Dear Reviewer 2,

Thank you for your thorough review. You find your comments (RC) in *italic* (your comments were copied from the supplement.pdf document), our responses (AC) in **bold and changes in quotation marks (" ").**

Kind regards,

Rena Meyer, on behalf of the authors

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

Anonymous Referee #2

Received and published: 28 June 2018

(1) Scientific significance. The paper presents a case study in which inferences about the regional distribution of groundwater travel times are based on 18 measurements of 14C at 7 locations. In addition to the measurements, an existing groundwater flow model and a voxel-based geologic model were available and used. Only porosity (in 7 zones) was optimized, using the existing flow model with advection-only particle tracking. The resulting porosity field was used in a direct age simulation to generate the mean travel time distribution throughout the aquifer system. The distribution of travel times was explained in the context of the geologic structure of the system, which in turn was extended to a discussion of the general vulnerability of the system to various forms of contamination (natural, sea water intrusion, and anthropogenic).The paper largely uses concepts and methods that are well known. Groundwater models have been calibrated to travel times (many examples). One aspect of the paper that is not well represented in the literature is the sequential calibration of an existing flow model to travel times using only porosity, but this this too has been used before, as for example in Starn, J.J., C.T. Green, S.R. Hinkle, A.C. Bagtzoglou, and B.J. Stolp. 2014. Simulating water-quality trends in public-supply wells in transient flow systems. Groundwater 52(S1): 53-62.

(2) Scientific quality. The methods and analyses are sound. The discussion of travel times in the context of the geology is especially good. Although the researchers reach a different conclusion than another study in the same geographic area, the differences are explained well and make good sense. Once the relation of travel time and geology was established (in this paper), the geologic voxel model was used to make broad statements about the susceptibility of groundwaters in the area. The paper is a good example of using relatively few data points, along with existing data, in a thoughtful way that should enhance proper management of the resource.

(3) Presentation quality. The graphs and tables could easily be made clearer. Suggestions on how to do that are included in an attached PDF document. Please also note the supplement to this comment:

https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-99/hess-2018-99-RC2-

supplement.pdf
Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-99, 2018.

Specific comments:

AC: Starn et al. (2014) is added to the introduction

Page 2:

RC#1: It would have been nice to include a sensitivity analysis – how sensitive are travel times to changes in porosity.

AC#1: We did not include a formal sensitivity analysis, but instead we show the effect of a distributed effective porosity field compared to a uniform one on a capture zone delineation application, which shows a significant change based on the effective porosity.

Page 3:

RC#2: Seems redundant.

AC#2: The sentence is important as it describes one dominant feature, a man-made drainage system that lowers the water table below sea level, disturbs the "natural" flow system by enhancing the inflow of "young" ocean water.

Page 4:

RC#3: what method? Should be stated here

AC#3: we add the reference to Goode (1996).

Page 6

RC#4: which equation is this? 2?

AC#4: Yes, we add the reference to equation 2.

RC#5: make a brief statement about the steady state assumption – over what time period; what is the evidence for steady state?

AC#5: The system, close to the coast, is not expected to be in steady-state over a very long period. Changes in sea level over the last thousands of years and human activity (drains and dikes) over the last centuries changed the hydraulic system, especially in the west, close to the sea. While upstream, to the east, where

most of our samples were taken, the system was more steady over the last thousands of years. We included a discussion about effects of transient conditions (see also our response to the short comment).

<u>Page 7</u>

RC#6: what physical features do these boundaries correspond to?

AC#6: The physical features are delineated along flow lines and watershed boundaries; we add this to the description:

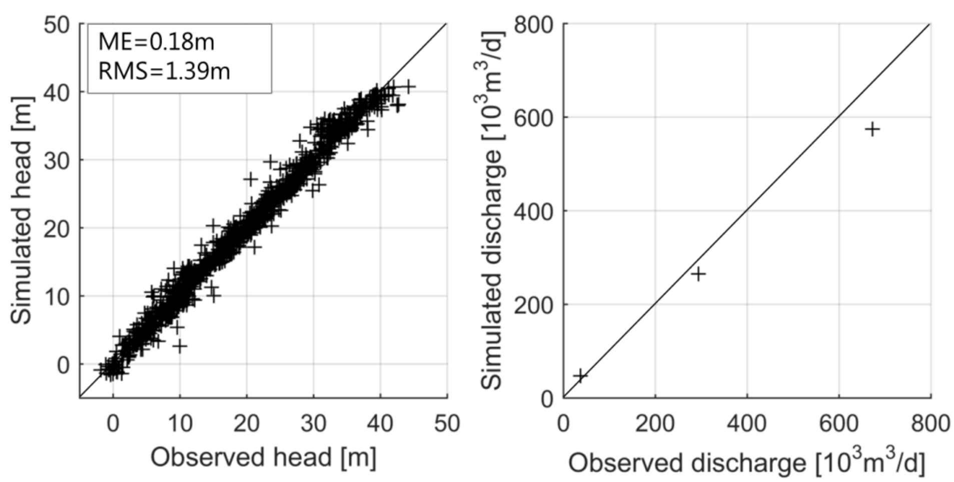
"No-flow boundaries were used along flow lines in the north and south and the water divide in the east and at the bottom, where the Palaeogene clay constitutes the base of the aquifer system."

<u>Page 8</u>

RC#7: What about the assessment of the existing model calibration? Could it be mentioned briefly here so the reader knows how good the model is?

AC#7: We add the calibration results by Meyer et al. (2018) in a new figure 3 and change the sentence to:

"The steady-state MODFLOW flow solution (calibration results summarized in Figure 3) forms the basis for the advective transport simulation using MODPATH."



"Figure 3: Calibration results of steady-state groundwater flow model that forms the basis for the advective transport model (modified after Meyer et al. (2018a). Left: simulated versus observed hydraulic head; right: simulated versus observed stream discharge. ME=mean error, RMS=root mean square."

<u>Page 9</u>

RC#8: Table 2 shows results; maybe save those for the result section.

AC#8: We move table 2 to the result section (section 4.2 calibration results).

<u>Page 11</u>

RC#9: What is the explanation for the porosities, i.e. are they reasonable given the description of each formation. Why do the clay units have relatively small porosities? Probably the estimated porosity is an effective transport porosity; this should be noted.

AC#9: The reviewer is right, we are estimating effective porosities that is the reason why the estimated effective porosities for clay are relatively small. We check throughout the manuscript and specify where it is missing.

RC#10: or of structural error in the number and boundaries of zones chosen, boundary conditions, ect. – many more possible causes than unsimulated heterogeneity

AC#10: The reviewer is right. We extended the explanation of the mismatches to:

"…mismatches can be, e.g., a result of small scale heterogeneity below grid resolution, errors in the model structure or uncertainties of parameter."

RC#11: clay typically has a larger porosity than sand

AC#11: This is right for the total porosity. We are dealing with effective porosities (see also AC#9)

RC#12: does

AC#12: Misspelling corrected. "does"

<u>Page 12</u>

RC#13: It seems that well C has several screen segments with short pathlines that should produce short travel times. It's not clear that only some results are excluded.

AC#13: We add screen numbers to be more precise on which wells were used for calibration.

"As mentioned above, only 14 C observations with an activity higher than 5pMC (Table 1) were used, which excludes results from well screens C1, C2, C3, D1, D2, F1 and F2."

<u>Page 13</u>

RC#14: A little more discussion on how SV are applied and interpreted here.

AC#14:

SVD operates on the sensitivity matrix, the Jacobian that relates parameters to observations, and divides the parameter space into a solution space and a null-space. Hereby parameters that are informed by the

observations are put in the solution space while those not informed by observations fall in the null-space. The truncation between these spaces is user-defined and should be at a level where observations do not further constrain parameters. The advantage of using SVD is that the number of estimated parameters is reduced and hence the inverse problem is well-posed. If too many singular values are included, the problem will be still ill-posed. If too few, the model fit might be unnecessarily poor. Singular values are ordered in a decreasing manner, meaning that a singular value of index 1 is more constrained by information contained in the observations than a singular value of index 2 (Anderson et al., 2015). In our study we truncated the SV at index 5 and in Figure 4 b) the identifiability of the parameters based on the SV index 5 is shown. For more details on singular value decomposition refer to, e.g., Anderson et al. (2015), Doherty and Hunt (2009) or Doherty (2015).

RC#15: does this mean that estimated porosities for clay units are not reliable?

AC#15: The reviewer is right, that the estimated effective porosities with a higher identifiability are more reliable, because they are constrained by the observations, compared to those with a small identifiability. However, it does not necessarily mean that the estimates with a low identifiability are unreliable. They are rather more dependent, or constrained, on the regularization and hence on the expert knowledge than by the observations.

RC#16: Consider color-coding the well designations on FIgs 5 and 6 and Table 3. This will make it easier to compare the information on each of these.

AC#16: We have considered color coding as suggested, but we think it is more confusing. The well screens are all numbered throughout the figures and the tables, which allow comparison easily.

RC#17: One problem with this type of plot is that some of the data are always obscured. Consider plotting each subplot on one or more 2D graphs.

AC#17: We changed the graph to a 2D normalized frequency distribution based on the former histograms.

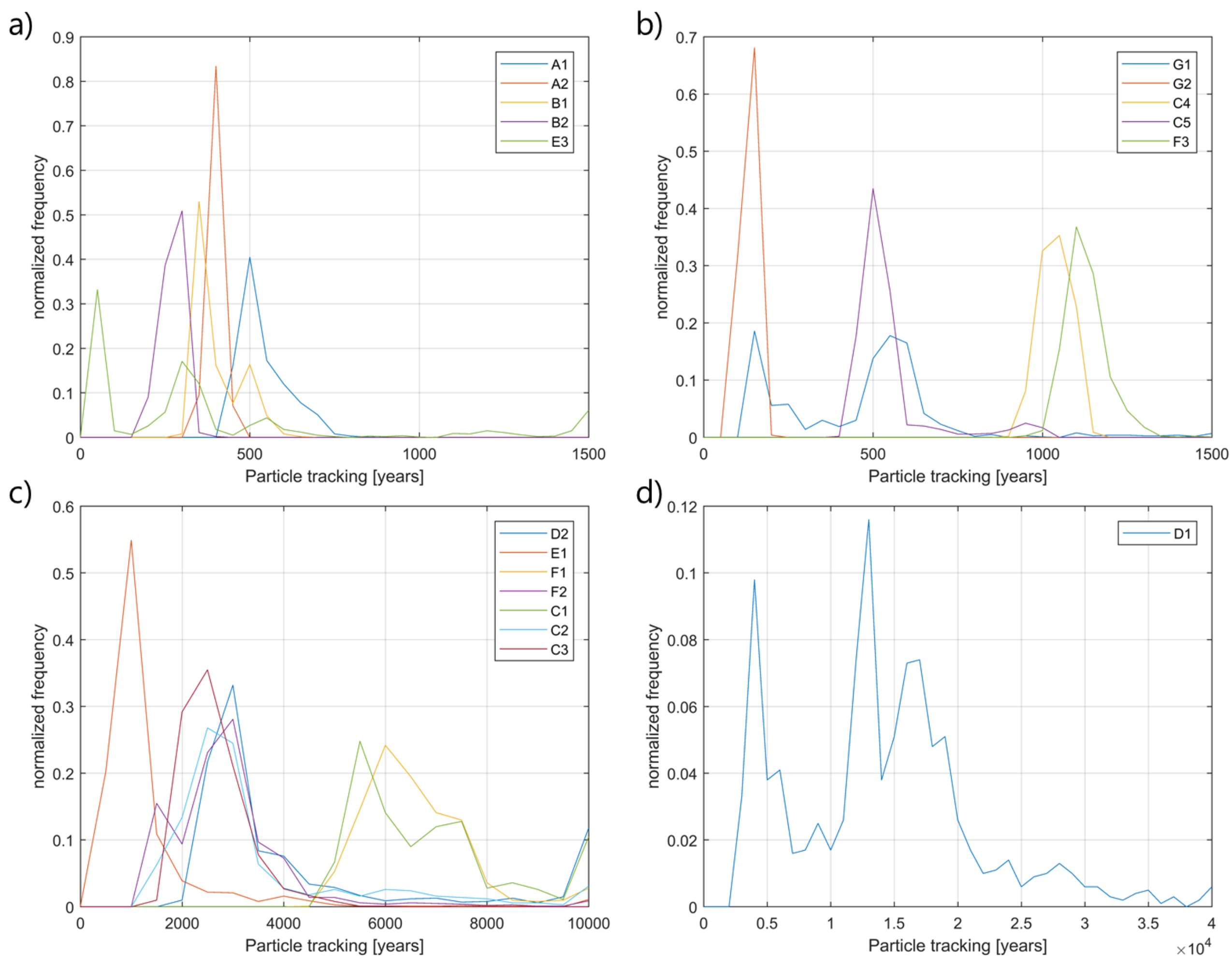


Figure 7 (before 6): Particle age distributions at sampling wells A-G (see Figure 1 for locations). a) and b) young waters (bin size = 50 years) show a narrow, unimodal distribution; c) old waters (bin size = 500 years) have broader and often multimodal distributions; d) multi-modal age distribution at sample location D1 (bin size = 1000 years), which shows the longest travel times.

<u>Page 15</u>

RC#18: Consider shading as Table 2 to show which samples were used in model calibration.

AC#18: We changed the shading as suggested.

RC#19: mean [this relation does not hold for the median]

AC#19: We specified the description to:

"The younger waters (mean age <1000 years) …"

RC#20: Consider using horizontal and vertical lines to show your thresholds of 1000 years and 10 and 20km path lengths.

AC#20: We have considered this. But if we do so, these lines should be on each subfigure. The axes of the subfigures are chosen to show best the distribution of the data. For some of the subfigures (wells A, B, F, G) these lines would be outside the figure, therefore we choose not to add these lines.

Page 16

RC#21: If you use alpha=0, you could have used particle tracking. This would avoid the complication of numerical dispersion and would allow you to talk about higher moments of the travel time distribution.

AC#21: We set alpha = 0 because the physical dispersion which we still have probably in a range of a few meters is overruled by numerical dispersion (see also AC#14 to comments by reviewer 1). We used particle tracking for the calibration at the sampled well location. But in order to get an idea of the age distribution in the entire model (1.2 mio cells) it was not feasible to produce ravel time distributions of 1000 particles in each cell (as we did for the cells where we had tracer samples). This is why we chose the direct mean age simulation to visualize the age structure in the entire aquifer system.

RC#22: Be clear this is mean age here and elsewhere in this section. Also, consider use the term travel time instead of age

AC#22: We checked the consistency and added 'mean' when it was missing. We considered using the term 'travel times'. To preserve the comparability between particle age and tracer-based apparent ages we chose the term 'mean age' instead of travel time.

Page 18

RC#23: Consider explicitly explaining why section e and f are different at the western boundary.

AC#23: We thank for this remark and add a detailed explanation of the two cross-sections and their differences.

"The two cross-sections e) and f) (Figure 10) differ in their geological connection to the sea-boundary (compare geological sections g) and h) Figure 10). In e) a buried valley connects the inland aquifer with the sea and here younger waters reach further inland due the relatively higher hydraulic conductivity and the inland head gradient as a result of the drainage system. Moreover, buried valleys constitute locations where the deep aquifer system, bearing old waters, connects with the shallow one and here upwelling of older water occurs due to the higher heads in the deep semi-confined (by the Maade aquitard) Miocene aquifer. In cross-section f) where the buried valley occurs further inland, the young ocean water penetrates the higher

permeable Miocene aquifer but is impeded in the low permeable sections and hence does not reach as far inland."

Page 19

RC#24: Be consistent with color schemes across all figures – that help the reader understand your points easily. Considering a 1000 year line instead of 100 because 1000 years is used in the discussion.

AC#24: We chose to have a different color scale on a) in order to better resolve the younger ages close to the surface. In order to prevent misinterpretation we add an explanation to the caption of the figure.

"Be aware that the color scheme in a) is different in order to better resolve younger ages close to the surface."

We chose the 100 year line because this is approx. the time span over which human activity (e.g. contamination with fertilizers) heavily started and contaminated groundwater might be expected. This is on what we base our interpretation and discussion groundwater quality and quantity issues on (section 5.3.)

Page 20

RC#25: It would be worth noting that regardless of human actions, stresses have not been steady over that period, either climate, sea level, or within earth's crust. If it takes that long to reach equilibrium under steady stress, the system is never in steady state.

AC#25: We thank for this remark. The reviewer is right, the system is over this period never in steady state. A similar remark was given in the short comment. We add the sentence here and further discuss this in the discussion.

"The steady state distribution of direct simulated mean groundwater age was reached after ~26000 years. Over this time span the system has been exposed to transient stresses from human activity, climatic changes (glacial cover, sea level, ect.). Therefore, the steady-state assumption is a notable simplification, which is further discussed in section 5.1."

RC#26: Review the porosity of Maade and how it was determined.

AC#26: The porosity of the Maade was estimated as 'Pleistocene clay (Maade formation)' e.g. Table 2 or Figure 4 (the new Figure 5).

RC#27: That seems to be older than what the pdf indicates.

AC#27: We have checked the mean groundwater ages derived from a moment analysis and the shown pdfs again. They are correct.

Page 21

RC#28: The direct ages are a function of the age mass of the volume of the model cell whereas the advective ages are a function of the well screen position within the cell. You wouldn't necessarily expect them to match.

AC#28: We add a section about the commensurability to the discussion. Here we discuss the differences in observed tracer ages, particle-based simulated ages and directly simulated ages.

"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 dispersivity was set to zero), which are included in the direct age approach, but neglected in simulating advective ages."

Page 22

RC#29: the dashed lines are not clear on these maps.

AC#29: We enlarged the figure, now the lines are better visible.

Page 23

RC#30: not clear what you mean by 'general behavior of the voxel system.' Maybe this could be reworded, for example, "properties averaged over hydrogeological units".

AC#30: We thank the reviewer for the suggestion and changed the sentenced accordingly to:

"The geology is highly complex and aquitard thickness and porosity distribution change spatially over the entire region, whereas the correction terms were based on the properties averaged over hydrogeological units."

Page 24

RC#31: Particle tracking can also be used to calculate flux-weighted residence times. The difference is how you choose to weight particles, whether by volume or by flux.

AC#31: To prevent confusion, also based on comments by reviewer 1, we decided to take this part out of the manuscript (see also AC#25 and 26 to reviewer 1)

RC#32: You can also assign weights to particles based on flux, which would give you a more comparable age to the direct method. You still have the difference that particles placed in a well screen have limited spatial distribution compared to those in a model cell.

AC#32: see AC#31.

References added:

Anderson, M., Woessner, W.W., Hunt, R., 2015. Applied Groundwater Modeling: Simulation of Flow and Advective Transport, 2nd ed. Elsevier. https://doi.org/10.1016/B978-0-08-091638-5.00001-8

Doherty, J., 2016. Model-Independent Parameter Estimation II. Watermark Numer. Comput. 217.

Doherty, J., 2015. Calibration and uncertainty analysis for complex environmental models. Wartermark Numerical Computing.

Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. J. Hydrol. 366, 119–127. https://doi.org/10.1016/j.jhydrol.2008.12.018

Starn, J.J., Green, C.T., Hinkle, S.R., Bagtzoglou, A.C., Stolp, B.J., 2014. Simulating water-quality trends in public-supply wells in transient flow systems. Ground Water 52, 53–62. https://doi.org/10.1111/gwat.12230