Dear Reviewer 1,

We very much appreciate your thorough review. You find your comments (RC) in *italic,* our responses (AC) in **bold and changes in quotation marks (" ")**.

Kind regards,

Rena Meyer, on behalf of the authors

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-99-RC1, 2018

Interactive comment on "Estimation of effective porosity in large-scale groundwater models by combining particle tracking, auto-calibration and 14C dating" by Rena Meyer et al.

Timothy Ginn (Referee)

tim.ginn@wsu.edu

Received and published: 20 April 2018

General comments

The paper details a significant modeling effort demonstrating the importance of carbon-14 dating in the calibration of spatially-distributed porosity. The study utilizes a previously calibrated 3D groundwater flow model of the site and selects 11 of 18 carbon-14 data as targets. I have two major concerns and several other concerns about the implementation of the inverse method and the conceptualization of the apparent ages. The latter are detailed in the specific comments and the former are:

1. the model assumes the conductivity field inherited from the (unpublished, at the time of this review) Meyer

et al. (2018a, and b which is in preparation); and 2. the data are prefiltered (e.g., eliminated) based on their coherence with the inherited model prior to the analyses. While I highly respect the authors' work in the field and I believe this work has a substantive contribution in the rarely touched world of porosity estimation, I think there are important elements that require consideration and careful address in the discussion. Details of my concerns follow.

RC# 1:

The very significant reliance on the unpublished groundwater flow model, and its fixed hydraulic conductivities, raises concerns about the current study. The current study seeks to identify porosities of 7 geological units by *fitting them so that the mean ("direct") ages match the apparent ages from carbon-14 corrected for dissolution and diffusion; however, there is no allowance for departures from the originally calibrated conductivities (from the unpublished Meyer et al. 2018a). Thus the porosities are treated as if they are independent of the hydraulic conductivities. This is not conventional and disagrees with current understanding of the properties of natural porous media, and needs to be addressed by the authors.*

AC#1:

The groundwater flow model that forms the basis for this study is now published in Journal of Hydrology as "Meyer et al. 2018: Regional flow in a complex coastal aquifer system: combining voxel geological modelling with regularized calibration", DOI: 10.1016/j.jhydrol.2018.05.020."

The effective porosities of the 7 geological units were calibrated using an advective particle tracking model (MODPATH) in the way that the mean average particle tracking time, based on 1000 particles released in each of the 11 cells where C-14 observations were available, (not the mean "direct age"), match apparent ages from C-14. These estimated effective porosities were subsequently used in a "direct age" simulation to analyze the age distribution in the entire aquifer.

We decided to approach the calibration of this large groundwater model in two steps procedure to enhance stability and well-posedness of the inverse problem. Simultaneous estimation of both flow and transport parameters resulted in stability problems where small changes in the weighting resulted in large changes in parameter values (unrealistic parameter estimates). Hence, following e.g. Carrera et al. (2009) the number of parameters were reduced by dividing the estimation problem into two stages (flow and advective transport). Hereby, realistic parameter estimates were obtained and an acceptable match to the targets were found.

Moreover, in our setup, the fluxes are pre-determined by boundary conditions such as recharge and constant heads in streams, drains and ocean. As a consequence, changes in hydraulic conductivities would come along with changes in the gradient, but would not necessarily change the fluxes dramatically and hence not influence the advective age (based on particle tracking). On the other hand, changing the effective porosity would have a more direct influence on the age.

RC# 2:

Multiple aspects of the inversion done in Meyer et al. are important here since that work laid the foundation flow model; for instance, the vertical anisotropy factors assigned from that work are 25 for sand and 85 for clay units, which are quite high, and qualitatively at least would seem to restrict vertical migration of water in a way that would definitely affect age.

AC#2:

We agree that the flow model is of major importance and by now the article Meyer et al. 2018 is also published and available (see AC #1). We refer to this paper for a discussion on the anisotropy factors?

RC# 3:

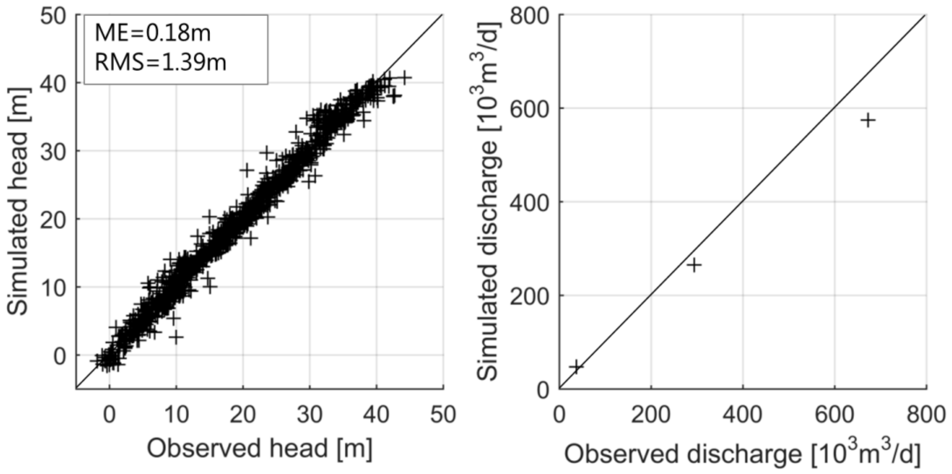
A more robust approach would have been to do a wholistic inversion, where the conductivity (and other flow and transport parameters) were calibrated at the same time as the porosity (and other transport parameters, including the dispersivity, set to zero here based on a brief local sensitivity), to the collective head and apparent age data. Why this is not done, and the potential constraints on the resulting two-stage inverse, should be discussed.

AC #3: We agree that a holistic inversion would have been desirable. However, given the model size (millions of nodes) and the runtime this is not possible. This is also the reason why we decided to use a step wise approach and only calibrate effective porosity at this stage, based on a calibrated flow field (Meyer et al., 2018)**, as this can be estimated using a particle model, which runs much faster than the full advective-dispersion model. It was not possible to perform a calibration on the full automated advective-dispersion model which requires several thousands of model runs. Of course our approach has limitations in a way that maybe information that is contained in the age observation about hydraulic conductivities is not fully exploited. Dispersion parameters are not possible to estimate using an advective transport model only. However, we think that our study still shows the benefit of estimating effective porosity instead of applying a uniform literature value.**

RC #4:

There are no error plots from the prior head-inversion of Meyer et al so the success of the calibration of the flow equation is unknown. More importantly for a subsequent inversion for porosity, there is no indication of the uniqueness of that first inversion. Even if that inversion gave good results, it may be non-unique, and it seems that there may be a different set of hydraulic conductivities and porosities which together might fit both the available head and carbon-14 data.

AC #4: Error plots and an uncertainty analysis of parameters are now available in the published article by Meyer et al., 2018 (see AC#1). We also add a new Figure 3, showing the calibration results for simulated vs. observed data.



RC #5:

The elimination of dispersivity appears not only somewhat arbitrary but also contradictory to the authors' overall argument for the importance of porosity (cf. specific comment on page 8 line 21). It appears they have replaced the modeling of mobile- immobile domain mass transfer in the model with the approximate diffusion-correction applied to the data. This could be justified based on pragmatic grounds but the discussion in this regard is lacking. The alternative to use effective mobile-immobile domain mass transfer seems potentially useful and pragmatic as well but is not discussed.

AC #5: We agree with the reviewer's argument that our approach is a simplification with regard to dispersivity and exchange between flow and stagnant zones. We base the calibration of the distributed effective porosity field on a steady state flow field and use a particle tracking scheme. This approach was needed to keep the computational effort down for the calibration (several thousands of runs). Even with our approach we gain still an important insight in the age distribution in a large scale complex multi-layer aquifer system. And it is shown that choosing a simple porosity estimation scheme is still beneficial compared with applying an uniform porosity.

RC 6#:

Very important is the unsupported elimination of 7 pesky carbon-14 data (cf specific comment on page 8 line 30). The focus only on the data which are consistent with the already partly calibrated model brings the entire study into question.

AC #6: See AC#15

RC 7#:

Why the recently developed methods for full distribution of age (e.g., several articles in J Hydrology, December 2016) are not used is not described; however, this may be attributed to the reliance on single radiometric tracer (carbon-14) concentration measurements, which precludes any inferrence of age distribution.

AC #7: Due to the complex nature of our hydrogeological model and the limitations to only one age tracer, as correctly identified by the reviewer. Moreover, our article focuses on the need to include effective porosity into groundwater model calibration which we demonstrate by the use of groundwater ages.

“The groundwater science community (de Dreuzy and Ginn, 2016) has a continued interest in the topic of residence time distributions (RTD) in the subsurface.“

“It would have been optimal to use RTD analysis (de Dreuzy and Ginn, 2016) to compare modelled and inferred groundwater ages in this study. But, due to the rather complex nature of our hydrogeological flow model, the inherent uncertainties associated with inferring an apparent age to ^{14}C, and the long computer runtimes, we have chosen to use the particle-based kinematic approach of simulating a mixed age at the well screen (or numerical cell with a screen).”

Specific comments.

RC 8#:

page 2 line 4. "Three different apporaches with specific benefits and disadvantages are commonly applied to simulate groundwater age..." The given list of commonly used methods is not complete (there are also the lumped-parameter approach, and the mixing cell model approach), and equally important are the new methods which are generally more robust [solving the actual equation of groundwater age, either by the Laplace method of Cornaton (WRR 2012) or by using reduced dimensions as in Woolfenden and Ginn (Groundwater, 2009)]. The review by Turnadge and Smerdon (JHydrology 2014) provides a more complete listing and assessment.

AC #8: We agree with the reviewer that there are more methods to calculate groundwater ages and their distribution. We add a sentence and include other methods.

“Turnadge and Smerdon (2014) reviewed different methods for modelling environmental tracers in groundwater including lumped parameter models (e.g. Maloszewski and Zuber, 1996), mixing-cell models (e.g. Campana and Simpson, 1984; Partington et al., 2011) and direct age models (e.g. Cornaton, 2012; Goode, 1996; Woolfenden and Ginn, 2009). Here, we explain three different approaches with specific benefits and disadvantages that are commonly applied to simulate groundwater age in 3D distributed groundwater flow and transport models (Castro and Goblet 2005; Sanford et al. 2017).”

RC 9#:

page 2 line 12. "A comparison of ages simulated using any of these methods with ages determined from tracer observations, referred to as apparent ages is desireable..." This is true but omits the very important point that "ages determined from tracer observations" are not equal to mean ages, especially as in the present case of decaying environmental tracers (e.g., carbon-14). The rest of this paragraph summarizes part of the way that "apparent ages" are misled by old carbonate dissolution, by diffusion, and by heterogeneity, following McCallum's work; however, it should also point out the fundamental difference between mean ages and radiometric ages described explicitly by equation 16 of Varni and Carerra (WRR 1998), and the general relation between distribution of age and the radiometric age given in Massoudieh and Ginn (WRR 2011).

AC #9:Thank you for this comment, we have added text;

"It is important here to distinguish between mean and radiometric ages as for example defined by Varni and Carrera (1998). The only way they can be directly compared in reality is if no mixing is taking place, i.e., if the flow field can be regarded as pure piston flow, which will give the kinematic age. " (introduction)

"The particle-based approach used in this study computes the kinematic age at a point. With 1000 particles released in each cell with a screen, we essentially get an age distribution of kinematic ages by perturbing the measurement location within the cell reflecting the mixing of waters from different origins. The C-14 ages have also been diffusion-corrected so that dilution or mixing due to loss of C-14 into the stagnant zones have been accounted for." (Methods page 8)

RC 10#:

page 2 line 238. "Bethke and Johnson (2002) concldued that the groundwater age exchange... is only a function of the volume of stored water." This is misleading because it is valid only for the mean groundwater age, and requires steady-state as detailed in Ginn et al. (Tranpsort in Porous Media, 2009). Also this point is made earlier and more precisely in Varni and Carerra (op. cit., page 3272), who points out that it is actually a result of Haggerty. The overall point by the authors that porosity is important to age modeling is valid.

AC #10: We agree and specify more clearly under which assumptions this is valid and include the suggested references:

"However, for steady state flow (Ginn et al., 2009) in a layered aquifer system, Bethke and Johnson (2002) concluded that the mean groundwater age exchange between flow and stagnant zones is only a function of the volume of stored water (Harvey and Gorelick, 1995; Varni and Carrera, 1998)."

RC 11#:

page 3 line 1. "neglecting dispersion effects seemed to be acceptable at large scale" is unsupported for the present application, results of cited Sanford and later Gelhar notwithstanding. See comments below (re: page 8 line 21 and the reliance on Sanford; page 10 lines 14-17 and Figure 8) for more discussion.

AC #11: see AC#14 and AC#18

RC 12#:

page 6 line 27. "Meyer et al. (2018b) simulated ....further details can be found in Meyer et al. (2018a)." Actually they cannot because Meyer et al. (2018a) is in submitted state (page 30 line 28). This is quite important because the present authors have chosen to rely upon the hydraulic conductivity field previously calibrated in that work, and here do not allow the conductivity values to be modified in the inversion using carbon-14 inferred ages (page 8 line 26).

AC #12:

The study Meyer et al. 2018 is published and available now (see AC#1).

RC#13:

page 8 line 2. "The resulting head distribution is shown in Figure 1." Figure 1 shows (it seems to me) only the shallow aquifer heads. It is well-known that the quality of an inversion of the flow equation (to determine hydraulic conductivities) depends on a broad spatial distribution of the heads, and it is unclear that such head data were available to Meyer et al. Also, there are no error plots showing the goodness-of-fit of the flow inversion to the measured heads, so it is impossible for the reader to evaluate how good was the flow equation inversion. Also it is impossible for the reader to evaluate the uniqueness of the flow equation inversion, which is commonly very poor.

AC#13:

More than 1000 head observations from different depths and aquifers were available and the information can be found in the published article Meyer et al. 2018. Calibration performance of the flow model in terms of goodness-of-fit, ME, RMS. Meyer et al. 2018 also contains an identifiability and uncertainty analysis of the estimated parameters as well as an evaluation and discussion of the non-uniqueness of the flow model.

We extended the description of the numerical model:

"The steady-state MODFLOW flow solution (calibration results summarized in Figure 3; Meyer et al. (2018a) also contains an identifiability and uncertainty analysis of the estimated parameters as well as an evaluation and discussion of the non-uniqueness of the flow model.) forms the basis for the advective transport simulation using MODPATH."

RC#14:

page 8 line 21 "According to Sanford (2011), neglecting hydrodynamic dispersion... on a regional scale is a reasonable approach when old-age tracers, such as carbon-14, are used as dispersion might not be crucial for these tracers." This sentiment is unclear because it suggests that there is something particular to the carbon molecule that frees it from dispersion, which is quite incorrect. It is also directly in opposition with the authors' claim (page 2 line 28ff) that porosity is important for groundwater mean age determination because "groundwater age exchange between flow and stagnant zones is only a function of the volume of stored water."

AC#14: In order to prevent any confusion we take the reference to Sanford out. Our intention was not to argue that we do not have physical dispersion in our system, but that we have a relative high resolution of geological heterogeneities resulting in flow scales of few hundreds of meters and hence physical dispersion of few meters (Gelhar). At the same time we expect some numerical dispersion due to the solver we used (standard finite difference) and the grid size. Hence, the numerical dispersion could overrule the physical one. This is why we set the physical dispersion to 0m.

RC#15:

page 8 line 30ff. The authors removed 7 data from their 18 carbon-14 measurements becaue the values did not match their conceptual model; 6 were deleted because the carbon-14 activities were below 5pmc, and one due to proximity to another sample with different value. The justification given for the first 6 is "it was assumed that the boundary conditions of the flow model ... were not representative for pre-Holocene conditions." This justification is unclear at best; the model is steady state so the initial conditions do not matter, and the boundary conditions are necessarily (by the steady-state assumption) constant. Thus the elimination of the low carbon data is unsupported. The elimination of the 7th datum is only weakly justified, as there appears to be nothing wrong with it other than its troubling value.

AC#15: It is right that the model is steady-state. However, the boundary conditions represent modern conditions. The eastern part of area has been affected by the Scandinavian Ice Sheet during the Weichselian. This ice cap probably induced a high hydraulic pressure with dramatic influence of the hydraulic system (e.g. Piotrowski, 1997) **and the boundary conditions in the East. We believe that the 6 C-14 measurements with C-14 activities below 5pMC might be influenced by these conditions and eventually recharged outside the modern eastern boundary. Therefore we decided to calibrate the model only based on the measurements that were recharged during similar hydraulic conditions as today.**

The 7th data point was excluded because there is an age inversion in the observations which might be a result of local heterogeneity and it would probably not be possible to reproduce this by the model with the current cell size. The age observations are located in neighboring cells, the younger one directly below the old one.

This would have caused troubles during the calibration. Therefore, we decided to exclude the data point from the calibration.

RC#16:

page 9 line 4-6. The weights on the data used in the inversion were all the same. They were based on an average uncertainly of apparent ages of ~102 years, as per "average of the standard deviation of the diffusion correction for the selected 11 samples..." This defeats the purpose of calculating individual standard deviations for individual data in the first place. The individual standard deviations (Table 1, last column) show a range of 8 to 310 years, so individual weights based on these values would have led to significantly different weights. Individualized weighting is rarely possible in groundwater flow model inversion but is often possible in transport inversion, and it seems to me that the authors have unintentionally limited the inversion by assigning equal weights to all apparent age data. The importance and utility of weighting is amply described in the books by John Doherty and Mary Hill, and could have been used to condition the data per their individual certainties; moreover it could have been used to condition - perhaps to good end - the pesky 7 data that were eliminated instead. In fact, the standard deviations of the 6 eliminated data range from 1323 to 2593 years, which would have led to quite significant reduction in the importance of these data as the weights are generally taken as the reciprocals.

AC#16: We had several calibration experiments including individual weights. However, the fit to the older ages, having larger standard deviations was worse, while the one to the younger ages not significantly improved. By applying a uniform weight we intentionally gave higher weight to the older ages than to the younger ones. We decided to not include the 6 data points as justified in AC#15.

RC#17:

page 9 line 27. "mean groundwater age is simulated in analogy to solute transport as an "age mass" (Bethke and Johnson 2008)." This "age mass" requires mathematical and physical definition; as pointed out in Ginn et al (2009, op. cit., section 2.2) the definitions of Goode and of Bethke and Johnson are not clear or consistent. The example of Bethke and Johnson involves an aquifer and an aquiclude with only immobile water, so that diffusion is the only mechanism by which exchange can take place. If it is eliminated, then the argument collapses.

AC#17: We are not sure what the reviewer means. Essentially eq. 8 is identical to the one in Goode (1996), which we refer to, or for that matter, Varni and Carrera (1998). We did not eliminate diffusion but physical dispersion (see AC#18).

RC#18:

page 10 lines 14-17 The numerical experiments to evaluate dispersion effects, described here, with results summarized on page 15 lines 10ff and in Figure 8, are apparently done on one model, that is, on one assignment of hydraulic conductivities and porosities. It is not clear which porosities were used. In any case, this is at best a local parameter sensitivity analysis and it would be more accurate to include the dispersivity values in the inversion. The argument that the 200mx200m grid cell size is sufficiently resolved to allow ignoring dispersion is unconvincing, because there are multiple modeling exercises where the effective dispersivity is proportional to the grid cell size, not zero. Figure 8 does not tell how the errors grew but only the total error - did the errors go biased ? If one were to guess, I would bet they did, because the dispersion would allow mass transfer laterally, causing generally older ages.

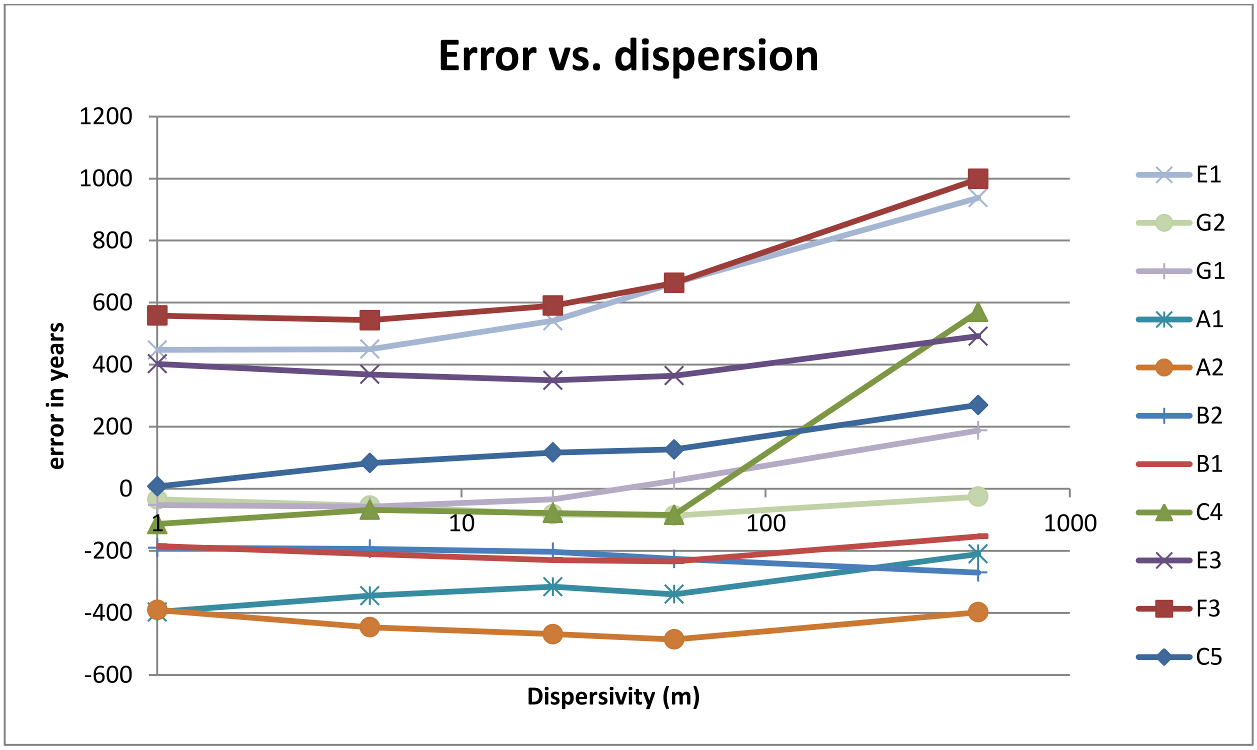
AC#18: The estimated porosities using the regularized inversion scheme and shown in Table 2 were used. The dispersity values were not possible to include into the inversion scheme as we used an advective particle tracking model (MODPATH) for the automated calibration due to the long run times of the full advective-dispersion model (MT3DMS). To still investigate the effect of physical dispersion we did as correctly mentioned by the reviewer a local sensitivity analysis. As we used the standard finite difference scheme, we expect a some numerical dispersion with our grid size. From figure 8 one can approximate a numerical dispersion in the order of several tens of meters. Our geological modelling approach and the transformation into a hydrogeological mode as detailed explained in Meyer et al 2018, resolves geological heterogeneities on a grid scale which is 200m to 400m and the flow at a similar scale. Hence, we assumed that physical dispersion at regional scale is accounted for by including a detailed description of the geological heterogeneity (≈ 200 m scale). Therefore, only local scale mixing processes needs to be described by the dispersivity concept, as larger scale processes are taken care of by a detailed description of geology. According to Gelhar et al., (1992) physical dispersion would at this flow scale range from one meter to several meters, which is also in accordance with studies in the Dutch polder system where dispersivity values of 2m are applied in similar sized models (e.g. Oude Essink et al., 2010; Pauw et al., 2012). In our system, we assume that numerical dispersion is in the same order of magnitude and therefore is sufficient to account for the local scale mixing processes not accounted for by the heterogeneities build into the model.

In the figure below the errors for the individual wells are illustrated as a function of dispersivity. The error is generally constant for dispersivities up to 50 m, while it increase when the dispersivity is increased to 500 m.

We changed the sentences:

"The very detailed voxel geological model resolves heterogeneities at a scale of 200m x 200m. Hence, it is assumed that mixing at scales larger than 200m is accounted for by the geological model. Therefore, the dispersivity should only describe the heterogeneity at a flow scale of several hundred of meters which justifies the use of a relatively small α_L. In accordance with Gelhar et al. (1992) flow scales of hundreds of meters result in α_L of magnitudes in the range of a few meters, which is also in line with studies in the Dutch polder system where dispersivity values of 2m were applied in similar sized models (e.g. Oude Essink et al., 2010; Pauw et al., 2012). On the grid scale of 200m to 400m and with the standard difference solver for the advection-dispersion equation a substantial numerical dispersion is expected. Since there is no sensitivity for

lower α_L (numerical dispersion dominates at this scale), a physical dispersivity was set to zero m in the following simulations of direct age. This does not imply that physical dispersion does not exist, only that physical dispersion is accounted for by numerical dispersion.”



RC#19:

page 11 line 6 "...as porosity does not impact the trajectory of the particle path..." this is true only via the assumption that the porosity and hydraulic conductivity are independent, which is not common.

AC#19: As correctly mentioned by the reviewer, our description is valid and limited to our assumption that in the approach we choose the hydraulic conductivity field is constant (see AC#1). To avoid misunderstanding we add these limitations to our description.

“Given that the hydraulic conductivity field is unchanged, no differences in the area of the whole capture zone are expected as porosity does not impact the trajectory of the particle path (Hill and Tiedeman 2007) and only affects the travel time.”

RC#20:

page 13 Figure 4a. The plot demonstrates in my view limited improvement for two reasons. First, the 5 older water samples (with carbon-14 corrected ages greater than 500 years) show significantly improved fitting in 3 cases, with one getting worse. Second, the plot is absent of confidence intervals (compare for instance to Figure 11) which could be it seems to me estimated based on the standard deviations of the corrected carbon-14 ages (Table 1), with additional uncertainty based on equation 16 of Varni and Carrera (op. cit.). The recognized uncertainty in the apparent ages should it seems be used to condition the results of Figure 4a.

AC#20: The calculation of uncertainty using equation 16 of Varni and Carrera requires the 2nd moment of the direct simulated age distribution which we do not have. In Figure 4 the age is calculated based on particle tracking not on direct age modelling (advection-dispersion equation), hence it is not possible to calculate uncertainty using equation 16 of Varni and Carrera as this requires the 2nd moment of the direct simulated age distribution using the advection-dispersion equation. However, we add the standard deviation derived from particle-based pdf at a well screen and age correction of the measurements to Figure 4a. As mentioned in the text (on page 11) we achieve a significant reduction in both ME and RMS compared to the uniform-porosity field model.

RC#21:

page 16 line 4ff "Hence, the dispersivity only describes the effect of heterogeneity at the grid scale, 200m. In accordance with Gelhar et al. (1992) this results in (dispersivity) with a magnitude of a few meters." I am unaware that Gelhar suggested this dispersivity value given (only) the size of the grid, please provide the page. Also in the intervening 25 years there has been extensive research and articles on the effective dispersivity for regional groundwater models, and more up to date referencing is called for. Notably, the model (including its effective parameters) at the 200m grid block scale tells only the expected or mean concentration in the grid block, that is, the concentration in the model is treated as a constant on the 200m x 200m x 5m grid block, while the carbon-14 data are collected from sampling wells on much smaller spatial scales - this issues should also be addressed or at least noted.

AC#21: We agree with the reviewer that the Gelhar plots refer to the flow scale and not the grid scale. This is actually what we meant. Thanks to the high resolution of the description of geological heterogeneities (cf. (Meyer et al., 2018) **we reach flow scales in the order of several hundred meters. According to Gelhar physical dispersivities would be in the order of several meters at this flow scale. To be more precise we changed the wording (compare AC#18).**

The problem of commensurability, the problem of comparing point measurements with a mean value for a large volume, is added to the discussion.

“The comparison of groundwater ages, estimated from tracer concentration in a water sample, and simulated groundwater ages, either derived by particle tracking or direct age modelling, bears the problem of commensurability, the comparison of a point measurement relative to the modelling scale. The water

sample represents the age distribution in the direct surrounding of the well screen which only makes up a few percent of the water in one model cell.

The differences between mean advective ages and directly simulated mean ages as described in section 4.4 can be related to the simulation methods. While particle tracking neglects dispersion, but allows simulating an age distribution in a cell (by perturbing the measurement location so to speak), direct age modelling allows to account for dispersion/diffusion, resulting in only the mean age at a cell. The mismatches between advective and direct age can be related to the diffusion and dispersion processes (here represented by numerical dispersion as dispersion was set to zero), which are included in the direct age approach, but neglected in simulating advective ages.”

RC#22:

page 17 line 1. "The age distribution is strongly affected by geology and is therefore in good agreement with the interpretation of the flow system by Meyer et al. (2018)." This statement is unclear: the age distribution is always strongly affected by geology. Figure 10 caption "Normalized probability distributions..." These are frequency distributions because there is no randomness in the model or its parameters.

AC#22: We agree and specify that “the age distribution is strongly affected by the heterogeneity in flow and transport through the aquifers geology”. The description of Figure 10 is changed as suggested to “frequency distribution of…”

RC#23:

page 21 line 12 (regarding Figure 11) "However, most of them lie within one standard deviation." Seven of the standard deviations here span several thousands of years while the means for all but one are less than 7000 years, so this is not a comforting result.

AC#23: In order to not give the impression to the reader that these fits are perfect we follow the reviewers comment and extend the description of the results to “However, most of them lie within one standard deviation; but please observe that the standard deviation spans several thousands of year at some locations, where particle travel time distributions show a multi-modal shape.”

RC#24:

page 23 section 5.1.2. This discussion clearly identifies the ways that individual particle path history of exposure to different geologic units differentiates the actual true correction of the carbon-14 from the simplified correction done in the paper; however, it still does not tell about the fundamental difference between the *apparent age and the mean age (cf. comment on page 2 line 12). That is, even if the correction were perfect, the apparent age would not equal the mean age.*

AC#24: Thank you. We have extended the discussion to reflect this;

"As mentioned in the introduction, the apparent age (or radiometric age) is not equal to the mean particle-based kinematic age. This introduces some additional, but unknown uncertainty. Ideally, one could develop an advection-dispersion equation for the 2. Moment and solve for the variance of ages (Varni and Carrera, 1998) and use that together with the directly simulate mean age (or first moment) to establish a relation between radiometric and mean ages. This has not been pursued as we believe the benefits from this would be masked by uncertainty in age dating C-14 (i.e. uncertainty on analyses, and corrections for effects of geochemical and physical processes)."

RC#25:

page 24 line 6. "While direct age corresponds to the flux-averaged mean, the particle tracking age is resident-averaged (Varni and Carrera, 1998)." I do not see where this statement is given in the cited reference, please clarify if so; furthermore, I do not believe the statement is correct. The mean age of the model of Goode is an Eulerian quantity, just like a solute resident concentration. The relation between resident and flux-averaged concentrations is given in a number of papers by Parker and van Genuchten and coworkers (1984) but the governing equations that result are mainly restricted to 1D cases.

AC#25: We take this part out from the manuscript as it has no further relevance for the overall study and leads to confusion.

RC#26:

page 24 line 9. The use of harmonic mean for particle ages is absent of a rational basis other than it seems to fit the data well, and a generic reference to Konikow (2008). The specific manner of averaging the particle ages should be physically-based and independent of how well it fits the data in a particular setting.

AC#26: We take this part out from the manuscript as it has no further relevance for the overall study and leads to confusion.

References

Cornaton, F. J. (2012), Transient water age distributions in environmental flow systems: The time-marching Laplace transform solution technique, Water Resour. Res., 48, W03524, doi:10.1029/2011WR010606.

Ginn, T. R., H. Haeri, L. Foglia, and A. Massoudieh (2009), Notes on groundwater age in forward and inverse modeling, Transport in Porous Media,79:117-134.

Massoudieh, A., and T. R. Ginn (2011), The theoretical relation between unstable solutes and groundwater age, Water Resour. Res., 47, W10523, doi:10.1029/2010WR010039.

Parker, J.and M. Th. van Genuchten (1984), Flux-Averaged and Volume-Averaged Concentrations in Continuum Approaches to Solute Transport, Water Resour. Res., 20(7):866-872.

Turnadge, C., and B. D. Smerdon (2014), A review of methods for modelling environmental tracers in groundwater: Advantages of tracer concentration simulation, Journalof Hydrology 519, 3674-3689.

Woolfenden, L., and T. R. Ginn (2009), Modeled ground water age distributions, Ground Water, 47(4), 547–557.

Varni, M., and J. Carrera (1998), Simulation of groundwater age distributions, Water Resour. Res., 34, 3271–3282, doi:10.1029/98WR02536.

AC# Added references:

Campana, M.E., Simpson, E.S., 1984. Groundwater residence times and recharge rates using a discrete-state compartment model and14C data. J. Hydrol. 72, 171–185. https://doi.org/10.1016/0022-1694(84)90190-2

Cornaton, F.J., 2012. Transient water age distributions in environmental flow systems: The time-marching Laplace transform solution technique. Water Resour. Res. 48, 1–17. https://doi.org/10.1029/2011WR010606

de Dreuzy, J.-R., Ginn, T.R., 2016. Residence times in subsurface hydrological systems, introduction to the Special Issue. J. Hydrol. 543, 1–6. https://doi.org/10.1016/j.jhydrol.2016.11.046

Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A Critical Review of Data on Field-Scale Dispersion in Aquifers. Water Resour. Res. 28, 1955–1974. https://doi.org/10.1029/92WR00607

Ginn, T.R., Haeri, H., Massoudieh, A., Foglia, L., 2009. Notes on groundwater age in forward and inverse modeling. Transp. Porous Media 79, 117–134. https://doi.org/10.1007/s11242-009-9406-1

Goode, J.D., 1996. Direct Simulation of Groundwater age. Water Resour. Res. 32, 289–296.

Harvey, C.F., Gorelick, S.M., 1995. Temporal moment-generating equations: Modeling transport and mass transfer in heterogenous aquifers. Water Resour. Manag. 31, 1895–1911.

Maloszewski, P., Zuber, A., 1996. Lumped parameter modles for the interpretation of environmental tracer data, in: Manual on Mathematical Models in Isotope Hydrology, IAEA-TECDOC 910. pp. 9–50.

Meyer, R., Engesgaard, P., Høyer, A.-S., Jørgensen, F., Vignoli, G., Sonnenborg, T.O., 2018. Regional flow in a complex coastal aquifer system : Combining voxel geological modelling with regularized calibration. J. Hydrol. 562, 544–563. https://doi.org/10.1016/j.jhydrol.2018.05.020

Oude Essink, G.H.P., Van Baaren, E.S., De Louw, P.G.B., 2010. Effects of climate change on coastal groundwater systems: A modeling study in the Netherlands. Water Resour. Res. 46, 1–16. https://doi.org/10.1029/2009WR008719

Partington, D., Brunner, P., Simmons, C.T., Therrien, R., Werner, A.D., Dandy, G.C., Maier, H.R., 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water-groundwater flow models. Environ. Model. Softw. 26, 886–898. https://doi.org/10.1016/j.envsoft.2011.02.007

Pauw, P., De Louw, P.G.B., Oude Essink, G.H.P., 2012. Groundwater salinisation in the Wadden Sea area of the Netherlands: Quantifying the effects of climate change, sea-level rise and anthropogenic interferences. Geol. en Mijnbouw/Netherlands J. Geosci. 91, 373–383. https://doi.org/10.1017/S0016774600000500

Piotrowski, J.A., 1997. Subglacial groundwater flow during the last glaciation in northwestern Germany. Sediment. Geol. 111, 217–224.

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

Varni, M., Carrera, J., 1998. Simulation of groundwater age distributions. Water Resour. Res. 34, 3271–3281. https://doi.org/10.1029/98WR02536

Woolfenden, L.R., Ginn, T.R., 2009. Modeled ground water age distributions. Ground Water 47, 547–557. https://doi.org/10.1111/j.1745-6584.2008.00550.x