Comment 1. The authors have brought together many analyses into a useful, integrated platform. As they point out, a fully distributed model that can capture all of the processes represented would be unwieldy at best and unattainable at worst. On this basis, I fully support the development of this tool.

Response 1. We thank Prof Ferre for taking the time to review the article, being supportive of the aims of our modelling endeavour, and for providing very helpful comments that will certainly improve the manuscript.

Comment 2. My only concern is that this model, while an improvement, still requires many simplifications. The authors have sufficient experience and expertise to know which module is best to use for specific conditions ... especially as they are familiar with the Walnut Gulch site and with modeling COSMOS data. Could they provide more guidance for less well-informed users? Which assumptions are most critical? Under what ranges of conditions should this model not be used? How could a user identify potential systematic errors from the model results? In short, simpler models are better - but, only if they aren't significantly flawed. Can the authors help to make sure that their model will be used appropriately and constructively?

Response 2. Thank you for raising this concern – the model does indeed require many simplifications, and we agree that it would be useful for us to add more to the revised paper to help make sure the model will be used appropriately.

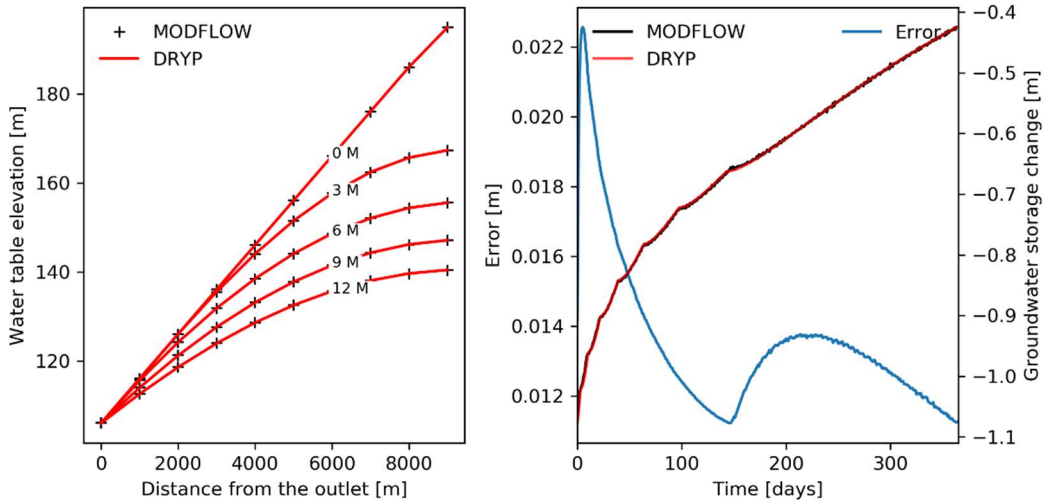
In terms of adding more guidance for users, we will re-iterate that one of the main advantages of parsimonious models is the ability to be efficiently (numerically/computationally) implemented. This allows more robust evaluation of the uncertainty of the model by considering combined structural and parameter uncertainty estimation, for example via Monte Carlo experiments. Such uncertainty analysis, which is made possible by DRYP's relative parsimony, can give insights into the potential systematic errors resulting from the model, and we will make it clearer in the revised paper that users are strongly advised to perform such analyses to ensure that the model is fit for purpose for their intended application.

In terms of the ranges of conditions for which the model is not applicable, we note that overland flow and routing processes in dryland regions are often highly dynamic mechanisms, both temporally and spatially, and greatly influence the water partitioning. DRYP has not been developed to capture these characteristics, which may be required for other applications such as flood forecasting. Rather, we aim to capture the long-term changes/processes of the main components of the hydrological cycle, such as drought, as stated in Lines 99-101: "*We do not intend for this model to accurately simulate event-based flood hydrographs, for example, for flood hazard analysis. Instead, we aimed to develop a model that captures the long-term behaviour of the water balance in dryland regions.*" Hence, in this context, an important way the model should *not* be used is if the user is looking to simulate the highly dynamic variation of flow in channels or on the surface. In more general terms, we don't see any particular restriction on the use of the model for other climatic conditions beyond drylands, as long as the main assumptions, especially regarding the infiltration and flow routing components, are considered valid to the particular application. We will update the revised manuscript to discuss these aspects in more detail.

Comment 3: If I have one other quibble, it is with the testing of the model. It is true that spatially distributed numerical models are expensive to run and I wouldn't suggest that they need to build and

run one to test the applicability of their simplified model. But, I would have thought that someone, somewhere, would have a model that could be used as a basis for comparison to give a better estimate of the magnitude and distribution of errors. It isn't my field, exactly, but I'm not sure that mass balance errors alone are sufficient to validate a model.

Response 3: Thanks for this suggestion. In response, we have now run some additional synthetic experiments against the industry standard software MODFLOW, to support the documentation of the ability of DRYP to represent the surface-water groundwater process interactions robustly. The results show excellent agreement between DRYP and MODFLOW models with respect to how the groundwater components interact with the drainage, as can be seen in the summary of the results shown below, and these experiments will be documented in detail in the revised paper.



a) Simulated head along the aquifer for different time steps estimated by DRYP (solid lines) and MODFLOW (dashed lines), and b) temporal variation of the mass balance error for DRYP.

Comment 4. All of these comments should be seen as suggestions only. It is entirely possible that I am out of touch with what is expected/required for models at this scale and especially for use in hydrometeorology. I leave it entirely to the authors' discretion to decide to consider or ignore my comments. Nice work! Ty Ferre

P.S. I did see one spelling error on a plot axis ... cummulative. I first thought that it might be a British spelling, but Google doesn't support that hypothesis.

Response 4. Many thanks – your comments were very useful and we hope you feel we have done them justice in our responses. We will correct the spelling error you have correctly spotted in the revised paper.