Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-436-SC6, 2019

Hydrology and Earth System Sciences
Discussions

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

Matthew Knowling

m.knowling@gns.cri.nz

Received and published: 18 November 2019

We thank the anonymous reviewer for their comments. We believe that their comments can be addressed where appropriate through some minor, yet important additions and clarifications to the manuscript text. The comments also provide us with an opportunity to revisit and expand on some of our findings and recommendations.

We first wish to clarify the following points related to the reviewer's summary of our work:

– Not only do we assess the apparent or "theoretical" information content of environ-

mental tracer observations for decision-support groundwater model data assimilation, as judged by rigorous data worth exploration, we also assess the potential for the assimilation of these data to cause unwanted effects such as forecast bias, by considering model error and using paired complex/simple models. To our knowledge, no other work has examined the assimilation of environmental tracers in the context of the bias-variance trade-off relevant to the use of imperfect models. We feel that this central aspect of our paper has been over-looked by the reviewer. Framing our findings and recommendations in the context of real-world decision-support modeling is fundamentally important to the purpose of our paper. The importance of this aspect was acknowledged by the other reviewers, e.g., "They make a great case that data, even if it inherently carries important information, can be misleading if it is viewed through the lens of an imperfect model."

– We wish to follow-up on the reviewer's comment that "in (our) eyes, flow models are typically too wrong for adequate physical representation of tracer behavior". Recent studies have showed that even seemingly minor model defects can cause significant ill-effects such as bias and uncertainty under-estimation. However, even more importantly, the outcome of these ill-effects depends on the purpose of the modeling analysis. The challenge is therefore to try to avoid these ill-effects in the context of the given modeling analysis and its purpose. We are advocating for careful and forecast-specific model design, to ensure that the rich information contained within tracer data can be properly assimilated. Potential use of more abstract means to assimilate these data into simpler models, is a promising model design option as it alleviates the need for increased model complexity and the costs associated with it. We suggest that this is an area of future work, as discussed in the Discussion and Conclusions section of the manuscript. These recommendations were explicitly valued by the other reviewers, e.g., "In this case, the group makes the point that increased complexity has a place in assimilating more and more varied data while recognizing that this increased complexity has limits for some applications.".

The reviewer's summary suggests that our conclusions are not warranted on the basis of two case studies and the consideration of one environmental tracer. We wish to point out that an exhaustive exploration of how and when environmental tracers can be most usefully assimilated into models represents a research field in itself. The purpose of our paper is to raise and illustrate the following two points: (i) the assimilation of environmental tracer observations may not always be worthwhile, depending on the forecast being made, and (ii) careful model design is central to the ability to assimilate environmental tracer observations into models. We now address each of the reviewer's comments below.

"Lacking information on model calibration"

The reviewer states that "information about the calibration procedure is not provided" for the first case study. This comment reflects a misunderstanding, and therefore a need to improve the presentation of this portion of manuscript. We employ FOSM techniques to quantitatively assess the worth of tritium-derived MRT along with other hydrologic observations by comparing forecast uncertainty changes following the notional data assimilation of different observations (e.g., see Lines 71-72, 75-78 and 128-130). The linearity assumption underpinning FOSM-based analyses means these powerful techniques can be applied to estimate the "theoretical" worth of data (see Lines 71 and 128), as measured by the change in forecast variance that would occur if uncertain parameters were conditioned using different combinations of observations. The efficiency of FOSM is recognized by assuming the action of a model from parameters to outputs can be characterized by a first-order sensitivity matrix. FOSM analyses do not rely on formal history matching or on the pre-existence of a "calibrated model". To be clear, no actual parameter estimation is undertaken as part of the first case study. FOSM techniques have been widely employed for data worth assessment in this notional context in many settings as it enables rapid exploration of the worth of many different combinations of conditional forecast variances in a computationally efficient manner (e.g., Wallis et al., 2014; Zell et al., 2018). We will address this comment by

revising Line 117 (e.g., add "notional", remove "via history matching") and by replacing the "History matching" sub-section heading with "Observations for assimilation".

We feel that these changes will address the reviewer's confusion regarding whether the iterative ensemble smoother was used for the first case study. The ensemble smoother was used only for the second case study, where we explore the potential for model simplification-induced forecast bias and how this may be exacerbated when assimilating tracer concentration observations. To further address this comment, we will make the distinction between the two different data assimilation approaches undertaken for the two case studies more explicit throughout the revised manuscript. We will also add the words "finite differences" after "Jacobian matrix was populated using" on Line 142.

The paired complex/simple model analysis undertaken for the second case study involves formal history matching and non-linear uncertainty quantification (via the iterative ensemble smoother) for various models with varying vertical discretizations. Each of these analyses are performed twice, once with and once without tritium concentration data for assimilation (i.e., Figure 4 is the result of six ensemble smoother experiments). This first-of-its-kind analysis for environmental tracer assimilation allows us to identify biases arising directly from the assimilation of these information-rich observations with imperfect, real-world groundwater models in a decision-support setting. This approach was necessary to explore otherwise invisible biases induced through assimilating tritium data. As discussed above, we feel this critical and novel part of the paper has been over-looked by the reviewer.

While the number of uncertain model parameters used in the second case study (for each of the 7-, 2- and 1-layer models) is provided in the now-published article White et al. (in press), we agree with the reviewer that this constitutes an important detail that, when absent, may obscure some details regarding the second case study. We will therefore add data assimilation details to Section 3.2 and to the Supplementary Information.

"Lacking information about observation data"

The reviewer is correct that the second case study does not contain information regarding observations used for history matching aside from tritium concentration observations. While the other hydrologic observations used for history matching are described in detail in White et al. (in press) (we will also add a summary of these other observations in response to other comments), we note this omission was purposeful in the original manuscript: the second case study does not compare the relative value of different data including tritium observations (as in the first case study). Instead, the second case study investigates an additional and equally-relevant aspect of environmental tracer assimilation concerning differences in the posterior forecast distribution in terms of first-moment (i.e., bias) and second-moment (i.e., variance) characteristics. These differences arise directly from the assimilation of tritium concentration observations using models that are progressively less equipped to assimilate this information. We will nevertheless add these details to the Supplementary Information to address this comment.

While more information regarding the tritium concentration observation data used for history matching in the second case study are presented in White et al. (in press) (and its SI), we agree with the reviewer that these details are warranted here. We will add details regarding the tritium concentration observations to the revised manuscript.

Responses to specific questions regarding data worth interpretations:

– For reasons described above, we will add details to the Supplementary Information regarding the other observations used alongside tritium for assimilation in the second case study.

– Tritium is indeed an "informative tracer" for the hydrologic settings in both case studies. We agree it is important to state this more explicitly. We will add a sentence to the end of the Introduction to address this comment.

– We agree with the reviewer that details regarding how tritium measurements were interpreted are warranted. This was also raised by Mr Chris Turnadge in his review comments—we refer the reviewer to his comments and our responses. We will add details on this to the description of the first case study (i.e., where MRT observations were used).

– Tritium concentrations are not "post-processed" into MRT for either of the case studies. To address this comment, we will add the following text to the revised manuscript for clarity: "The first case study involves the assimilation of tritium-derived MRT observations (following use of lumped-parameter exponential and binary mixing models). The second case study involves the assimilation of tritium concentration measurements through history matching simulated tritium concentration outputs."

– We will indicate in the revised manuscript the field techniques used to measure fluxes.

– For the first case study, the spring discharge measurement uncertainty is assumed to be proportional to the absolute flux magnitude (i.e., a heteroscedastic error model; e.g., Sorooshian, 1980). The assumed measurement uncertainty for MRT observations are also assumed to be proportional to the MRT magnitude, and are generally lower than those of spring discharge observations (primarily reflecting the difference in magnitudes). However, as described on Lines 144-146 and 300-301, in order to approximately account for the role of model error in reducing a model's ability to fit observations (i.e., we should not be fitting observations to a level commensurate with measurement noise given the presence of model error), we use the model-to-measurement residuals as a basis to adjust the uncertainty surrounding observations (e.g., Doherty, 2015).

– Observation weights assigned to different observations used for assimilation indeed reflect uncertainty—specifically epistemic uncertainty (accounting for both measurement and structural sources of uncertainty)—as described above. This is stated explicitly on Line 143.

"Relevance of findings are over-stated"

The purpose of this paper is to show, through two real-world decision-support models, that the assimilation of environmental tracers is not a panacea to all the ills of environmental simulation, and that care in the model design is essential to ensure that the information contained within these tracers is not squandered. While we will endeavour to make this purpose more clear in the text to avoid misinterpretation, we note that the comment that "the relevance of our findings is over-stated" is in direct contrast to the technically detailed comments by Mr Turnadge, who explicitly stated the carefulness with which our conclusions were drawn: "Having re-read the manuscript, I now believe that the authors have been careful to state that their conclusions regarding the applicability of environmental tracer observations are indicative, rather than comprehensive". We have taken care to ensure that all general statements are accompanied by appropriate caveats.

While we agree that a "systematic study on this topic is potentially interesting and useful" (the reviewer's words), a comprehensive analysis into the value or otherwise of assimilating environmental tracers, and the models required to do so, can never be systematic because it will always be context specific. This context reflects a combination of important factors, such as the decision-support quantities of interest, the hydrologic setting, the complexity (or otherwise) of the model, the other available observations, among others.

On the representativeness of tritium.

We do not "conclude" that tritium is representative of tracers in general. Instead, we consider tritium as a representative environmental tracer in the context of the potential outcomes of its assimilation into imperfect models in general. Our consideration of tritium simply reflects that it is one of the most widely used environmental tracers in younger groundwater systems, as were the subject of the two case studies.

The purpose of the paper is not to assess the representativeness of tritium relative to

other tracers. Instead, it is to demonstrate, in general terms, that the usefulness or otherwise of any information-rich data, including environmental tracers, is related to the forecasts being made, and that such data may induce (undetectable) forecast bias, when assimilated into imperfect models. We believe that these issues are relevant and transferrable to the assimilation of environmental tracers in general. This reflects that we consider the primary barrier to appropriate assimilation of tracer data to be the difficulties associated with extracting information from spatially-discrete concentration observations when using upscaled or simplified representations of hydraulic properties within a model that simulates tracer concentrations using the advection-dispersion equation. To the extent that the simulated output corresponding to observed tracer concentration(s) are sensitive to model details or parameters that are "missing" in a simplified model (e.g., White et al., 2014), inappropriate parameter compensation will occur. Then, to the extent that the forecast of management interest is dependent on these biased parameter estimates, the forecast will be biased, potentially leading to resource mismanagement. Therefore, we believe that these factors, which all real-world decision-support model analyses share, will also cause issues for tracer data assimilation and assimilation of diverse data in imperfect models more generally.

To address comments related to how our findings can be applied to other tracers, we will focus on providing more detail in the Discussion and Conclusions section. Despite this, we believe that the larger challenge for the transferability of our work is the specificity of different decision-support contexts, and the infinite spectrum of model design, as we discuss in the Discussion and Conclusions section.

On citing existing literature.

We note that the reviewer states that only "aspects" of our study have been demonstrated in other studies. We agree. However, as discussed in detail above, the aspects that the reviewer is referring to do not encapsulate the thrust of our paper.

We agree with the reviewer that numerous studies have explored the value in environmental tracer data for history matching groundwater models, and we accept that we have not cited every paper on this in our literature review. However, we do not feel that it is necessary to do so, as the literature we cite in the Introduction collectively expresses the state of the science: that studies have "identified the large benefits of environmental tracers for groundwater model calibration in general" (to use the reviewer's words).

However, our study (or at least our first case study) provides an important demonstration of how "variable" the worth of these data may actually be, e.g., in the presence of other data and when making water quantity-related forecasts. We provide a series of detailed explanations for this—we refer the reviewer to Lines 235-254 of the Discussion and Conclusions sections. We feel that such "worked examples" are important for the community to see—this perspective was strongly supported by the other reviewers, e.g., "This is really important work and I feel that HESS is just the right target audience", "this is just the sort of result that could spark important conversations in our field" and "the group takes advantage of its abilities to provide useful guidance for the community" (Prof. Ferre), and "this manuscript provides a valuable and timely contribution towards guidance in the use of subsurface environmental tracers" and "investigations of when benefits may be obtained (or perhaps more importantly, when not) from these additional data types (and therefore specific guidance in their use) has been limited" (Mr Turnadge).

We note that the references suggested by the reviewer do not tackle the entangled issues of model error, data assimilation and predictive reliability—in contrast to our study. For example, following consideration of the references suggested by the reviewer, we consider questions such as: (i) how could these studies explore the potential for forecast bias given that history matching was undertaken using only a single model?; and (ii) how can the observation data responsible for inducing forecast bias through assimilation be identified?. These questions illuminate how our study differs from previous studies. This also provides an insight into the importance of the context is which our

paper is framed—and how this differs to groundwater modeling practice in general terms. We were very careful to make this point clear, e.g., "Modeling for the purpose of decision support is the context in which the remainder of this paper is framed" (Line 29), and "The benefit or otherwise of direct assimilation of tritium concentration data in other decision contexts, or for more general system understanding and conceptual model development, is therefore not the focus of the current study—this study is concerned with a model's ability to "predict" (in two decision-support contexts) rather than "explain" (observed system behavior), as contrasted by Shmueli (2010)." (a revised version of Line 231). Nevertheless, to address this comment, we will add the references provided by the reviewer to the Introduction.

We agree with the reviewer that it has been reported in the literature that it is a preferred approach to simulate tracer concentrations (involving solution of the advective-dispersive equation) and history match to tracer concentrations directly, rather than simulate residence times (involving advective-only particle-tracking schemes) and history match to derived quantities such as MRT. We state this on Line 40-43. However, in real-world modeling, the compounding challenges associated with simulation of tracer concentrations, e.g., computational demands, can complicate the assimilation of tracers. These challenges are why the latter approach is still popular in the industry (e.g., Turnadge and Smerdon, 2014).

It is interesting that the reviewer states that model "structural inaccuracies" generally explain why residence times cannot be reliably simulated. We contend that a significant degree of "structural inaccuracy" will persist, even when dispersive and decay processes are simulated explicitly (i.e., rather than simplified advective-only simulations), as is demonstrated explicitly in the second case study. The difficulty in simulating spatially and temporally distributed tracer concentrations (or more generally, simulating advective-dispersive transport) has been discussed by many (e.g., Zheng and Gorelick, 2003; Riva et al., 2008). Our second case study demonstrates how discretization-related model error combined with assimilation of discrete-point concentration observations can induce considerable biases and ultimately corrupt resource management.

We agree that the specific finding from the first case study that the spring discharge observations are of most "worth" when considering spring discharge forecasts is not surprising. We state this explicitly on Lines 243-245: "The worth of MRT observations relative to various hydraulic potential and discharge observations across the different forecasts are, in general terms, similar to those reported by Zell et al. (2018)". To address this comment, we will add the Hunt et al. (2006) reference to this sentence, and also add a sentence to the Results section (of the first case study) citing the Masbruch et al. (2014), La Vigna et al. (2006) and Oehlmann et al. (2015) references.

References

Doherty, J.E.: PEST and its utility support software: Theory, Watermark Numerical Publishing, 2015

Riva, M., Guadagnini, A., Fernandez-Garcia, A., Sanchez-Vila, X., Ptak, T.: Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site. Journal of Contaminant Hydrology 101, 1, 1-13, doi: 10.1016/j.jconhyd.2008.07.004, 2008

Sorooshian, S.: Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors âĂŤ The noninformative data case. Journal of Hydrology 52, 1–2, 127-138, doi: 10.1016/0022-1694(81)90099-8, 1981

Turnadge, C. Smerdon, B.D.: A review of methods for modelling environmental tracers in groundwater: advantages of tracer concentration simulation. Journal of Hydrology 519, 3674–3689, doi: 10.1016/j.jhydrol.2014.10.056, 2014

White, J.T., Doherty, J.E., Hughes, J.D.: Quantifying the predictive consequences of model error with linear subspace analysis. Water Resources Research 50, 2, 1152-1173, doi: 10.1002/2013WR014767, 2014

White, J.T., Knowling, M.J., Moore, C.R.: Consequences of model simplification in risk-based decision making: An analysis of groundwater-model vertical discretization, Groundwater, doi: 10.1111/gwat.12957, In Press

Wallis, I., Moore, C., Post, V., Wolf, L., Martens, E., Prommer, H.: Using predictive uncertainty analysis to optimise tracer test design and data acquisition. Journal of Hydrology 515, 191-204, doi: 10.1016/j.jhydrol.2014.04.061, 2014

Zell, W.O., Culver, T.B., Sanford, W.E.: Prediction uncertainty and data worth assessment for groundwater transport times in an agricultural catchment. Journal of Hydrology 561, 1019-1036, doi: 10.1016/j.jhydrol.2018.02.006, 2018

Zheng, C., Gorelick, S.M.: Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. Groundwater 41, 2, 142-155, doi: 10.1111/j.1745-6584.2003.tb02578.x, 2003

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