Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-436-RC2, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "On the assimilation of environmental tracer observations for model-based decision support" *by* Matthew J. Knowling et al.

Anonymous Referee #2

Received and published: 13 November 2019

In the present manuscript, Knowling et al. aim to demonstrate that environmental tracer observations in general are not as informative for groundwater model data assimilation as previously thought because, in their eyes, flow models are typically too wrong for adequate physical representation of tracer behavior. The authors base their conclusions on only two case studies involving groundwater model calibration against only one environmental tracer (i.e. tritium, in one case study using tritium-derived groundwater residence times and in a second case study using tritium concentrations directly). The authors specifically identify errors in groundwater model vertical discretization as a reason for why data assimilation of groundwater model with tritium concentrations is prone to result in biased model predictions.

While a systematic study on this topic is potentially interesting and useful, the present study lacks the necessary rigor in experimental design and standard in scientific reporting to be able to demonstrate what the study aims to demonstrate and to be a valid contribution to HESS.

Shortcomings include: Failure to properly describe (1) the model calibration procedures, (2) the observation data, and (3) the models and assumptions used to derive residence times from tritium concentrations. The authors also fail in properly referencing scientific literature which already demonstrated aspects of the present study. Moreover, misleading statements are made about existing studies, and the general conclusions that were drawn on the value of environmental tracer observations for groundwater model calibration in general are not justified from the results of the simple experimental setup and use of tritium alone. Due to the lack in reporting, it isn't even possible to fully understand, assess or reproduce the findings.

Below I elaborate on some of the shortcomings of the study which I see as reasons for rejecting of this paper.

——— The manuscript lacks key information on model calibration:

The present manuscript doesn't sufficiently explain the observation data, models which were used to derive the different observation types or calibration procedures.

In the first case study, the value of observations of tritium-derived groundwater residence times are compared to the value of groundwater levels and spring discharge observations for the reduction of the predictive uncertainty of spring discharge predictions. However, information about the calibration procedure is not provided, i.e., it isn't clear whether an ensemble-based data assimilation procedure (i.e., the iterative ensemble smoother as mentioned in the abstract), or whether a classic history matching calibration procedure (i.e., based on a weighted, multivariate maximum likelihood estimation procedure as described in a referenced modelling report) is used. Even though in the abstract it is stated that iterative ensemble smoother was used in the present study, the method isn't explained in the methods section of model study 1.

One can either assume that it was the same as for model study 2, i.e. Iterative Ensemble Smoother. This is suggested by the wording of the abstract and the term 'data assimilation via history matching' (line 117). An Iterative Ensemble Smoother approach, and ensemble-based data assimilation procedures in general, would however make the direct application of linear predictive uncertainty analysis based on FOSM impossible because to the jacobi matrix isn't calculated by these approaches. Or, one could assume that data assimilation was not conducted but instead classic history matching after reading a referenced modelling report (however, Rakowski and Knowling, 2018, is not referenced in the respective model calibration and uncertainty quantification methodology section (2.4)). Using classic history matching would be a contrast to what was stated in the abstract and make the data worth assessment difficult to compare to the findings of modelling case study 2. The authors should also explain in detail what they mean by how the jacobi matrix was populated.

For the second modelling case study, in section 3 after the description of methods and results of model study 1, it is explained that an Iterative Ensemble Smoother with 100 realisations was used. While for model study 1 it was stated that 882 parameters were calibrated, for model study 2 one does not learn how many parameters were calibrated. While for model study 1 there is a referenced modelling report available, the report referenced for model study 2 was not accepted or published at the time of the article submission and therefore not available for checking (on lines 184-185 it is stated: 'The model, and the vertical-discretization simplification analysis, is described in detail in White et al. (forthcoming)' and the said study is listed in the bibliography as 'accepted, subject to minor revisions').

Key information in the calibration procedure is essential when the purpose of the study is to demonstrate the value of different observation types, as the calibration procedure strongly influence the data worth results.

——— The manuscript lacks key information about the used observation data:

Observation data which were used for the modelling study are not provided, even though this is critical information to understand and reproduce the reported findings.

While for model study 1 at least the different observation types which were used are mentioned, for model study 2 it is completely unclear what observations were used alongside tritium. It isn't clear how many observations of tritium, what uncertainty these observations are associated with, and the study which probably contains such information was not accepted at the time of submission and is not available.

Key questions that should be addressed before data worth can be objectively assessed are: What data were used alongside tritium? Is tritium an informative tracer for each of the two given systems, i.e., is the groundwater residence time in both catchments sensitive to tritium? How was tritium analyzed and which equations were used to post-process tritium concentrations into groundwater residence times? How were flux measurements obtained? What is the uncertainty of spring discharge observations? Are the uncertainties comparable to tritium-based residence time uncertainties? What are the weights that were used during calibration and do they reflect the uncertainty of the different observations? None of this is described in the manuscript.

This information is needed for the readers to assess whether the results of the present study are correct and meaningful.

——— The relevance of the authors' findings is over-stated:

It is unclear why it is concluded that tritium is representative of environmental tracers in general. The manuscript lacks an important number of references which have already published similar results on the value of spring discharge or tritium or which have shown, in much more systematic and rigorous experimental approaches, that environmental tracers are highly valuable for groundwater model calibration.

While the title is very broad, i.e.: 'On the assimilation of environmental tracer observations for model-based decision support', the present study does not generally assess the value of environmental tracer data in a data assimilation context. It appears as if only for one of the two modelling studies formal data assimilation has been conducted (however, as outlined in the previous comment, it is not entirely clear what calibration approach was used in the first modelling case study). Furthermore, only one single environmental tracer is used: tritium. Tritium is certainly not reflective of all environmental tracers and for many groundwater systems, tritium is not a useful tracer because groundwater residence times are of an order on which tritium isn't sensitive. The wording of abstract, introduction, discussion and conclusions strongly suggests that the authors believe that their two case studies of tritium are representative of the wider worth of environmental tracer data for groundwater model calibration (e.g., Lines 268-271) : 'We consider this recommendation to be in stark contrast to the common belief that "calibrating to more data improves the model and its predictions". We therefore also consider this recommendation to be of significant implication to decision-support environmental modeling practitioners. It is expected that this finding can be extended to the general approach of assimilating diverse observation types in environmental modeling.'

Tritium is not representative of environmental tracers in general, as it requires more complex mathematical simulation procedures to do its complex decay and production pathways justice. Showing that a simple one-layer model cannot properly represent tritium transport and therefore calibrating it against tritium results in biased predictions is not generating insights representative for environmental tracer value in general. Numerous previous studies have much more systematically analysed and identified the large benefits of environmental tracers for groundwater model calibration in general, but the large majority are not referenced in the present manuscript. Here are a few examples:

Carniato et al. (2015), Highly parameterized inversion of groundwater reactive transport for a complex field site. DOI: 10.1016/j.jconhyd.2014.12.001.

Delsmann et al. (2016), Global sampling to assess the value of diverse observations in conditioning a real-world groundwater flow and transport model. DOI: 10.1002/2014WR016476

Hunt et al. (2006), The importance of diverse data types to calibrate a watershed model of the Trout Lake Basin, Northern Wisconsin, USA. DOI: 10.1016/j.jhydrol.2005.08.005 (cited in the present manuscript)

Rasa et al. (2013), Effect of different transport observations on inverse modeling results: case study of a long-term groundwater tracer test monitored at high resolution. DOI: 10.1007/s10040-013-1026-8 (cited in the present manuscript)

Xu and Gomez-Hernandez (2016): Characterization of non-Gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures. DOI: 10.1002/2016WR019011

Oehlmann et al. (2015), Reducing the ambiguity of karst aquifer models by pattern matching of flow and transport on catchment scale. DOI: 10.5194/hess-19-893-2015

Masbruch et al. (2014), Hydrology and numerical simulation of groundwater movement and heat transport in Snake Valley and surrounding areas, Juab, Millard, and Beaver Counties, Utah, and White Pine and Lincoln Counties, Nevada. DOI: 10.3133/sir20145103

What is demonstrated in the first modeling case study, i.e., the complicated nature of using residence/travel time observations derived from tritium for groundwater model calibration, is very well known and was already subject of multiple much more systematic and thorough comparisons and reviews, some of which are even referenced in the present manuscript (e.g., Turnadge and Smerdon 2014 (DOI: 10.1016/j.jhydrol.2014.10.056), McCallum et al. 2014 (DOI: 10.1111/gwat.12052) and 2015 (DOI: doi:10.1111/gwat.12237), Schilling et al. 2019 (DOI: 10.1029/2018RG000619), Sanford 2011 (DOI: 10.1007/s10040-010-0637-6)).

All these studies concluded already that it is better to calibrate a flow model against environmental tracer concentrations, or yet even better, direct flux observations, rather than against residence times due to the fact that the simulation of residence times is often faulty due to structural inaccuracies in the numerical groundwater model.

Specifically, the fact that spring discharge observations contain the largest amount of information for spring discharge predictions is neither surprising nor new. Exchange fluxes in general, be it groundwater discharging as spring water or into a surface water body, or surface water infiltrating into the subsurface, have been demonstrated to not only be more valuable data for groundwater model calibration than travel/residence times observations, but also to be much less prone to bias due to straightforward implementation into flow model calibration compared to the more complex physical underpinnings required for groundwater residence times simulations. The authors even reference one study which has demonstrated this systematically in comparison to groundwater residence time observations: Hunt et al. (2006, DOI: 10.1016/j.jhydrol.2005.08.005, already cited in the manuscript) compared the worth of several different flux observations to the worth of hydraulic heads, environmental tracer concentrations and travel time information, and found that groundwater exfiltration onto the surface (providing baseflow of a stream) was the most information rich data type overall, and that many other flux observation types were also more informative than travel time observations.

The authors failed to reference studies which have already demonstrated the high importance of spring discharge more specifically: Masbruch et al. (2014, DOI: 10.3133/sir20145103, not cited in the manuscript) systematically compared the information content of spring discharge to observations of groundwater levels, temperature and environmental tracers, and found that spring discharge observations were the most informative overall data type. A similarly high importance of spring discharge observations was identified by La Vigna et al. (2006, DOI: 10.1007/s10040-016-1393-z, not cited in the manuscript), who systematically elaborated the worth of spring discharge

observations for the calibration of groundwater flow models in comparison to hydraulic head observations. Oehlmann et al. (2015, DOI: 10.5194/hess-19-893-2015, not cited in the manuscript) systematically analysed the calibration of karst groundwater models against observations of spring discharge, groundwater residence times and groundwater levels. They identified that spring discharge observations provide indispensable information for karst groundwater model calibration, but also showed the large information content of residence time observations. The use of all three observation types together was the most beneficial approach for groundwater model parameterisation.

The authors' literature review is unbalanced, misses many key references, and makes incorrect statements about findings of key studies.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-436, 2019.