Final response

Referee comment on "Calibration of groundwater seepage on the spatial distribution of the stream network to assess catchment-scale hydraulic conductivity" by Ronan Abhervé et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-175-RC1, 2022

Referee comments are shown in black*. Our responses in blue italic.*

RC2: 'Comment on hess-2022-175', Ciaran Harman #2, 25 Oct 2022

The authors present a methodology for estimating a catchment-scale effective hydraulic conductivity by matching the surface expression of a 3D groundwater model and an estimate of the stream network.

This is a valuable contribution, and very interesting. I expect there will be more work in this area in the coming years. I have one major issue that I believe the authors should address, and only a few minor concerns. My recommendation of major revisions reflects only the one major issue I mention below. Otherwise only minor revisions are needed.

Answer:
We are thankful to the reviewer for the summary of the outcomes of our work and the positive assessment of the manuscript. The comments and suggestions raised by the reviewer were helpful to improve the manuscript.

Major issue

- I strongly recommend that the paper be changed to characterize the approach as providing estimates of the transmissivity, not the conductivity. I realize the authors may not welcome this suggestion and will likely be tempted to argue that the difference is trivial and not worth the effort to modify so many figures and text. I would urge them to consider the recommendation seriously though.

· Firstly, the transmissivity is the physical property that most controls the surface drainage expression, not the conductivity. The results of Litwin et al 2022 (see in particular figure 4, and note that gamma is a dimensionless transmissivity, and Hi is better understood as related to K/R) show that under geomorphic equilibrium the drainage density is most closely related to the transmissivity, not merely the conductivity. The drainage network appears where the groundwater table needed to transmit water downslope just reaches the surface, so it is the depth-integrated conductivity that matters, i.e. the transmissivity. Because of this, the drainage network contains information about the transmissivity, not about the underlying conductivity.

· At present, the thickness is set to 30m in the model, which seems arbitrary and unjustified given the likely variability in permeable thickness across the region. One might reasonably ask whether changing this (to, say, 20m) would affect the estimates of K? It surely would. In fact, it would almost certainly result in an increase in the estimates of K by just enough so that the product of K and 20m would be the same as in the case where the depth was 30m (i.e. K would be about 50% larger for the 20 m case than it was for the 30 m case). In other words, I would guess that there is a roughly linear sensitivity of K to the choice of thickness.

· Such sensitivity will be an impediment to efforts to understand the results and make comparisons between studies. Conductivities from different studies will not be comparable if they are based on models with different assumed thicknesses.

· However, if the authors reported that their model estimates transmissivity, they would likely find that those estimates are not so sensitive to the choice of thickness. Varying the thickness will likely not change the estimated transmissivity nearly as much. This will make the results more robust, and easier to compare with other future studies.

· The approach presented here is important and will probably be taken up by others in the future. It would be in the long-term best interests of the discipline that this important issue be clarified early on. I urge them to consider it.

Answer:

We fully agree with the reviewer. The spatial distribution of groundwater seepage is indeed a function of the transmissivity (T) as shown by previous studies (Bresciani et al., 2014; Litwin et al., 2022; Luijendijk, 2021). This means that the seepage extent area varies not only with the ratio K/R but is also sensitive to the saturated aquifer thickness (d) and the characteristic hillslope geometry (slope, length). On a technical side, the hydraulic conductivity (K) and the total aquifer thickness appears to be the input parameters for the model while the transmissivity is actually a model result as it includes the modelled saturated thickness. Because the saturated thickness is highly variable across the modelled domain, estimating the actual transmissivity in this 3D and unconfined contexts is challenging.

We extensively analyzed the relationship between transmissivity, considering the saturated thickness bellow the stream (maximum thickness), and the actual hydraulic conductivity (see Figure RC2 1). In line with the reviewer comments, we found that for thick-enough aquifers, estimations of T remain stable independently of the imposed thickness (d) (Dupuit-Forchheimer assumption, in this case for d ≥ 30m). However, when the saturation thickness is significantly reduced, estimation of T increases as expected. These results show that by considering thick-enough aquifers, the conversion from T to K by assuming an actual aquifer thickness is valid. In our case, we have considered a total aquifer thickness of 30 meters as it is observed from extensive field analysis performed in the studied region (Dewandel et al., 2006; Kolbe et al., 2016; Mougin et al., 2008; Roques et al., 2016).

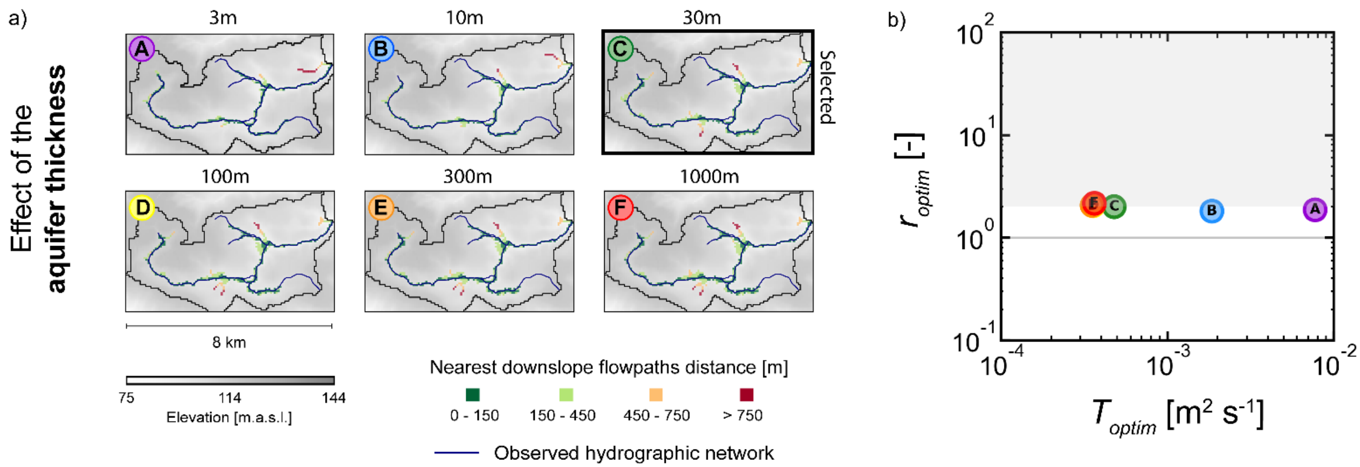

Figure RC2 1. Sensitivity analysis of the method for the Canut catchment (Site 6) to the aquifer thickness. a) Mapping the downslope flowpaths distances of the simulated seepage areas projected onto the reference hydrographic network for K_{optim}. Graphs b) show the r_{optim} as a function of T_{optim}. The optimal transmissivity T_{optim} is obtained by considering the specific thickness of the aquifer applied to the model: 3, 10, 30, 100, 300 and 1000 m.

In order to cover these challenges in the manuscript, we propose to add dedicated result sections in which we will assess the relationship between K, aquifer thickness and transmissivity (Figure RC2 1). We also propose to present both estimates of T and K for the ensemble of catchments as pictured in the Figure RC2 2. This will enable to compare the estimated hydraulic properties with other values obtained independently by other methodologies that are classically given in hydraulic conductivities. This is also in line with the recommendations of the first referee. We also state in the figure legends and results section that estimations of K and their comparison with other methods are conditioned to the assumption of a 30 m aquifer thickness. We will provide a detailed description of the validity of the approach to estimate K from T in the discussion (see comment from referee RC1).

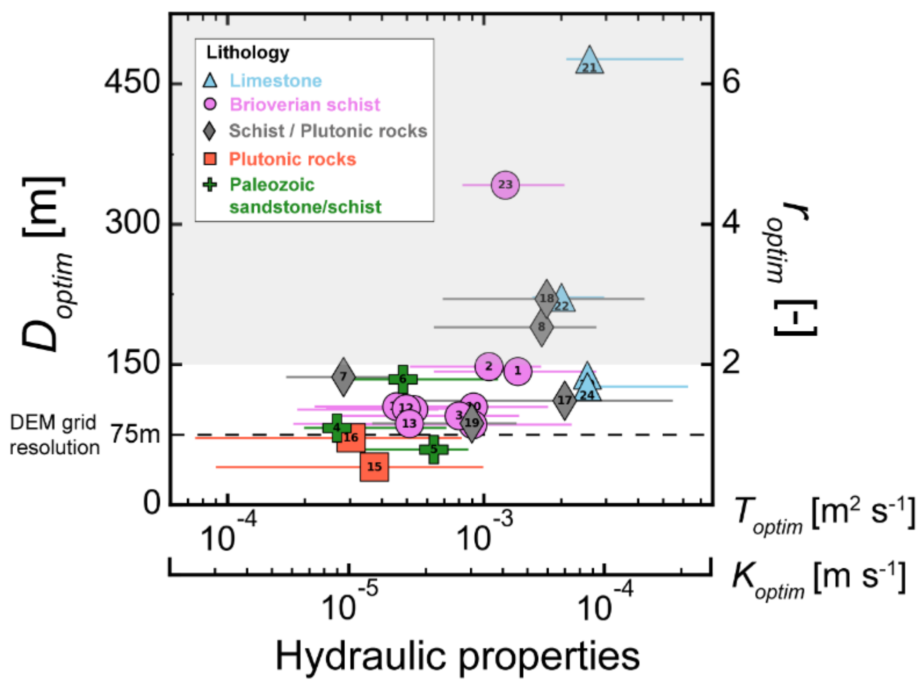

Figure RC2 2. D_{optim} and r_{optim} criteria as a function of K_{optim} estimated for the 24 sites. The optimal transmissivity T_{optim} is obtained by considering the aquifer thickness of 30m applied for each model. The shaded area corresponds to sites with $r_{optim} > 2$. The DEM resolution is 75 m and the aquifer thickness is 30 m. The error bars correspond to the estimated K_{optim} considering the DEM resolution as an uncertainty indicator.

Minor issues

- It isn't clear to me why the recommendation of using lower resolution DEMs is justified. I understand that registration errors are enhanced with a high resolution DEM, but that doesn't seem like a problem that arises from the DEM itself. Wouldn't it be better to create a buffer around the stream network to account for uncertainty in its location?
Furthermore, it is not clear why r_optim should be less than or equal to 2 "considering that the mismatch cannot exceed the resolution of two pixels" (Line 210). I don't follow the reasoning, and the restriction is clearly violated in the results. What is the purpose of normalizing by pixel size? Doing so will always result in a larger 'error' for high resolution DEMs. The larger values of r_optim for the 5m and 25m don't reveal a clear deficiency to me.

Answer:
Thank you for raising these very important points. We will modify the discussion accordingly.

We agree that the deficiency noted with the r_{optim} criterion for the higher resolution 5 m and 25 m DEMs is not caused from the DEMs themselves. The criterion is not the issue. It rather comes from the mismatch between the too high resolution of the DEM compared to the resolution of the observed stream network. In this line, we will put in perspective the consideration of uncertainties in the location of the stream network (buffer as suggested), considering that this error is known for a given stream product (function of the DEM used to map/generate the stream network).

The D_{optim} criterion, initially based on distances, is normalized by the pixel size, in r_{optim}, in order to compare results based on different DEM resolutions. The choice to consider that the mismatch should not exceed 2 pixels was driven by analyzing the results obtained for the 24 catchments studied. We will provide a more detailed discussion on the sources of errors emerging from the resolution of the DEM and the observed reference hydrographic network (refer to next question/answer).

- The discussion of the errors from line 246 on is important and ought to be expanded. Consider:

- The assertion that the differences "come essentially from the data" is confusing. It also seems inaccurate, since some of the errors are due to deficiencies in the model (where the assumption of a uniform K strays too far from reality, such as Site 8) and some are due to deficiencies in the data (where the mapped stream network does not adequately capture a more complex reality, such as site 18), and in some cases it is not immediately clear which is in error (Site 23 -- is the true stream in fact offset from the lowest point in the topography, or is there a registration error in the alignment of the stream location data with the DEM data?).

Answer:
These very technical and precise comments are clearly in line with our reflections on the subject. We propose to add discussion and clarification to each of these points. We recognize that the formulation "the differences come essentially from the data themselves rather than from the model" [Line 247] is confusing and should be clarified for each sites presented.

Since the model simulates the seepage zones at the origin of the streams in the valley bottoms of a given DEM, the errors identified may come from:

- *Incomplete observed stream network (simulated but not observed, for example the presence of a water body not represented on the observed stream layer as Site 18) or "unnatural" sections (observed but not simulated, for example with a true stream in fact offset from the lowest point in the topography)*
- *Spatial mismatch between the observed stream network and the DEM valley bottoms (for example, registration error in the alignment of the stream location data with the DEM data as Site 23), depending mainly on the resolution of the products.*
- *Model deficiencies, especially considering a homogeneous aquifer. For example, in Figure 6 of the manuscript, we identify the effect of this lateral heterogeneity for Site 8. On this site, we*

found that in the northern part of the catchment our simulation results fail in modelling the observed stream network on the granite lithology, while it overestimates its extent on the schist/sandstone lithology. This suggests that the obtained bulk hydraulic conductivity is slightly overestimated for the granite-dominated part of the catchment while being underestimated for the schist part.

These previous points will be discussed and added to the discussion, as also requested by the first reviewer.

· In the case of site 21 and 22 does 'non-reported subsurface flow' refer to a karst conduit (i.e. a channel, but underground) that is known to exist (i.e. data exists showing that it is there)? Or could the down-valley flow be through porous media with a higher K than other areas? One might argue that these are quite different sorts of errors. If the former, it suggests that improved accuracy would come from including conduit flow in the observed stream networks. If the latter, it suggests improved accuracy would come from allowing for spatial variability in the K-field of the model. It may not be possible to distinguish between these in practice.

Answer:
For sites 21 and 22, the more accurate observed stream mapping, including all streams, i.e. perennial and intermittent, shows a discontinuous streamline at the identified error areas (white polygon in the Figure 6 of the manuscript). Indeed, a short section of intermittent stream is displayed in the database in the upstream part of the white square, before disappearing downstream. This suggests the existence of a subsurface flow in a potential conduit, consistent with the limestone geological formation and the context studied.

In this case, we are able to identify the source of error due to the stream network used. Nevertheless, as the reviewer highlight, it could be linked to the lithological heterogeneity. In this sense, we confirm that the perspectives of the method should consider the spatial variability in K across the catchments. As mentioned by reviewer, it is possible to include heterogeneity and apply the method independently of lithologies. This point will be discussed in the future version of the manuscript.

- It might be worth discussing a geomorphic justification for the approach. For example, the ideal stream network data is mapped extent of flowing streams, not merely the existence of a blue line on a topographic map, which are often based on purely geomorphic criteria (e.g. minimum upstream area). Further errors would occur where the surface drainage is not in geomorphic equilibrium (e.g. where relic channels remain from historical periods of erosion following deforestation, but they do not carry flow today). The authors should caution that incorrect results would arise from using stream network data that does not actually represent the extent of the flowing stream.

Answer:
We share this point of view on the data qualified as "observed hydrographic network". We have used the most standard stream network layer available at the scale of France, and it does not represent

exactly the mapping of the flowing streams. The definition of a flowing stream is not even standardized in such database at the national level. Thus, it may indeed be that some parts of the drainage surface that are not in geomorphic equilibrium are source of errors. The example mentioned by the reviewer will be included in the discussion in the new version of the manuscript.

In our opinion, to clarify this point, instead of listing the potential errors of the hydrographic network that has been exploited in this study, it would be relevant to alert the readers to the major interest to use, if available, a hydrographic network that represents best observed/mapped flow extension. This supports the importance of considering a DEM and observed hydrographic network with similar spatial resolution.

- Can the authors provide better justification for the performance criterion? Why this and not something else?

Answer:
Following this comment and the ones from RC1, we provide a more detailed description in line 313 and 317 of the manuscript, for example as following:

"The method calibrates the dimensionless parameter K/R by matching the modelled groundwater seepage zones to the observed stream network minimizing their respective distances. The nearest downslope flowpath distances (D_{so} and D_{os}) improve the Euclidean distance of the cell-by-cell and cell by-neighbourhood analysis (Franks et al., 1998; Güntner et al., 2004) by constraining the observed-to-simulated and simulated-to-observed stream networks to the topographical structures (Mardhel et al., 2021). the performance criterion proposed allows for a compromise between over- and under-estimation."

- Does the groundwater model include overland flow and reinfiltration? I.e. losing and gaining reaches?

Answer:
Our groundwater flow model does not include overland flow and reinfiltration, i.e. the stream reaches are only gaining. This model limitation will be clearly specified in the new version of the manuscript.

- It is confusing that the description of the recharge rate estimation is included in section 2.1.5 and not in section 2.1.2 (where the estimated recharge rate is presumably used, unless I misunderstand)

Answer: *These equations are moved to the recommended section.*

- A uniform recharge rate is used -- this limitation of the method should be acknowledged. We know that riparian areas receive more recharge than uplands, and that evapotranspiration can be drawn from the groundwater.

Answer:
This limitation is added in the new version of manuscript while emphasizing that it is possible to apply a heterogeneous recharge to the model.

"Just as the application of spatial variability in the K-field as a function of lithology is possible in the model, heterogeneous recharge can also be applied at the surface. This could be achieved by coupling the current simplified subsurface flow-based approach with a partitioning of precipitation into recharge from evapotranspiration and runoff or using more complex models coupling surface and subsurface."

- Line 203-204: I think D_os and D_so are switched here

Answer: *Corrected.*

- Figure 5: please expand the figure so that the horizontal error bars are not cut off.

Answer: *Modified.*

References of the final response document

Bresciani, E., Davy, P., & De Dreuzy, J. R. (2014). Is the Dupuit assumption suitable for predicting the groundwater seepage area in hillslopes? *Water Resources Research*, *50*(3), 2394–2406. https://doi.org/10.1002/2013WR014284

Dewandel, B., Lachassagne, P., Wyns, R., Maréchal, J. C., & Krishnamurthy, N. S. (2006). A generalized 3-D geological and hydrogeological conceptual model of granite aquifers controlled by single or multiphase weathering. *Journal of Hydrology*, *330*(1–2), 260–284. https://doi.org/10.1016/j.jhydrol.2006.03.026

Franks, S. W., Gineste, P., Beven, K. J., & Merot, P. (1998). On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research*, *34*(4), 787–797. https://doi.org/10.1029/97WR03041

Güntner, A., Seibert, J., & Uhlenbrook, S. (2004). Modeling spatial patterns of saturated areas: An evaluation of different terrain indices. *Water Resources Research*, *40*(5), 1–19. https://doi.org/10.1029/2003WR002864

Kolbe, T., Marçais, J., Thomas, Z., Abbott, B. W., de Dreuzy, J.-R., Rousseau-Gueutin, P., Aquilina, L., Labasque, T., & Pinay, G. (2016). Coupling 3D groundwater modeling with CFC-based age dating to classify local groundwater circulation in an unconfined crystalline aquifer. *Journal of Hydrology*, *543*, 31–46. https://doi.org/10.1016/j.jhydrol.2016.05.020

Litwin, D. G., Tucker, G. E., Barnhart, K. R., & Harman, C. J. (2022). Groundwater Affects the Geomorphic and Hydrologic Properties of Coevolved Landscapes. *Journal of Geophysical Research: Earth Surface*, *127*(1), 1–36. https://doi.org/10.1029/2021JF006239

Luijendijk, E. (2021). *Transmissivity and groundwater flow exert a strong influence on drainage density*. *April*, 1–32.

Mardhel, V., Pinson, S., & Allier, D. (2021). Description of an indirect method (IDPR) to determine spatial distribution of infiltration and runoff and its hydrogeological applications to the French territory. *Journal of Hydrology*, *592*. https://doi.org/10.1016/j.jhydrol.2020.125609

Mougin, B., Allier, D., Blanchin, R., Carn, A., Courtois, N., Gateau, C., & Putot, E. (2008). *SILURES Bretagne (Système d'Information pour la Localisation et l'Utilisation des Ressources en Eaux Souterraines)*. https://infoterre.brgm.fr/rapports/RP-56457-FR.pdf

Roques, C., Bour, O., Aquilina, L., & Dewandel, B. (2016). High-yielding aquifers in crystalline basement: insights about the role of fault zones, exemplified by Armorican Massif, France. *Hydrogeology Journal*, *24*(8), 2157–2170. https://doi.org/10.1007/s10040-016-1451-6