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.

Chris Turnadge

chris.turnadge@csiro.au

I would like to thank the authors for their invitation to provide comments on this manuscript. I believe this manuscript provides a valuable and timely contribution towards guidance in the use of subsurface environmental tracers to improve the predictive capability of groundwater flow and transport models. Many publications have implored researchers and practitioners to include a range of non-standard observation types such as temporal differences (Peeters et al., 2011), temperatures (Anderson, 2005), isotope concentrations and activities and inferred residence time (or "ages") (Schilling et al., 2019), and/or geophysical data (Hinnell et al., 2010) in model inversion

and prediction uncertainty minimisation. However, 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.

On first reading of the manuscript I questioned whether the authors were "over-reaching" in their conclusions. 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. More generally though, I do believe the manuscript will benefit from some revisions. Specifically, I would like to provide three major criticisms of the manuscript. These mainly relate to the suitability of the experimental design, in terms of the suitability of testing the hypotheses presented. These are followed by a number of minor criticisms that I believe the authors should also consider addressing. These minor criticisms mostly relate to the interpretation of environmental tracers, or to descriptions provided of the parameterisations of the numerical models used.

I hope my comments are helpful to the authors in improving the manuscript and I am more than willing to provide further clarifications off-line, if they are needed.

Major comments:

1. Tracer observations are of limited additional value when direct observations of fluxes are already available

For the Heretaunga Plains example, I do not believe that the addition of environmental tracer observations (in this case, tritium) when flux observations (in this case, spring discharge) are already available provides an ideal (i.e. fair) test case. Tracer concentrations are proxies for fluxes (either recharge, lateral or discharge) so they are often measured (and subsequently included in model inversion) when direct observations of fluxes are not available. Assessing the value of tracer observations when flux observations are not available would provide a fairer test case and would be of greater interest. For the Heretaunga Plains example, this could be implemented by simply

omitting spring discharge observations from all model inversion. âĂČ

2. Tracer observations are of limited additional value when collected upstream of prediction locations

For the Hauraki Plains example, subsurface tritium observations were recorded far upstream of nitrate discharge predictions. Since environmental tracers such as tritium integrate information along flow paths, it is intuitive that these tritium observations would contain limited information of value to predictions of nitrate concentration located far downstream. If subsurface tritium concentrations are available from locations in the vicinity of the predictions of interest, then I suggest that it would be more relevant to include these in the case study.

3. Conservativeness and management thresholds when presenting prediction uncertainty

For the Hauraki Plains example, I believe that the authors criticise simple modelling approaches without providing any discussion of whether modelled predictions are conservative. Prediction histograms are presented in the absence of a management threshold; in this case, an upper permissible limit for nitrate discharge. I suggest that the authors present a relevant management threshold when presenting prediction uncertainty results. This would allow the authors the additional benefit of exploring whether predictions produced using simple and complex parameterisation approaches were conservative.

Minor comments

1. Mean residence times

Mean residence time (MRTs) values were used to quantify the age of groundwaters. However, MRTs are equivalent to groundwater ages only under very strict assumptions; specifically, requiring highly simplified conceptualisations. The latter are the basis of mixing functions used in lumped parameter models. In most (complex, real-world)

cases, MRTs act as a fitting parameter. This is especially true if MRTs are derived from binary mixing models, which arbitrarily blend two mixing models, which generally undermines their physical bases.

As a solution, and rather than the use of lumped parameter models to derived mean residence times, where possible I suggest simulating the reactive transport (or at least, non-reactive transport with first order decay) of environmental tracer concentrations (or activities. Admittedly, for some tracers (such as carbon-14) the complexity of reactions may make reactive transport simulation prohibitive. However, the examples presented by the authors already feature combined numerical flow and transport models. Additionally, the authors' examples also feature a relatively simple tracer requiring only the simulation of first order decay, so I would assume that reactive transport simulation would be feasible.

2. Binary mixing models

The authors state that, as part of lumped parameter modelling, binary mixing models (BMMs) were used to derive mean residence times from subsurface tritium concentrations. BMMs are a linear combination of two other (ideally) physically-based mixing models. I suggest that the authors describe which mixing models were combined using the BMMs, the relative contribution of each, and the physical meaning of the combined result.

3. Atmospheric tritium concentrations

The authors state that the historic record of atmospheric tritium concentration features a "shape [that allows] for unique interpretation" (lines 69–70). It is unclear what the authors mean by the term "unique" in this context. Unlike SF6, it would not be true to state that the historical atmospheric tritium record is consistently monotonic. Tritium values increased initially due to nuclear weapons testing and have declined ever since. In some cases, the historical atmospheric tritium record features more than one peak value after the cessation of nuclear weapons testing. In addition, historic records of

atmospheric tritium concentrations at various locations are typically quite noisy, unlike SF6. See Figure 1 in McCallum et al. (2014) for an example of a noisy, two-peak example of a historical atmospheric tritium record. For this example, and following correction for radiometric decay in the subsurface, a tritium observation of 20 TU would correspond to any of four different recharge times (and therefore ages). In comparison, a historical atmospheric SF6 record is also shown by McCallum et al. (2014), which increases monotonically and would therefore permit unique interpretations of "ages" from measured concentrations. I suggest that the authors could state instead that tritium is a popular tracer for the identification of young age groundwaters (i.e. <70 years old) for the following reasons. Unlike CFCs, it is not affected by microbial degradation or contamination and, unlike SF6, it is not affected by potential subsurface sources. The authors may wish to cite Beyer et al. (2014), who provided a comparison of traditional (e.g. 3H, CFCs, SF6) and emerging (e.g. Halon-1301, SF5CF3) young age tracers.

4. Non-reactive modelling of contaminant transport

The Hauraki Plains model simulated nitrate as a non-reactive constituent. In practice, nitrates in the subsurface are subject to a range of processes: assimilation, nitrification/denitrification, volatilisation, sorption/desorption and retardation (Kendall and Aravena, 2000). For this reason, I suggest that the authors explain why nitrate was not simulated as a reactive constituent in the forward model.

5. Screen lengths of wells sampled for environmental tracers

When interpreting subsurface concentrations of environmental tracers, knowledge of the length of screened sections in sampled groundwater wells is crucial. If lumped parameter models are used, this affects the choice of mixing model. For example, if sampled wells are open holes or fully screened then the exponential mixing, exponential–piston flow, or dispersion models may be appropriate. Alternatively, if sampled wells are partially screened then the partial exponential model may be appropriate. Given the importance of this information to tracer interpretation, I suggest that the authors

describe the screen extents of each sampled well. The authors could also state how this information was used to select an appropriate mixing model for lumped parameter modelling, from which mean residence times were calculated.

6. Pilot point parameterisation

Pilot point parameterisations of the Heretaunga Plains and Hauraki Plains models are not described explicitly in the manuscript. Specifically, it is not clear which parameters were parameterised using this method. I suggest that the authors describe explicitly which model parameters were implemented on a cell-by-cell basis, or using pilot points, zonation or using spatially uniform values, including horizontal and vertical hydraulic conductivity, specific yield/storage, recharge and, for transport models, porosity.

7. Variogram definitions

It is not clear whether, for a given model, the same variogram was used to implement pilot point parameterisation for one or many parameter types. For example, I would not expect that the spatial correlation between hydraulic conductivity values to be the same as for recharge rates. I suggest that the authors state explicitly which variogram parameter values (e.g. correlation length, range, sill, nugget) were used to define which model parameter values, and describe the spatial analyses used to quantify spatial correlation between parameter values. Given that the degree of model complexity (particularly in relation to the ability of a model to assimilate observed data) is a key focus of the manuscript, I believe that detailed descriptions of the model parameterisation used are relevant.

8. Bias and underestimation

The authors state that the assimilation of tritium can induce "biased first moments or underestimated second moments" (line 55). I suggest that the authors could unpack this statement by providing simple examples to support this statement, for both bias and underestimation. The authors could also state explicitly the nature of the bias; i.e.

whether prediction mean values were under- or overestimated.

9. Ensemble size representativeness

The authors state that their implementation of the Iterative Ensemble Smoother featured ensemble sizes of 100 (lines 194-195). This value appears to have been selected arbitrarily, likely based on logistical constraints (e.g. forward model and inversion computing times). Was bootstrapping or other representativeness/convergence testing methods used to assess whether an ensemble size of 100 representative, and/or whether ensemble statistics converged as the ensemble size approached 100? I suggest that the authors demonstrate that the ensemble size used was representative.

10. Vertical coarsening of model grid

The authors provide limited explanation of why vertical coarsening of the model grid led to fewer relatively long flow paths. Since this observation is crucial to the interpretation of the authors' results, I suggest that the authors expand their discussion of this key point.

References cited:

Anderson, M. P. (2005). Heat as a ground water tracer. Groundwater, 43(6), 951-968.

Beyer, M., Morgenstern, U., and Jackson, B. (2014). Review of techniques for dating young groundwater (< 100 years) in New Zealand. Journal of Hydrology (New Zealand), 53(2), 93-111.

Hinnell, A. C., Ferré, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J., and Kowalsky, M. B. (2010). Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. Water Resources Research, 46(4).

Kendall, C., and Aravena, R. (2000). Nitrate isotopes in groundwater systems. In: Environmental Tracers in Subsurface Hydrology. Cook, P. G., and Herczeg, A. L. (Eds.). Springer, Boston, Massachusetts, USA, pp. 261-297.

C7

McCallum, J. L., Engdahl, N. B., Ginn, T. R., and Cook, P. G. (2014). Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time?. Water Resources Research, 50(3), 2022-2038.

Peeters, L. J. M., Rassam, D., and Lerat, J. (2011). Improving parameter estimation in transient groundwater models through temporal differencing. In: Chan, F., Marinova, D., Anderssen, R.S. (Eds.), MODSIM2011, 19th International Congress of Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand; Perth, WA, Australia; 12–16 December 2011, pp. 3959-3965.

Schilling, O. S., Cook, P. G., and Brunner, P. (2019). Beyond Classical Observations in Hydrogeology: The Advantages of Including Exchange Flux, Temperature, Tracer Concentration, Residence Time, and Soil Moisture Observations in Groundwater Model Calibration. Reviews of Geophysics, 57(1), 146-182.

C8