LSM in terms of physics schemes used. While out of the scope of this paper, these have a very clear influence on the results of the model, and should be listed.**

We thank the referee for raising this. We have added Appendix Table A1 listing the Noah-MP parameterisation options used for the simulations in the revised manuscript.

Page 5, lines 130-131: "The Noah-MP parameterisations used are the defaults in WRF-Hydro (Gochis et al., 2020) and are listed in Appendix Table A1."

Table A1. Noah-MP parameterisation options used for the simulations

Noah-MP physics parameterisation	Option selected
Dynamic vegetation	4 – Leaf area index/stem area index from lookup table; maximum vegetation fraction from climatology
Stomatal resistance	1 – Ball-Berry formulation
Soil moisture reduction for stomatal resistance controlling	1 – Similar to original Noah based on soil moisture
Runoff	3 – Infiltration excess surface runoff and free drainage subsurface runoff
Surface exchange coefficient	1 - Monin -Obukhov similarity
Frozen soil	1 – Hydraulic properties from total soil water and ice (Niu and Yang, 2006)
Supercooled liquid water in frozen soil	1 – General form of freezing-point depression equation (Niu and Yang, 2006)
Radiative transfer	3 - Two-stream approximation applied to vegetated fraction
Snow albedo	2 – From land surface scheme CLASS
Partitioning precipitation into rainfall and snowfall	1 – Formulation as in Jordan (1991)
Lower boundary condition of soil temperature	2 – Fixed lowest soil temperature from input
Temperature time scheme	3 – Semi-implicit but split by snow fraction
Surface resistance to evaporation/sublimation	4 – Sakaguchi and Zeng (2009) for non-snow and separate snow resistance for snow fraction

8. **Line 115: The "eight seasons" seems to be obfuscating the amount of analysis done. Why not just "2 years of results, broken into individual seasons" or something similar?**

Thanks for seeking clarification. The results are presented for four individual seasons from two years of simulations (Figure 4, Figure 5, Figure 6), rather than eight seasons. The wording of this sentence is likely confusing, and we have revised it to convey this clearly.

"We simulate years 2013 to 2017 and analyse the changes in ET during 2015-12 to 2017-11 broken into individual seasons, discarding the first ~3 years as spin-up." (Page 5, lines: 133-134).

9. **I am being pedantic here, but please define monthly streamflow; is this an average or an accumulation over the whole month? I assume it is an accumulation, but could not find it confirmed in the text.**

Thanks for seeking clarification. It is cumulative monthly streamflow, and we have clarify it in in the text in Section 3.1.1 (page 9, line 217), and in the captions of Figure 2 and revised Figure A4.

10. **Figure A3 panel a: why is there a single dot in the middle of the panel behind all of the text. Is this an erroneous plot? Also, please move the Bias and NSE results so that they do not overlap any of the lines. It is hard to read!**

We thank the referee for spotting this. The single dot behind the text was erroneous and we have now corrected the figure. The figure has also been updated so that the text listing the Bias and NSE results do not overlap any of the lines. The revised figure (Fig. A4 in the revised manuscript) is shown below.
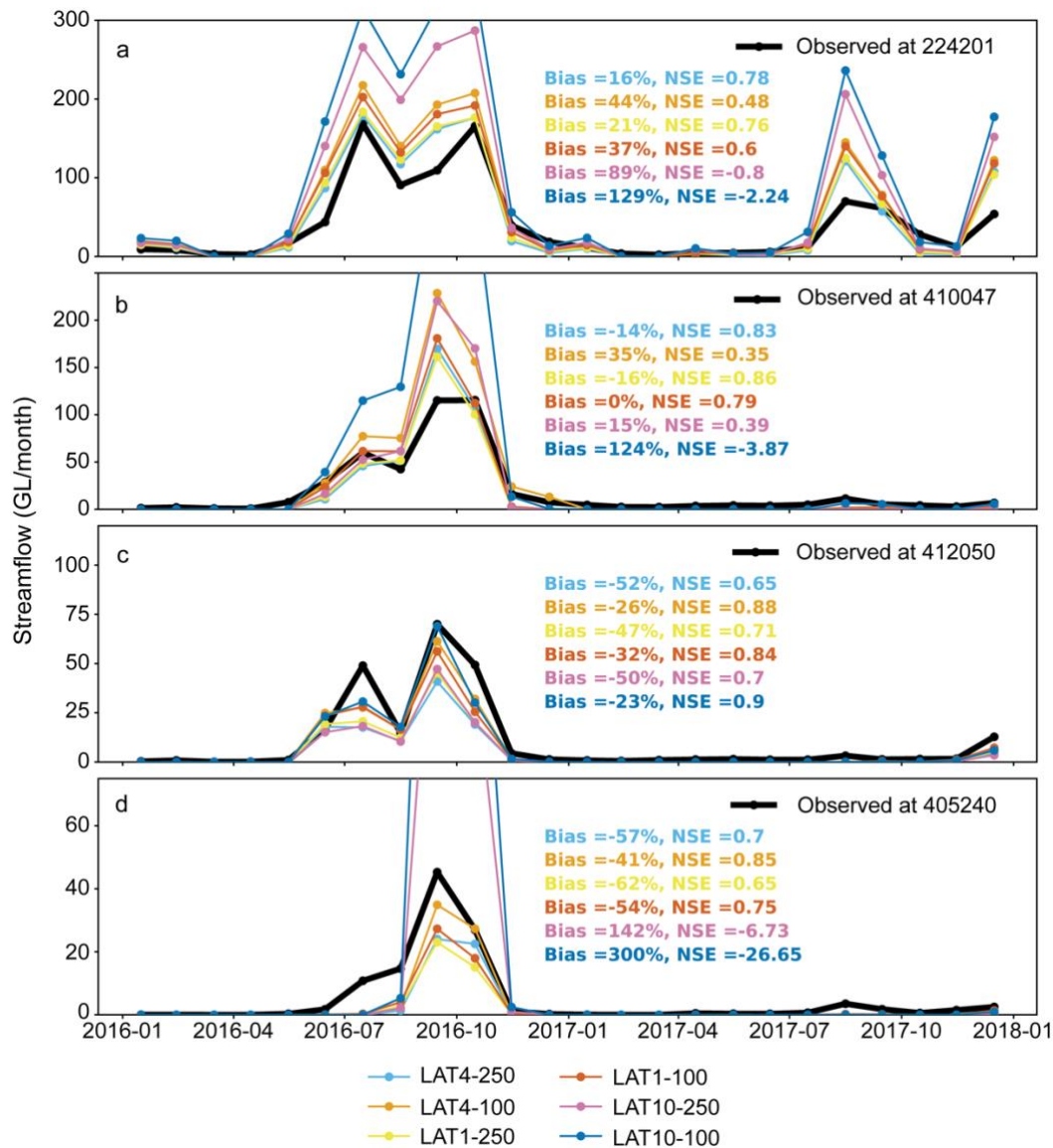
Figure A4: Simulated cumulative monthly streamflow using the calibrated parameter values compared to observed streamflow (in GL/month) at the gauge locations in simulations at different resolutions.

11. **Please revise "*The simulated timeseries of ET are within the range from the DOLCE product most of the time, except in 6 out of 24 months where the simulations are slightly outside this range.*" 25% of the time being outside of the uncertainty range is a pretty significant amount to be outside of the uncertainty estimates.**

Thanks for the suggestion. We have revised this sentence to state that: "The simulated timeseries of ET are within the range from the DOLCE product 75% of the time, and slightly outside this range the rest of the time (6 out of 24 months)." (Page 10, lines 246-247).

12. **For ET in Figure 3: Please specify whether or not this is over the full domain in Figure 1b or just within the sub-catchments of interest somewhere in the text.**

This figure shows the ET averaged over the full domain in Figure 1b. We have clarified this in the text as: "We compare the simulated domain average monthly ET (domain shown in Fig. 1b) during 2016-17 with…" (page 10, line 244).

Referee 2

The results of this paper are significant for the hydrologic modeling and LSM communities, as they demonstrate the impacts lateral flow in LSMs and the possible implications for atmospheric fluxes. The methods of this manuscript are generally strong, and I anticipate that the results of this work could have implications for the hydrometeorological community.

However, I have some technical concerns on the methods that I ask the authors to clarify in order for this manuscript to be accepted.

We thank the referee for the encouraging comments, and for seeking clarification on some of the technical aspects.

Major Comments:

1. **Section 2.1, Lines 100-110: I appreciate this discussion on the significance of horizontal routing with LSMs. I also recommend discussing the impacts of LSM depth, as the standard 2-m Noah-MP depth often does not capture groundwater processes and has resulted in dry biases with ET in some regions.**

We thank the referee for raising this and agree that the 2-m soil depth may contribute to LSM biases. In the revised manuscript, we have elaborated on this point as below (page 5, lines 123-127).

"Here we use the standard Noah-MP LSM which has a constant soil depth of 2 m with vertically homogeneous soil parameters. This formulation can contribute to biases in runoff and evapotranspiration, which may be ameliorated by incorporating variable and higher soil depths, groundwater processes, and vertical soil heterogeneity (Gochis et al., 2010; Barlage et al., 2015; Wu et al., 2021; Yimam et al., 2025) in the modelling framework. These aspects are outside the scope of the work presented in this study and have not been explored here."

References added:

Barlage, M., Tewari, M., Chen, F. *et al.* The effect of groundwater interaction in North American regional climate simulations with WRF/Noah-MP. *Climatic Change* **129**, 485–498 (2015). https://doi.org/10.1007/s10584-014-1308-8

Gochis, D. J., Vivoni, E. R., and Watts, C. J.: The impact of soil depth on land surface energy and water fluxes in the North American Monsoon region, Journal of Arid Environments, 74, 564–571, https://doi.org/10.1016/j.jaridenv.2009.11.002, 2010.

Wu, W.-Y., Yang, Z.-L., and Barlage, M.: The Impact of Noah-MP Physical Parameterizations on Modeling Water Availability during Droughts in the Texas–Gulf Region, Journal of Hydrometeorology, 22, 1221–1233, https://doi.org/10.1175/JHM-D-20-0189.1, 2021.

Yimam, Y. T., Neely, H. L., Morgan, C. L. S., Kishné, A., Gross, J., and Gochis, D.: Evaluation of Noah-MP performance with available soil information for vertically heterogenous soils, Agrosystems, Geosciences & Environment, 8, e70048, https://doi.org/10.1002/agg2.70048, 2025.

2. **Section 2.2 and 2.3, Calibration: Please clarify why the authors use 3-day averaging for streamflow validation and calibration? This could likely underestimate major surface runoff events leading to flash flooding. Furthermore, calibration to 45-day periods may not capture the full range of processes that lead to hydrologic response. I recommend demonstrating that these parameters are consistent with a longer range calibration period. If this is not easily feasible within project constraints, I alternatively recommend discussing the assumptions and possible limitations behind this methodology decision.**

We thank the referee for the question and recommendations. Preliminary work was undertaken to arrive at the calibration settings documented in the manuscript. As computational resources preclude us from demonstrating consistency with a long-range calibration period, we have included the following information in the revised manuscript to elaborate on the calibration choices and its implications (pages 7-8, lines 196-206).

"Calibration of WRF-Hydro is computationally intensive and involves choices that may be aligned to the purpose of the simulations. Here we study the influence of lateral flow on seasonal timescales and hence the main purpose of calibration is to obtain better streamflow outcomes on monthly to seasonal timescales, rather that improved simulations of daily scale streamflow events. The streamflow in the domain primarily occurs in the cool season (May to October), and model simulations using default parameter values exhibit biases during these high flow months (Fig. 2). However, preliminary results showed that event-based calibration to high flow days did not translate to improved monthly flows indicating that it is necessary to use a period at least of the order of a month, which includes both high and low flow days at the four gauges for calibration. As a 45-day period is computationally feasible, and yields reasonable outcomes at monthly to seasonal timescales, this length of time is chosen for calibration. The daily streamflow data is smoothed by aggregating to 3-day flows to dampen the effect of individual high flow days."

Minor Comments:

3. **Figure 1: Consider adding the channel network (even if it is only higher order channels) to help the reader visualize the hydrologic connectivity.**

We thank the referee for the suggestion to improve the figure. We have revised Figure 1 as shown below to include the channel network.
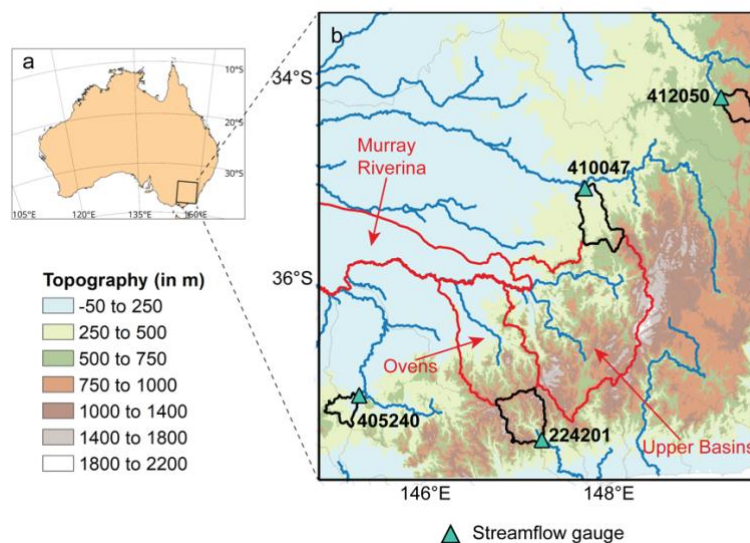

Figure 2: (a) The WRF-Hydro model domain in southeast Australia. (b) Features in the domain. Background shading indicates topography, and the surface water catchments (black outlines) that drain into the streamflow gauges (blue triangles) used for calibration are marked on the map. The basins outlined in red (Upper Basins, Ovens, and Murray Riverina) are used to analyse the influence of lateral flow in basins with varying topographic characteristics. The network of major rivers (blue lines) based on data from Geoscience Australia (https://pid.geoscience.gov.au/dataset/ga/42343) are shown in panel b.

4. **Section 2.2.1, Geographic Data: I find it surprising that the TERN dataset produced worse streamflow compared to the default soil dataset. Is this something that could eventually be improved with calibration?**

We thank the referee for the question. We agree that it is surprising that the TERN dataset produced worse streamflow, and we had briefly discussed this in the original manuscript (lines 135-145). Based on our results, rather than improving through calibration, we speculate that that it may be possible to obtained improved streamflow outcomes by utilising the TERN soil data in conjunction with regional soil parameter measurements.

We have improved upon lines 135-145 from the original manuscript as below.

Revised manuscript page 6, lines 153-161

"The better streamflow simulations obtained using the default soil dataset rather than the TERN dataset may be surprising, as the TERN data, which utilises regional observations (Teng et al., 2018), likely provides more accurate soil information over Australia. This is possibly because the modelled influence of soils on surface water partitioning in each land column relies on soil parameters in addition to soil type. The land model uses parameters (such as moisture at saturation, field capacity, wilting point, saturated hydraulic conductivity) for each soil type, the default values for which are defined based on scarcely available field observations in various regions (Kishné et al., 2017). Our results suggest that regional measurements that can be used to refine the default parameters values in conjunction with the regional soil datasets, such as the TERN dataset, may be necessary to obtain improved simulations."

5. **Section 3.1.2, lines 215-220: The negative ET bias might reflect the limits of the 2m LSM. I recommend connecting back to this point in the discussion.**

We thank the referee for the suggestion to include the point about soil depth in the discussion. We have included this point in the paragraph detailing avenues of future work. The text now reads as below (page 19, lines 389-391).

"…, more accurate representation of processes including baseflow and the feedback from the channels to the soil columns may be relevant in arid and semi-arid regions, such as southeast Australia. These processes are not represented in our simulations. Our calibration results suggest that spatially varying baseflow may improve lateral flow representation in our domain. Incorporating soil layers deeper than the 2-m depth modelled in standard Noah-MP have reduced surface flux biases in some cases in other semi-arid locations (Gochis et al., 2010; Barlage et al., 2015) and may be explored in future work. Incorporating channel seepages have improved streamflow simulations in semi-arid Arizona (Lahmers et al., 2021), but the feedback of this process to the soil column has not been modelled yet."