Review of Oldham et al: Evidence-based requirements for perceptualising intercatchment groundwater flow in hydrological models.

<table>
<tr><td colspan="3">RC3: 'Comment on egusphere-2022-529', Anonymous Referee #3, 11 Oct 2022 reply</td></tr>
<tr><td>1.</td><td>It is a common fact that water actively exchanges between surface divides in regions with carbonate fractured aquifer outcrops. This have challenged the ‘watertight substratum’ assumption that is the foundation of many existing catchment rainfall-runoff models for long time without appropriate solutions and model conceptulizations. The main aim of this manuscript is to improve our perceptual models of intercatchment groundwater flow. The authors took advantage of their wealth of data, densely gauged river network, and geological variability from national meteorological, hydrological, hydrogeological, geological and artificial influence datasets to develop a perceptual model of intercatchment groundwater flow (IGF) and to show how it varies spatially and temporally in 80 subcatchments of the River Thames, United Kingdom (UK). The water balance, presence of gaining/losing river reaches and intra-annual dynamics were investigated through a water balance analysis.</td><td>The authors would like to thank the reviewer for their detailed and constructive comments on our paper.

In response:</td></tr>
<tr><td>2.</td><td>The study is important for hydrological predictions and water resources management in groundwater-activated catchments. However, the method adopted by the authors can only provide site-specific results about qualitative water balance, it is still difficult to represent regional inter-catchment groundwater dynamics as they could not provide some essential functions that describing how groundwater between neighbor units exchanges according to different conditions of groundwater levels, different lithology, human abstractions and so on. In order to couple IGF processes into existing hydrologic models, it is important for the authors to derive the IGF functions quantitatively describing how IGF varies with time, groundwater levels and abstractions, ….</td><td>The reviewer has made an important point, however we would argue that we provide a broad summary of these influences in quantifiable terms, rather than being wholly qualitative. We have not dealt with the detail, as that is beyond the inference possible with the methodological approach we have taken. Our overall aim was to identify if water balances at the reach scale are anomalous, not by how much in specific directions of flow, as we cannot make that inference from the information we have. We wanted to produce a high-level perceptual model that could be used by a hydrologist to focus model development flexibly, rather than provide absolute threshold limits for subsurface fluxes. In addition, our human influences data is highly uncertain and we did not feel comfortable reporting exact values of IGF derived from them – water balance analysis can only take us so far.

We have made edits to the introduction and analysis sections in order to clearly state our aims and objectives regarding identifying (and not explicitly quantifying) IGF, in order to better manage the expectations of the reader</td></tr>
</table>

		and justify our methodological approach, including the limitations of defining exact amounts.
3.	The water balance equations (1)-(4) adopted are also not rigorous as discussed by the authors themselves in Section 6.3 that input data uncertainties can lead to large computational uncertainty. In fact, equations (1) or (3) represent multiyear water balance instead of single year water balance. So dS/dt=0 is not strictly true, and a empirical 100 mm/yr was used by the authors to help to identify the non-conservative reach water balance. As the IGF fluxes could not be measured directly in catchment scale, empirical estimation is inevitable. However, the fudge factors e.g., 100 mm/yr as well as the physical meaning of S (groundwater storage or soil water storage?) should be discussed in depth.	As per the wider IGF literature (e.g. Le Moine 2007, Bouaziz 2018, Fan 2019), S has been used by us as a general term to incorporate all storage in a catchment, i.e. groundwater storage, soil storage, vegetation storage etc. Over the long-term the change in storage in a catchment can indeed be considered to be negligible. For this reason we have not tried to quantity IGF at the intra-annual scale. Our water balance calculations based on equations (1) and (3) are indeed at the annual scale and have been used by many other authors. We have added some text on the limitations of assuming the change in storage to be negligible at the annual timescale in the Methods section. Regarding the uncertainties in input data, we will present an uncertainty evaluation of the P, Q and AET time series estimates. This will use a simple error model to generate multiple time series for an example catchment as per Lloyd et al. 2016 and calculate the resultant water balance uncertainty range. From this we will be able to state more categorically why our thresholds (i.e. 100 mm/yr) for considering a water balance anomalous, and thus attributed to IGF type processes, can be stated. We have also included further discussion on our decision-making process regarding the threshold value in the discussion section. Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H. and Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows, Hydrol Earth Syst Sc, 22, 6415-6434, https://doi.org/10.5194/hess-22-6415-2018, 2018. Fan, Y.: Are catchments leaky?, Wires Water, 6, https://doi.org/10.1002/wat2.1386, 2019. Le Moine, N., Andreassian, V., Perrin, C. and Michel, C.: How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, Water Resour Res, 43, https://doi.org/10.1029/2006wr005608 Lloyd, C.E.M., et al., Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams. Hydrological Processes, 2016. 30(1): p. 135-152. DOI: 10.1002/hyp.10574

4.	The quality of many figures could be further improved e.g., to fully show the reach units are subdivided and to accompany their figures tightly with the text words to upgrade the readability. The reviewer suggests that reach units can be subdivided into two categories, the headwater reach and the internal reach. The water balance of reach units from the headwater areas which is recharged singly by precipitation in conservative catchments should be highlighted in order to identify the leakage recharge from outside catchment.	Reaches have been subdivided into headwater and non-headwater (internal) reaches in Fig 6 and its accompanying text. We also reflect on the importance of whether a reach is a headwater or not in Fig 2 and its accompanying text.
5.	P4L120: Here annual average precipitation for the whole basin and its spatiotemporal distribution is needed. Discharge volume of main gauges also should be provided.	We have added text on catchment rainfall, PET and discharge in the Study Area section and added a figure in our new Supplementary Information document showing the spatial variation in catchment P, PET and discharge.
6.	P6L144-146: "reach as the catchment area between river gauging stations. The analysis undertaken in this study is developed at the river reach scale rather than at the sub-catchment scale". However, the title of this manuscript is "…perceptualizing inter-catchment groundwater flow…". What is the difference between reach drainage area and sub-catchment?	We have defined a reach as "the catchment area between river gauging stations". The drainage area of a sub-catchment is therefore much larger – incorporating all upstream reaches. We acknowledge that we use the term "inter*catchment* groundwater flow" rather than "inter*reach*", but feel that we should continue using the well known term despite this discrepancy.
7.	Are the units presented in Figure 1c the reach units? I suggest that the authors provide reach units distribution map in terms of the river gauging stations.	The units presented in Figure 1 (and also now Figure S1) are catchment units, not reach units. We have edited the figure caption to stress this. Reach calculations are a part of our analysis. We have used Figure 1 to show catchment information and data that is freely available from external sources to provide a general background to the study area. We feel that the separation between catchment data in the general background section and reach data in the results and discussions sections mirrors the development of our analysis.
8.	P7L175: Provide the cells adopted for water balance computations.	On the assumption that the reviewer is referring to "wells" rather than "cells":

		We show the location, distribution and average groundwater level of the 151 wells that we use to develop our monthly groundwater level profiles and support our groundwater flow direction reviews in Figure 3a.
9.	P8L175: "A limiting factor of 70% of the total reach area was assigned as an indicator of reach coverage." What is the meaning of 70% here.	We have reworded the sentence to improve clarity for the reader.
10.	P8L217-219: S represents many storage components, e.g., groundwater storage, soil water storage, vegetation water storage, etc. how do they calculate groundwater exchanges without eliminations of other terms. It is possibly due to this reason I guess that an empirical factor 100mm/yr was adopted (see in Lines 573-576), which helps to filter disturbance from other terms? In addition, equation (1) or (3) can represent multiyear water balance instead of single year water balance. So the authors should explain the limits of using these equations.	We have added text to the Uncertainties section in the Discussion to explicitly reference the limits of using equations 1 and 3 and assuming the change in storage to be negligible. See our response to comment #2 regarding the selection of the 100 mm/yr empirical factor.
11.	P12L315-316: "Due to the high storage (Table 1)". In table 1 lower greensand aquifers are with the lowest average (0.005) storage coefficients? Why you claimed the high storage in the main text? Similar expressions can be seen also in P11L290, P21L479.	We have made edits to the text in Section 4 to ensure our descriptions regarding hydrogeological physical properties of the aquifers are valid. This includes corrections to descriptions of storage.
12.	P13L340-341: "The three lowest main river reaches show particularly large naturalised water balance losses (>1000 mm yr-1)". I noticed that the average annual precipitation of Thames basin is only about 710 mm (Gabriel et al., 2022). Why so much losses of water (>1000 mm yr-1) in the main reaches in the River Thames?	The unusual water balance results stem from the combination of a number of different factors, all highlighting the challenges when undertaking such an analysis. Firstly, we are calculating water balance at the reach, not catchment, scale. Significant differences between the topographical surface water catchment and the underlying groundwater catchment are exacerbated when discretising datasets based on topographic boundaries. In addition, the uncertainties associated with the location of, and scale of, human influences are considerable when assigning reach-scale impacts. We discuss how the >1000 mm/yr results in the Lower Thames are likely as a result of surface water abstraction and discharges in section 5.1, directing the reader to the more detailed discussion on the topic of naturalisation in section 6.3.

13.	Do you have the losses averaged over the reach units? In P14L349-357, other values about water losses or gains seems to be regular. However, I don't understand how do you convert the water losses into water depth. I suggest that the authors may use water losses volume in m3 yr-1 instead of water depth since the reference reach unit area is quite difference and upstream inflow is also different from up to down river reaches.	The water balance variables of reach P, reach AET and reach Q are averaged over the reach area to obtain values in mm/year. We purposefully chose to report our water balance analysis in depth (mm/year) rather than volume for both consistency across different climatological and hydrological variables and to account for the differences in reach areas. We feel that this remains the best choice.
14.	P14L362: what is the raio of 622 mm yr-1 annual loss in total volume of precipitation in the Kennet headwater reaches. As we know, headwater reaches do not receive upland surface inflow, so the net loss of 622 mm is large compared to the annual precipitation 710 mm over the whole basin.	Some of the non-conservatisms we find on aquifer outcrop areas can be significant, and highlight the impact that regional groundwater systems (i.e. those operating across reach topographical catchments) can have on river reach water balances. However, 622 mm/yr was a sum of all the Kennet headwater reaches' non-conservatisms. 710 mm/yr is a whole Kennet catchment average for rainfall, so the two values are not comparable. The average non-conservatism of the Chalk headwater reaches in the Kennet is 125 mm/yr and, in comparison to the average rainfall in those reaches, is actually 15% of precipitation. We have amended the text to report average losses and added reference to its ratio to the average precipitation. The use of a cumulative mm/yr headwater reaches loss was erroneous and we thank the reviewer for drawing this to our attention.
15.	P19: I suggest that the total amount of groundwater exchange should be marked in Figure 8. And how do you judge the flux directions? From the method in Section 3, I do not find related algorithm for estimating the flux directions.	As we discuss in our response to comment number 2, we made the research decision not to explicitly quantify IGF as our overall aim was to identify where water is moving that is not controlled by topography, not by how much. In 3.1.2 we state that our groundwater level data were used to confirm groundwater flow directions We have added text to Section 3 to further explain our method for estimating flux directions, whereby it is based on comparison of water balance results against our groundwater level data and published groundwater level contours.
16.	P20L454-455. "The Chalk of the Thames Basin can be locally sub-karstic, but fracture and fissure flow remain the primary groundwater flow". It maybe true as you claimed, however, if the IGFs should occur in the relatively less passageways of karstic conduits?	There is one case in the literature of IGF between topographical catchments occurring in the Chalk of the Thames via sub-karstic flow (see Maurice et al. 2022) but fracture flow is by far the dominant flow process.

17.	P21L503. It is true that not including IGF as a model flux will result in many models overestimating river flows or actual evapotranspiration. But the key question may be to describe how groundwater between neighbor units exchanges according to different conditions of groundwater levels, different lithology, human abstractions and so on.	This would indeed be interesting, and necessary, further work but at this stage we feel this is outside of our scope.