Reviewer 1

We greatly appreciate the comprehensive, detailed and useful feedback provided by the reviewer. Below we address all the comments and suggestions provided in order.

1. 1. The manuscript presents the updated version 2.0 of the hydrological modelling software DRYP 1.0 described in GMD in 2021:

Quichimbo, E. A., Singer, M. B., Michaelides, K., Hobley, D. E. J., Rosolem, R., and Cuthbert, M. O.: DRYP 1.0: a parsimonious hydrological model of DRYland Partitioning of the water balance, Geosci. Model Dev., 14, 6893–6917, https://doi.org/10.5194/gmd-14-6893-2021, 2021.

The modeling software aims to represent important hydrological processes in the dry regions of the globe, including transmission losses in channels, local and non-local groundwater recharge, and streamflow intermittency, processes difficult to simulate, particularly in large-scale hydrological models. The update includes the consideration of additional water storage compartments (interception, ponds, and lakes) as well as an improvement in run-time due to more efficient algorithms and the use of Fortran for some software parts. The model is applied for a few synthetic cases and the Horn of Africa dryland (HAD) region (2 million km^2 with a spatial resolution of 1 km^2 and hourly time steps). The uncalibrated model, which is able to represent water flows between streams and groundwater as well as between lakes and groundwater based on hydraulic head gradients, is compared to remote sensing-based estimates of evapotranspiration, soil moisture, and terrestrial water storage anomaly.

Overall, the description of the model enhancement lacks clarity. Also, the three model applications are insufficiently described in the manuscript. The application for HAD lacks information about which lakes are taken into account in this application, and the resulting exchange flows between lakes and groundwater, and streams and groundwater. And what information was used to simulate interception and ponds, and what is the impact of these new components on the model results?

The main structure of the model and all of its components is discussed in detail in the first version of the model (Quichimbo et al 2021 DRYP 1.0). In this manuscript we briefly outlined all enhancements in the main section and provided a detailed description of the model updates in the appendix. We also refer to the first version of the model and its relevant component within the text where a full description of each of the components is described. However, we appreciate the reviewer's comment, and we will provide more clarity on how the new enhancements link to the model structure and to the new results.

I disagree with the statement in L10 of the abstract, “The results highlight the ability of the model, even without calibration, to reproduce global remote sensing data such as soil moisture, actual evapotranspiration, and total water storage”. Only correlations were assessed, and the correlations for soil moisture were mostly rather poor, with the majority of the area showing a Pearson correlation coefficient below 0.6. The manuscript does not show how meaningfully and successfully a model simulating hydraulic head gradients and including many parameters that are difficult to derive by model calibration can be applied at a large scale. The synthetic experiments only cover small regions, and the HAD application is not well described and evaluated enough to be convincing.

We will soften the language in the abstract – we will remove the word “correlation”, state that the model is uncalibrated but that it can broadly capture the magnitude and timing of fluxes and stores. We used the uncalibrated correlation to show readers the ability of the model to capture the dynamics of fluxes and stores represented by the model. However, calibration is out of the scope of this manuscript although, as part of the conclusion section we provide some guidance for future work and suggest regionalization approaches to improve the parameterization of the model.

I suggest that the HAD application is run with v1 and v2 of the DRYP model, and differences in the fit to the three remote-sensing-based variable values (plus streamflow) are analysed.

Thank you for this suggestion. However, since DRYP v1.0 was developed for basin scales while DRYP v2.0 was extended for regional scales, inherently the two versions are only comparable at basin scale.

Main comments

Introduction

I suggest expanding the introduction regarding the following:

1. Explain clearly that 1) modeling of bidirectional water flows between surface water bodies and groundwater requires the simulation of hydraulic heads both in the surface water body and the groundwater, 2) most global-scale models are storage-based and can therefore not simulate bidirectional flows dynamically.
2. Instead of only stating that regional to global-scale models lack processes such as bidirectional water flows between surface water bodies and groundwater, I suggest describing those regional to continental models that are, at least in principle, capable of simulating bidirectional water flows because they compute hydraulic heads in surface water bodies, soil, and groundwater (using difficult to solve partial differential instead of ordinary differential equations). These include ParFlow and HydroGeoSphere, who have been applied at the regional to

continental scale. How do they compare to DRYP? And what about land surface models, which sometimes compute the groundwater table elevation.

We acknowledge that there is indeed a wide range of models including those mentioned by the reviewer (ParFlow, HydrgeoSphere), as well as coupled modelling systems that integrate bidirectional interactions between different hydrological stores including lakes. However, as the reviewer mentions, these models usually use computationally demanding approaches that we seek to overcome using a simplified approach that preserves critical physically-based assumptions such as lateral groundwater flow driven by the hydraulic gradient. Following the reviewer's suggestions, we will expand the introduction and clarify the approaches adopted within existing models and their constraints.

Methods

The description of the model (algorithms) is not sufficiently clear. Both the synthetic experiments and the application to HAD need a lot of clarification to make them well understandable by the reader (see my specific comments below). In particular, no information is provided on the integration of lakes and ponds in the HAD application in section 4.2, nor on how the newly introduced interception was parameterized in this application (was it even included?). The appendix provides detailed descriptions of the new model components, which are, however, difficult to follow. In particular, information on the values or sources of data is not provided for all the parameters of the algorithms. Just as an example, in A3, it is stated in L486 "Here, it is assumed that the parameters a, hmax, and Amax are known". But no information is given in the application to HAD about the values/sources used for these three parameters. And this is exactly the challenge of large-scale modeling: How to find suitable values for, e.g. a or Amax? And how to decide which surface water bodies interact with the groundwater (here: lakes) and which not (here: ponds)?

We will edit the manuscript to clarify the description of the synthetic experiments and the application to the HAD. We will also present the data sources used for model parameterization in the Methods section.

Regarding the ponds and lake parameters, hmax and Amax can be obtained from existing datasets at regional and global scales such as GLOBathy and local information in different regions, as was stated in line 221. Remote sensing datasets can also be used to derive maximum extent of water bodies at different scales. Additionally, available depth and volume information can be used to derive the parameter "a" for pond calculations. We will expand this section to include additional references to guide the reader on the parameterization of the lakes and ponds.

Results

Also the information provided in the results section is not precise enough to be readily understandable (e.g., regarding the temporal resolution and extent of the evaluation data used for correlation analysis in HAD). Besides, it is methodologically incorrect to evaluate the correlation with GRACE TWSA for spatial units of 0.25° due to the inherent low spatial resolution of GRACE signals (see my comments to section 3.3 below).

We will clarify the description of the results as well as limitations of the approach. We understand the reviewer's concern about the resolution of GRACE TWSA and the constrains of the evaluation approach. We have only used the correlation at grid-by-grid approach to show the spatial variability of the model outputs. In fact, to show the consistency of the model capabilities with a remote sensing dataset we added the basin scale evaluation of the Juba basin, which uses a consistent aggregation of GRACE TWSA data. The size of the Juba basin is well above the recommended area for GRACE data evaluation. Therefore, in order to avoid misleading the reader on the use of GRACE TWSA, we will remove the correlation plot between the simulated and GRACE TWSA.

Specific comments

The reference Quichimbo et al. (2021) should cite the GMD paper, not a Zenodo page.

Agreed, we will make the change.

L9: Write "about 2.1 million km^2" not "2,000,000 km^2"

OK, we will make the change

L21: I think "streamflow intermittence" is a more suitable term than "stream intermittence".

OK, we will make the change

L31: In standard hydrological models, groundwater is assumed to be discharged to the river and maintain low flows. So what is described here as non-local hydrological connectivity is normally represented by hydrological models (and not omitted, as written in L25).

As the reviewer points out, baseflow or groundwater discharge is normally represented in hydrological models. However, we refer here to the dependency of groundwater (focused) recharge via transmission losses on streamflow generated by runoff in upland areas. We will rewrite this paragraph to more clearly describe the limitations of large-scale hydrological models in representing focused groundwater recharge.

L54: Clarify why you use the term "process-based" here

We refer here to models that use a process-based approach but that lack ability to represent dry conditions and therefore cannot capture flow intermittence. For physically-based models, such as ParFlow, representing dry conditions is more challenging due to the numerical constrains (e.g. small-time steps).

Fig. 1 correct typos such as throughall, Developemement, …. Include lakes in Figs. 1 b and c.

We will make the edits.

L187: "River cells are characterised by the riparian unsaturated zone, while non-river cells are characterised by the hillslope unsaturated zone." Does this mean that a cell either has at store UZ (in Fig. 1) or a store RUZ (in Fig. 1) but never both? I.e. there is no diffusion groundwater recharge in river cells? If so, please indicate this in Fig. 1, e.g. in Fig 1c by making whole cells either blue river cells or green hillslope/non-river cells.

A cell always has UZ but the RUZ is applied in addition when a stream is added to the cell. We will clarify this in the text accordingly.

L191: For easier understanding, I suggest having two equations instead of eq. 7, one for the river cells and one for hillslope/non-river cells.

OK, we will make the change

L215: Is QPND routed as overland flow to the downstream cell or as streamflow? Please clarify in section 2.5.1 or before which type of surface water bodies is represented as "pond" by Eq. 10. What about the reservoirs and small lakes mentioned in the title of section 2.5.1? Hamonize with the title for A3.

Q_PND is the excess flow and is routed as overland flow to the next downstream cell. We will update this statement for clarity.

L234 and Fig. 2: Please explain whether it is assumed that a lake does not have a surface water outflow? What precisely is Q? If you assume no surface water outflow of a lake, under what circumstances is such a representation suitable for representing reality?

In addition to the fluxes across the wetted area of the lake, any water surplus resulting from water reaching the top of the lake, is considered as output. We will modify the figure to clarify this point, as well as expand the description of the component in the methods section.

Section 2.2.7 In Eq. 11., "Kaqh" is included, and not transmissivity. Please clarify under what circumstances the aquifer transmissivity is described by a time-dependent K*h and when by the transmissivity options described in section 2.2.7. Besides, I understand that there is only one layer of saturated grid cells. So what does a depth-dependent transmissivity as shown in Fig. 3 mean? Do you mean a depth-dependent hydraulic conductivity of the aquifer that results in different T of the saturated layer dependent on the C, L or E function? This needs to be explained much better. Please provide a suitable equation with transmissivity.

We will clarify this in the corresponding section. K_aqh is the K in the transmissivity term. We will fix all the text and equations using this term.

Regarding the transmissivity functions, the transmissivity varies with changes in hydraulic conductivity, which is depth dependent in the model. We will clarify this in text as well as provide a more detailed description of this model capability.

Multi-transmissivity options must be selected by the user by providing a raster file containing the spatial distribution of transmissivity types. The file should contain a code at each cell of one of the transmissivity functions available (0: constant, 1: linear, and 2: exponential). By the default the model uses a linear dependent approach. We will clarify this in the text.

The exponential function available in this version of the model has already been described in DRYPv1.0, and has been extensively used in other large-scale models (Fan, et., al. 2013) We will clarify this in the text and add the corresponding references, we will also expand the description in the Appendix.

Fan, Y., Li, H., Miguez-Macho, G., 2013. Global Patterns of Groundwater Table Depth. Science 339, 940–943. https://doi.org/10.1126/science.1229881

Sections 2.3.1, 2.3.2: Please indicate all parameter values used for the synthetic experiments (maybe in a supplement). The requirement is that the reader could recreate these experiments. Also the boundary conditions are not clearly indicated in the figures and the text.

We will add all the corresponding parameters in the manuscript as well as improve the figure to clearly indicate the boundary conditions.

L271: "Initial conditions assumed dry soil, with water content equal to the wilting point. For groundwater, the initial water table was specified as a flat surface located 1 m below the lake bottom elevation (95 m above the reference datum)." I cannot recognize this in Fig. 4. Also, Fig. 7 does not show a flat gw table at t-0, and the water table is exactly at the lowest point of the lake. Very confusing! It seems that Figure 4 b is incorrect regarding the initial water table; also indicate the elevation of the constant head boundary in m (<84 m). And indicate what the role of the maximum lake extent (Fig. 4) or maximum lake depth is in the modeling. It is not reached at t-4, will it be reached at steady-state?

Figure 4 is a schematic of the geometry of the lake model and it is not scaled, therefore, all dimensions are only for reference. However, following the reviewer's comment we will update figure 4 to clearly show the initial condition and the boundary conditions accordingly as indicated in the text.

Regarding figure 7, panel 7b shows the temporal variation of the water table at different locations along the y-axis, one point is located in the deepest point of the lake whereas the other further away. Figure 7b shows the increase in the water table and the lake emergence when it interacts with the water table. For the period shown in the figure, the system has reached a dynamic steady state, where the precipitation, actual evapotranspiration/evaporation, and flow at the boundary conditions have reached the dynamic balance. The deepest point in the lake (p4), as expected, is the one that slowly increases due to the high storage capacity of the lake (S=1), whereas at the other location the change is considerably high given the low storage capacity (0.1) of the aquifer in relation to the lake. This is the reason why P1, P2, and P3 show a rapid increase in the water table in relation to the P4. It also should be noted that P1 slowly increases in relation to P2 as a result of the influence of the boundary condition. For this test case, the maximum lake elevation or lake maximum extend is not reached, as the water level in the lake increases only one meter.

We will edit this section to provide a clear description of the figure as well as the process.

Figure 5: I do not see the flow boundary. And what is indicated with “model domain”?

We will update the figure accordingly to clearly show the boundary conditions.

Regarding the model domain, the model is the representation of an aquifer system where the ratio between the length and the width is very large, so it can be assumed as a 2-D flow. Thus, the “model domain” refers to the cross section of the aquifer. We will clarify this in the figure caption.

Line 281: Figure number missing.

We will fix this.

Figure 6: Indicate Shabelle River as well as the Ogaden area in Fig. 6

We will add this to the figure 6.

Table 1: hlake: what exactly is meant by hlake; and what about the location of extent of lakes? Were ponds considered? Please describe how hlake is derived by Khazai et al. (2022). What is the source of Kch = 10.9 mm/h?

Ponds were not included in the model and we will clarify the manuscript in this regard. Lakes were included, and the parameterization was achieved by generating raster files from the lakes parameters datasets.

hlake is the lake bathymetry, we will clarify this model description section.

Rasterisation was achieved by using the publicly available tool "*Generate_Bathymetry_Rasters.py*" available via Google Earth Engine. We will expand the description of the model parameters to explain our approach to this.

Section 2.4.2: Also compare to in-situ observed streamflow, e.g. the data available at GRDC, at least roughly, even if observed data are only available before 2000. If only correlations between the ESA CCI soil water content and the DRYP value are analyzed, is a normalization necessary? Provide information on the temporal resolution used for the correlation analysis in 3.3 (or provide this information in 3.3.)

We evaluated the model at the monthly time scales for all variables. Model outputs were spatially aggregated using the average to match the resolution of the dataset used in the evaluation.

Regarding the GRDC, following the reviewer comment, we will add the comparison of the simulated streamflow with observed streamflow available for the Juba basin.

Regarding to the normalization of ESA-CCI and model outputs, we performed it in order to compare the soil water from the model and ESA-CCI. This is because soil moisture in the model represents water content in the rooting zone whereas ESA-CCI covers only the shallow layer (< 10mm). We will add additional justification for this to the paper.

Section 2.4.3, L330 Indicate with what forcing the 20 years were driven.

The forcing dataset for the warming up period is the IMERG dataset, which is described in section 2.4.2. The start and end dates are 01/07/2000 and 01/07/2021, respectively, corresponding to 20 years starting from the start date of IMERG dataset. We will modify the text to clearly explain this in the manuscript.

Section 3.1: Please provide the water balance of the model domain, in particular, which fraction of the precipitation becomes evaporation from the lake and the soil, and which fraction discharges via the constant head boundary. Please explain whether groundwater table elevations are shown in Fig. 7b, c, and d. If so, what is the water table of the lake? And why is the (groundwater)table in the lake lower than at P2? Is the lake a gaining lake, and under what circumstances would the lake discharge water into the saturated zone and thus towards the constant head boundary? And explain better the relevance of the model experiment for the simulation of the real world.

We will add the water balance for all the modelled stores and fluxes. The example shows the emergence of a gaining lake. The main purpose of this example is to show the mechanism in place during the lake emergence due to the interaction with groundwater. Thus, one of the main fluxes into the lake is lateral flow and saturation excess flow from zones close to the lake. Figure 7b shows the temporal variation of the water table at different locations and over time. Time series at P2 is on the side of deepest point in the lake, however, it is not yet connected to the lake. The connection will only happen when the lake water level and consequently lake extend reaches P2. As the lake is not connected, the increase in water table will depend on the recharge and hydraulic properties of the aquifer. Therefore, for the given conditions, high recharge values, and low hydraulic conductivity, the water table rises above the lake surface elevation. A losing lake condition will only occur when the water table falls below the lake water elevation. A mixed condition may also happen when some sections of the water table around the lake are above and some are below the lake water elevation.

To clarify the process, we will expand the description of this component in the method section as well as in the results section.

Section 3.3

While GRACE terrestrial water storage anomalies in the form of CSR-M mascons (Save et al., 2016) are provided with spatial resolution, evaluation can only be done for spatial aggregates of at least about 100,000 to 200,000 km2 (see also https://www2.csr.utexas.edu/grace/RL06_mascons.html). Therefore, 0.25° grid cell values cannot be compared to the DRYP output (Fig. 9c), but both CSR-M and DRYP values have to be first aggregated to regions of such a size. Save et al. (2016) write "While these mascon solutions are estimated on geodesic grid roughly 120 km wide, that is not the resolution of these GRACE solutions. These GRACE mascon solutions are limited by the band-limited nature of GRACE, with an approximate resolution of around 250–300 km near the equator."

We appreciate the reviewer's comment that the figure is potentially misleading so we will remove the correlation plot and only show the aggregated results over the Juba basin as example of model performance.

Fig. 9. It is the temporal correlation that is shown. Provide, in the figure caption, the temporal resolution and time period of the correlation analyses.

The temporal resolution of the figure is monthly, we will add the resolution and the evaluation period to the caption.

L401: While correlation is high, the temporal variability of AET is underestimated by DRYP and it looks like the long-term mean evapotranspiration is underestimated, too.

This underestimation is due to the model being uncalibrated, but it also reflects the fact that the GLEAM data are also a modelled output (with its own uncertainties). Calibration is out of the scope of this manuscript but we will expand the result section to clearly state this limitation.

Appendix A4: I do not see the relevance of the synthetic experiment. In addition, the applied precipitation amount is unrealistically high for almost all regions of the globe.

Appendix A4 only provides additional guidance to understand model parameterisation and application; it provides an example of the dynamics of the process and how it is represented in the model. We acknowledge that values are considerably high, however, they are intended only as example of simulation. They provide a first view of how the changes in key parameters affect the performance of the simulation. The high values of precipitation allow for the generation of enough water as runoff in a short period of time to fill the ponds despite their different sizes. This enables us to observe the temporal variation of storage, area, and stage changes for the three analysed cases. We consider that this example, although simple, provides an impression of the potential impact on the model representation of ponds and consequently the performance of the model. We believe these ponds are an important (and overlooked) element in a region where people and livestock rely on shallow ephemeral water bodies.

Code and data availability: To make all described modeling exercises reproducible, scripts, input data (in particular parameter values) and output data files have to be provided. It is not enough to provide the code and the sources of the input and validation data.

We strongly agree with reviewer comments, we are committed to the key principles of reproducibility and will provide access via permanent public repository to all scripts and datasets required to run the model. We should emphasize the all the data used in the manuscript are already publicly available. Although the forcing datasets are too large for sharing and more importantly requires further permission from the owner to add them to another repository, we will provide to the user all the scripts required to access all the information and dataset along with the main code of the model.