Dear Editor,

we thank you for your helpful comments. You find our responses (AC) in **bold, changes in quotation marks (" ")** and your comments (EC) in *italic*.

Kind regards,

Rena Meyer, on behalf of the authors

EC#1

Comments to the Author:
Thank you for your thoughtful and fairly comprehensive responses to reviewers' comments. I believe your revised paper is much improved and will be of significant interest to the hydrologic community. In reading all the reviews, your responses and the revised manuscript, however, I have two concerns that I believe you can address fairly easily. (1) The statement on p. 19 line 9: "This does not imply that physical dispersion does not exist, only that physical dispersion is accounted for by numerical dispersion." is problematic because you do not even assert that the magnitude of numerical dispersion might similar to that of the sub-grid scale (200 m) dispersion. It sounds like you hope that the numerical dispersion will mimic the real thing, but you do not really know. Unless you have some evidence that the numerical dispersion might fortuitously mimic the real thing, I suggest you remove this sentence and discussion of sub-200-m scale dispersion, and rely on the argument that your resolution of 200-m scale heterogeneity accounts for the dominant dispersion on the scale of this problem and leave it at that. You could cite Weissmann et al. (2002) on this, as well as LaBolle & Fogg (2001) which is cited by Weissmann. The obvious fix would have been to use the more accurate MOC or TVD solution schemes in MT3DMS, and apply actual grid-scale dispersivities, rather than hoping numerical dispersion will cover it for you. If those methods created some other problems, such as excessive execution times, making them impractical, you should say so in the methods.

AC#1:

We agree and removed the sentence and instead rely on the argument that on our modelling scale dispersion is dominated by facies-scale heterogeneities that are accounted for in the geological model in a scale of 200m to 400m. We add two sentences on page 19:

"Similar to Weissmann et al. (2002) and LaBolle and Fogg (2001) the simulations showed little sensitivity to local scale dispersivity because at the modelling scale of tens of kilometers, dispersion is dominated by facies-scale heterogeneity which is captured by the detailed, highly resolved geological model."

"Choosing the TVD or MOC solver scheme for the advection-dispersion equation would have been more accurate in terms of less numerical dispersion, but would have required excessive running times which made it impractical to use in this study."

EC#2:

(2) As suggested by reviewer 2, the low clay porosities that you estimated will look quite implausible to most reviewers. A true clay will indeed always have higher porosity than a sand, unless perhaps the clay has been heavily lithified, say by compaction due to glacial loading, in which case they may be more like claystone formations than clay. Another possibility is that the clays have heterogeneities within, possibly including fractures, that provide preferential flow paths that would increase the apparent effective porosity. So if it really is plausible for the clay effective porosities to be that low, you should explain it through geologic arguments concerning the actual nature of those clays. My first thought, however, was that the calibration was forced to dramatically lower the clay porosity to compensate for too-high K values from your flow model calibration. The latter is also quite plausible because calibrating aquitard K values in a flow model is tough unless you have really good 3D head data. At this point, I just suggest you give one or two geologically plausible explanations for how the clay porosity could be so low.

AC#2: As we replied to reviewer 2, we calibrated effective porosities, defined as the pore space which allows the fluid to travel through. The effective porosity can be smaller for clay than for sand (e.g. Hölting and Coldewey, 2013, page 13 Fig. 4) because the pores might be less connected and the water adhesive to the clay minerals. However, we agree that the small effective porosity of the Miocene clay of deep marine origin might be due to compaction as a result of glacier load during several glaciations.

Moreover, we cannot rule out that there might be some compensation of the porosity values due to uncertainties in k values because we treat the flow and the advective transport calibration independently.

We add the argument to page 14:

"The relatively small effective porosities for clay units might be due to compaction as a result of glacial loading in the course of several glacial periods during the Pleistocene."

References

Hölting, B., Coldewey, W.G., 2013. Hydrogeologie, 8th ed. Springer-Verlag, Berin, Heidelberg. https://doi.org/10.1007/978-3-8274-2354-2

LaBolle, E.M., Fogg, G.E., 2001. Role of Molecular Diffusion in Contaminant Migration and Recovery in an Alluvial Aquifer System. Transp. Porous Media 42, 155–179. https://doi.org/10.1023/A:1006772716244

<u>**Reply to Reviewer 1**</u>

Dear Reviewer 1,

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

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 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).

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).

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 <u>3D</u> 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, <u>Bethke</u> 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. 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.

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."
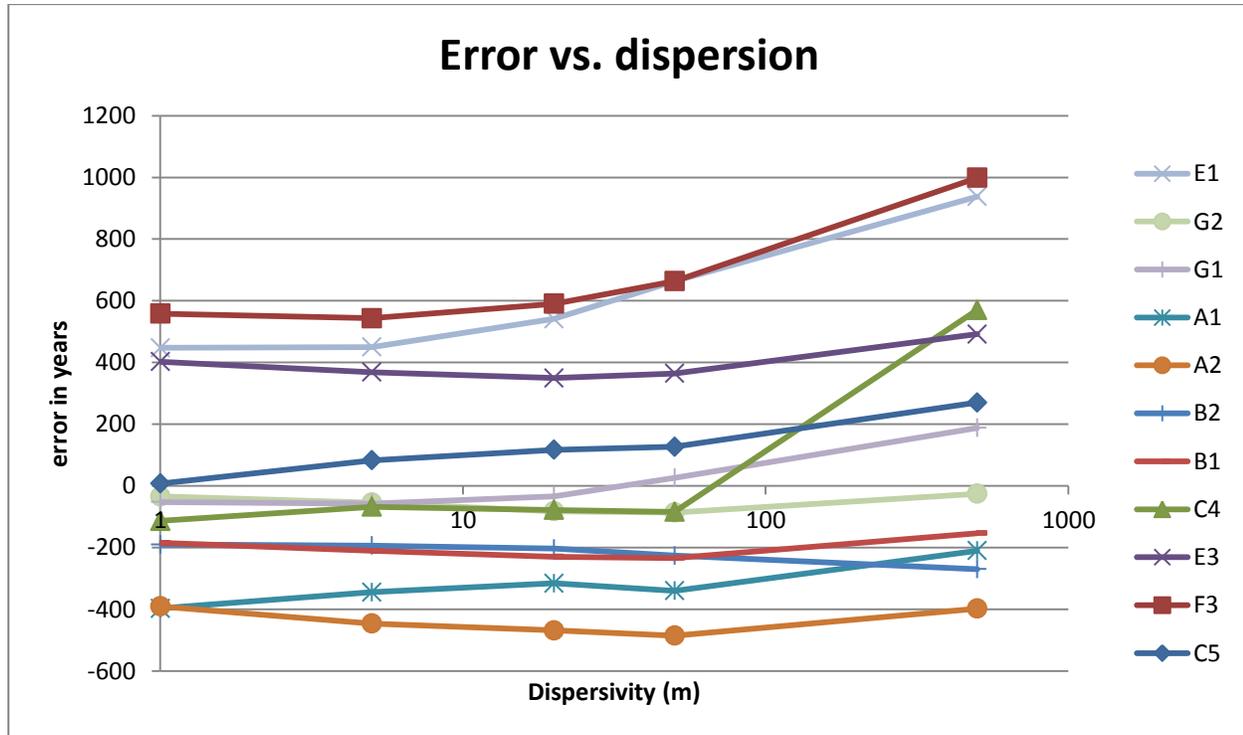
RC#19:

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 extra, but unknown uncertain. 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 (laboratory uncertainty and dilution-diffusion-correction)." (page 24)

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 timeaRmarching 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

<u>**Reply to reviewer 2**</u>

Dear Reviewer 2,

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

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-

<u>Specific comments:</u>

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

<u>Page 2:</u>

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.

<u>Page 3:</u>

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.

<u>Page 4:</u>

RC#3: what method? Should be stated here

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

<u>Page 6</u>

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).

Page 7

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."

Page 8

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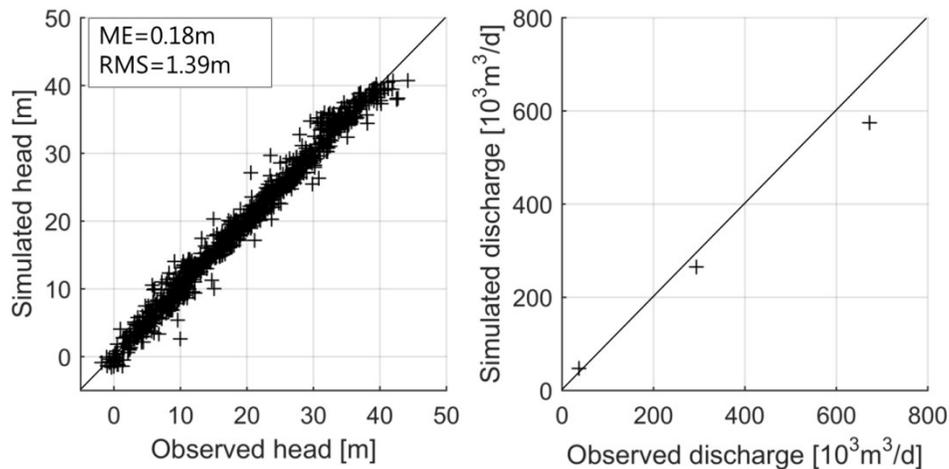
"Figure 3: Calibration results of steady-state groundwater flow model that forms the basis for the advective transport model (modified after Meyer et al. (2018a). Left: simulated versus observed hydraulic head; right: simulated versus observed stream discharge. ME=mean error, RMS=root mean square."

Page 9

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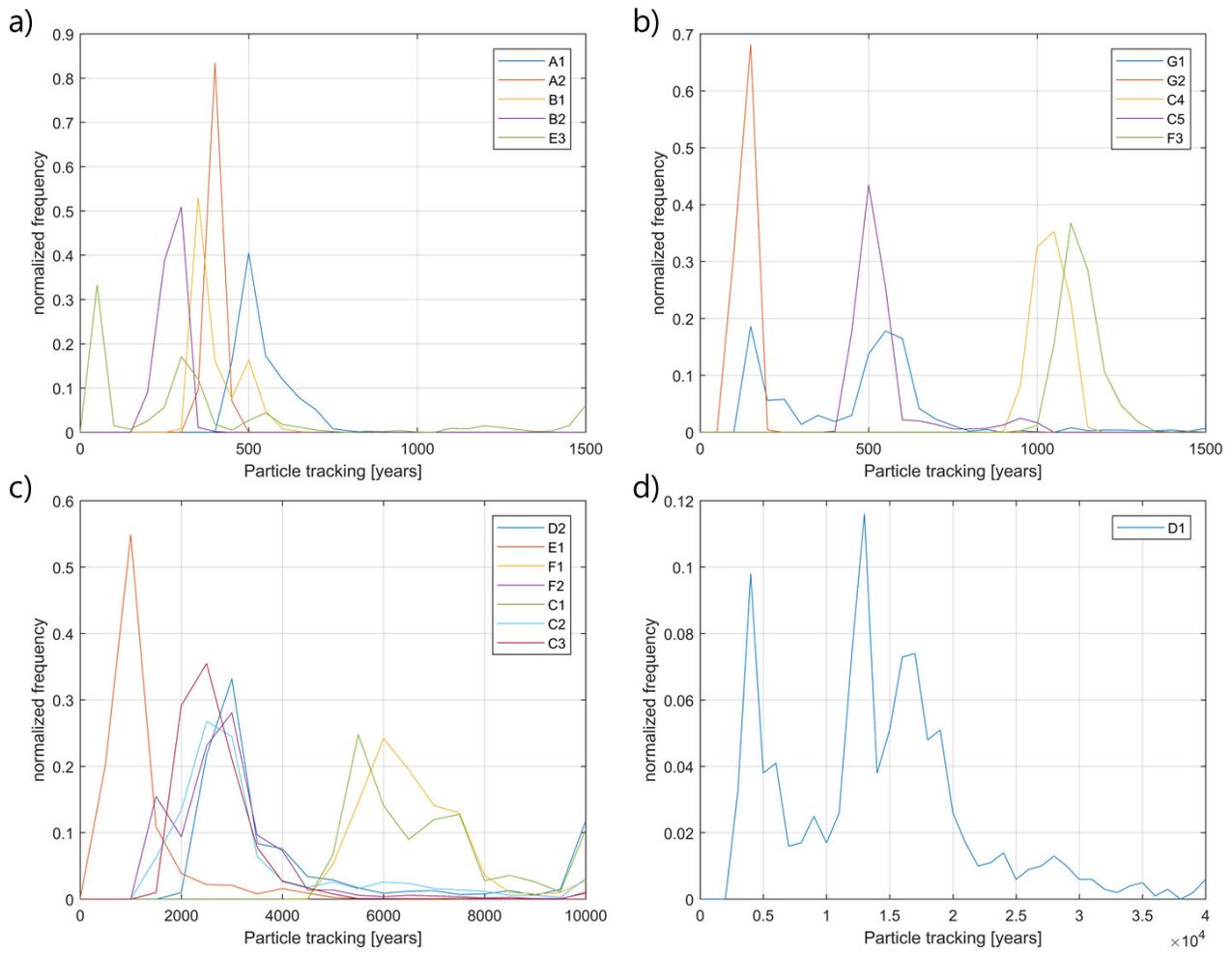
Figure 7 (before 6): Particle age distributions at sampling wells A-G (see Figure 1 for locations). a) and b) young waters (bin size = 50 years) show a narrow, unimodal distribution; c) old waters (bin size = 500 years) have broader and often multimodal distributions; d) multi-modal age distribution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

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."

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.)

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.

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."

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

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

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."

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

Marked up manuscript

Estimation of effective porosity in large-scale groundwater models by combining particle tracking, auto-calibration and ^{14}C dating

Rena Meyer[1], Peter Engesgaard[1], Klaus Hinsby[2], Jan A. Piotrowski[3], and Torben O. Sonnenborg[2]

[1]University of Copenhagen
[2]Geological Survey of Denmark and Greenland
[3]Aarhus University

Correspondence: Rena Meyer (reme@ign.ku.dk)

Abstract. Effective porosity plays an important role in contaminant management. However, the effective porosity is often assumed constant in space and hence heterogeneity is either neglected or simplified in transport model calibration. Based on a calibrated highly parametrized flow model, a three-dimensional advective transport model (MODPATH) of a 1300 km^2-large coastal area of southern Denmark and northern Germany is presented. A detailed voxel model represents the highly heterogeneous geological composition of the area. Inverse modelling of advective transport is used to estimate the effective porosity of seven, spatially distributed units based on apparent groundwater ages inferred from 11 ^{14}C measurements in Pleistocene and Miocene aquifers, corrected for the effects of diffusion and geochemical reactions. By calibration of the seven effective porosity units the match between the observed and simulated ages is improved significantly resulting in a reduction of ME of 99% and RMS of 82% compared to a uniform porosity approach. Groundwater ages range from a few hundred years in the Pleistocene to several thousand years in Miocene aquifers. The advective age distributions derived from particle tracking at each sampling well show unimodal (for younger ages) to multimodal (for older ages) shapes and thus reflect the heterogeneity that particles encounter along their travel path. The estimated effective porosity field, with values ranging between 4.3% in clay and 45% in sand formations, is used in a direct simulation of distributed mean groundwater ages. Although the absolute ages are affected by various uncertainties, a unique insight into the complex three-dimensional age distribution pattern and potential advance of young contaminated groundwater in the investigated regional aquifer system is provided, highlighting the importance of estimating effective porosity in groundwater transport modelling and the implications for groundwater quantity and quality assessment and management.

1 Introduction

The age of groundwater, i.e. the time elapsed since the water molecule entered the groundwater (Cook and Herczeg 2000; Kazemi et al. 2006) is useful (i) to infer recharge rates (e.g. Sanford et al., 2004; Wood et al., 2017) and hence to sustainably exploit groundwater resources, (ii) to evaluate contaminant migration, fate and history (Bohlke and Denver 1995; Hansen et

1

al. 2012) and predict spread of pollutants and timescales for intrinsic remediation (Kazemi et al. 2006), (iii) to analyze aquifer vulnerability or protection to surface-derived contaminants (e.g. Manning et al. 2005; Bethke and Johnson, 2008; Molson and Frind, 2012; Sonnenborg et al., 2016) and indicate the advance of modern contaminated groundwater (Hinsby et al. 2001a; Gleeson et al. 2015; Jasechko et al. 2017) and groundwater quality in general (Hinsby et al. 2007), and (iv) to contribute to the

5 understanding of the flow system, e.g. in complex geological settings (Troldborg et al. 2008; Eberts et al. 2012).

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

10 and Ginn, 2009). Here, we focus on 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).

Particle-based advective groundwater age calculation utilizing travel time analysis is computationally easy, but neglects diffusion and dispersion. The full advection-dispersion transport simulation of a solute or an environmental tracer is computa-

15 tionally expensive and limited to the specific tracer characteristics (McCallum et al. 2015, Salmon et al. 2015), but accounts for diffusion, dispersion and mixing. The tracer independent direct simulation of groundwater mean age (Goode 1996; Engesgaard and Molson 1998; Bethke and Johnson 2002) includes advection, diffusion and dispersion processes and yields a spatial distribution of mean ages. A comparison of ages simulated using any of these methods with ages determined from tracer observations, referred to as apparent ages, is desirable as it can improve the uniqueness in flow model calibration and

20 validation (Castro and Goblet 2003; Ginn et al. 2009) and it potentially informs about transport parameters such as effective porosity, diffusion and dispersion, that are otherwise difficult to estimate. However, the approach is far from straight forward as environmental tracers undergo non-linear changes in their chemical species (McCallum et al. 2015) and groundwater models only represent a simplification and compromise on structural and/or parameter heterogeneity. In a 2D synthetic model, McCallum et al. (2014) investigated the bias of apparent ages in heterogeneous systems systematically. McCallum et al. (2015)

25 applied correction terms, e.g. diffusion correction for radioactive tracers, on apparent ages to improve the comparability to mean advective ages. They concluded that with increasing heterogeneity the width of the residence time distribution increases and that apparent ages would only represent mean ages if this distribution is narrow and has a small variance. 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,

30 which will give the kinematic age.

Flow and transport parameters such as hydraulic conductivity, conductance of streambeds and drains, recharge and dispersivities have gained more and more focus in calibration of groundwater models, recently also on large scales, where the combination of head, flow and tracer observations are widely used as targets (McMahon et al. 2010). However, effective porosity has not received nearly as much attention and especially its spatial variability is often neglected, except for Starn et al. (2014). The

35 lack of focus on calibrating distributed effective porosity on a regional scale might be related to the common assumption that

recharge in humid climates can be precisely estimated and porosity of porous media is relatively well known from literature (Sanford 2011). 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). Thus, the groundwater age exchange is directly related to the porosity. Yet, the calibration of a spatially distributed effective porosity field and its application to simulate groundwater ages and infer capture zones has not gained much attention.

The uniqueness of the presented study lies in the calibration of a three-dimensional, spatially distributed, effective porosity field in a regional-scale complex multi-layered heterogeneous coastal aquifer system. The aim is (i) to use apparent ages inferred from dissolution- and diffusion-corrected ^{14}C measurements from different aquifer units as targets in auto-calibration with PEST of seven unit-specific effective porosities in an advective (particle tracking) transport model ~~A particle-based simulation scheme (MODPATH) was evaluated as suitable in terms of the computational time while neglecting dispersion effects seemed to be acceptable at large scale using radiogenic old-age tracer (^{14}C) (Sanford 2011)~~; (ii) to assess the advective age distributions at the sampling locations to obtain information on the age spreading; (iii) to apply the estimated seven effective porosities in a direct age simulation (Goode, 1996) to gain insight in the three-dimensional age pattern of the investigation area and (iv) to assess the effect of using the heterogeneous effective porosity model compared to a homogeneous effective porosity model for differences in capture zones via particle back-tracking, which is a water management approach to define wellhead protection areas or optimize pump-and-treat locations for remediation of pollution (Anderson et al. 2015).

2   Study area

The $1300\,km^2$-large investigation area is located adjacent to the Wadden Sea in the border region between southern Denmark and northern Germany (Fig. 1). During the Last Glacial Maximum (LGM; 22 ka to 19 ka ago, Stroeven et al. 2016), the area was the direct foreland of the Scandinavian Ice sheet. The low-lying marsh areas (with elevations below mean sea level) in the west were reclaimed from the Wadden Sea over the last centuries and protected from flooding by a dike for the last $\approx 200\,a$. A dense network of drainage channels keeps the groundwater level constantly below the ground surface, thus, mostly below sea level. The water divide near the Jutland ridge with elevations of up to 85 m a.s.l. defines the eastern boundary of the study area.

The aquifer systems are geologically complex and highly heterogeneous spanning Miocene through Holocene deposits. The bottom of the aquifer system is defined by low-permeability Palaeogene marine clay. The overlying Miocene deposits consist of alternating marine clay and deltaic silt and sand (Rasmussen et al. 2010). The Maade formation, an upper Miocene marine clay unit, with a relatively large thickness in the west while thinning out to the east, is located below the Pleistocene and Holocene deposits. Buried valleys filled with glacial deposits, mainly from the Saalian glaciation, cut through the Miocene and reach depths up to 450 m below surface. They are important hydrogeological features as they may constitute preferential flow paths and locally connect the Pleistocene and Miocene aquifers.

In our previous studies (Jørgensen et al., 2015; Høyer et al., ~~2016a~~2017; Meyer et al., 2018a), the available geological and
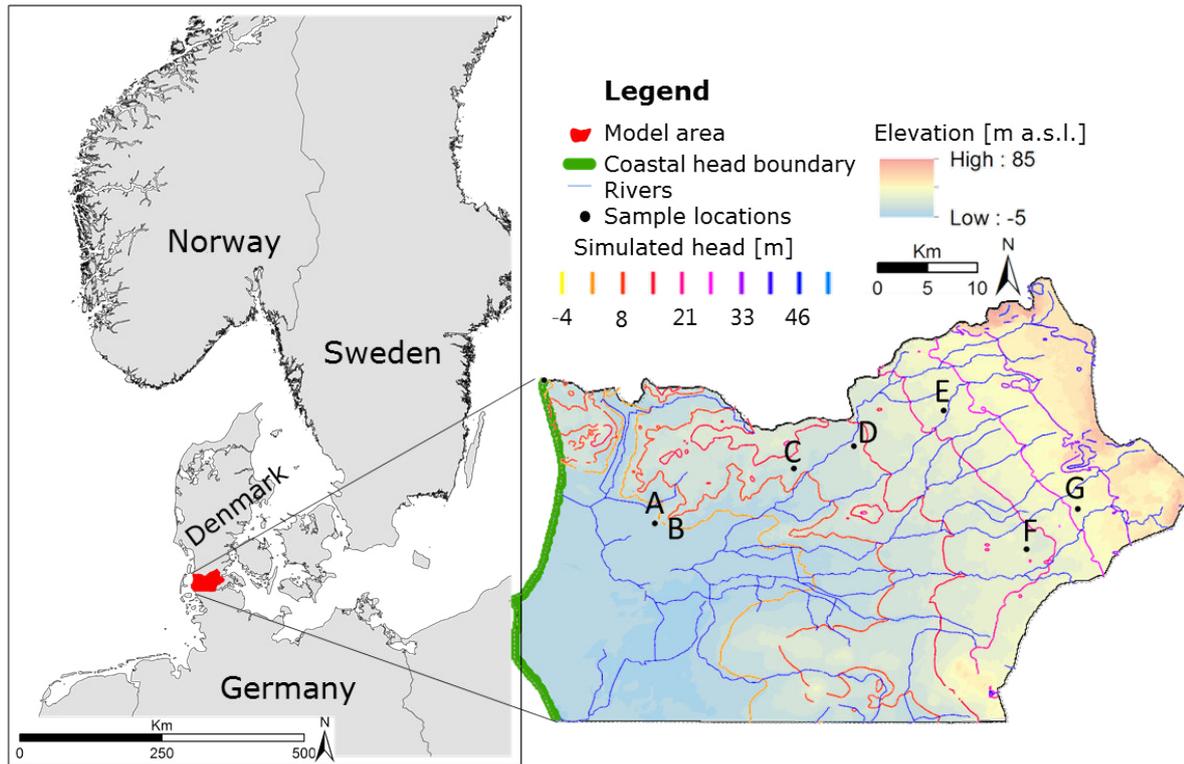
Figure 1. Investigation area at the border between Denmark and Germany. Simulated hydraulic heads are from the shallow aquifer (Meyer et al. 2018a). Topography, ^{14}C sample locations (A-G), river network and coastal head boundary are indicated.

geophysical information including borehole lithology, Airborne Electro Magnetic (AEM) and seismic data were assembled into a heterogeneous geological voxel model comprising 46 geological units with raster sizes of ~~100 x 100~~ 100 m x 100 m x 5 m. Manual and automatic modelling strategies, such as clay fraction (CF), multi-point simulation (MPS) and cognitive layer approach, were complementarily applied. Meyer et al. (2018a) investigated the regional flow system and identified the most
5 dominant mechanisms governing the flow system comprising geological features and land management that are visualized in a conceptual model in Figure 2. Extensive clay layers separate the Miocene and Pleistocene aquifers, buried valleys locally cut through the Maade formation and connect Miocene and Pleistocene aquifers allowing groundwater exchange and mixing. The large drainage network, established in the reclaimed terrain keeping the groundwater table constantly below the sea level, acts as a large sink for the entire area. In the deeper aquifers, significant inflow from the ocean occurs at the coast near the marsh
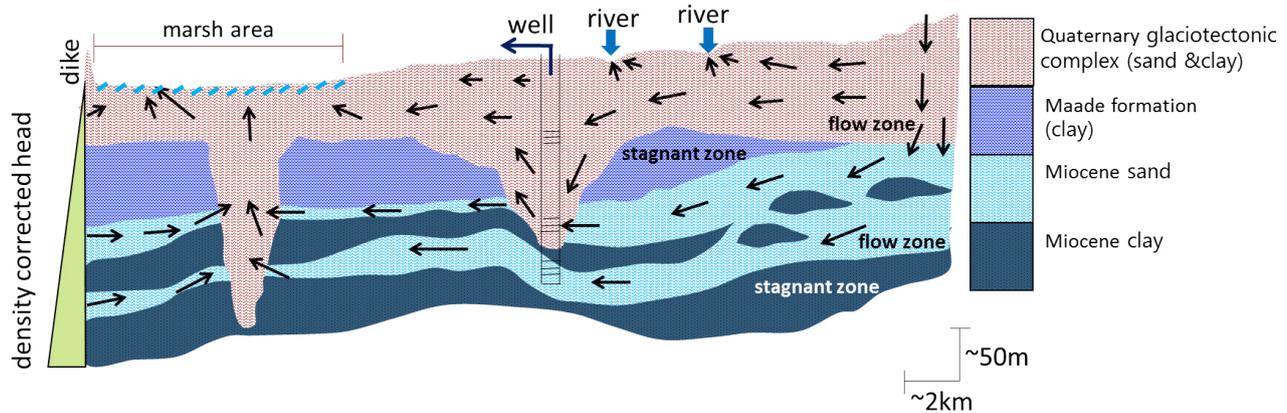10 area as a result of a landward head gradients induced by the drainage.

Figure 2. Conceptual regional model showing a simplified geology featuring buried valleys and groundwater flow and stagnant zones (as used for the diffusion correction). Arrows indicate general flow field of groundwater. Also shown are the boundary conditions, i. e. density corrected coastal boundary, drained marsh area, rivers.

3 Methods

The age simulation and calibration of effective porosities builds upon the calibrated regional-scale groundwater flow model (MODFLOW) of a highly heterogeneous coastal aquifer system by Meyer et al. (2018a). First, advective transport simulation using MODPATH (Pollock, 2012) was used for the calibration of effective porosities of seven different geological units. ^{14}C

5 observations were corrected for carbon dissolution and diffusion and subsequently used as calibration targets during inverse modelling with PEST. 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 from ^{14}C analysis, 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).

10 Secondly, the analysis of advective age distributions at ^{14}C sampling locations provided an insight in the ranges of travel times and distances and hereby the complexity of groundwater age mixing. Thirdly, the estimated effective porosities were used in a direct age simulation (Goode, 1996) in order to investigate the spatial groundwater age distribution in the regional aquifer system. Finally, the impact of using a seven-porosities model compared to a constant porosity model on capture zone delineation at two well locations was assessed.

3.1 ^{14}C measurements

During a field campaign in February 2015, 18 groundwater samples were collected from wells at seven sites with screens at different depths and in different aquifers (Figure 1, Table 1). The wells were pumped clean three times their volume to prevent the influence of mixing with stagnant water. In situ parameters (pH, EC, O2) were measured and after they stabilized
5 samples for radiocarbon analyses were collected in 1-liter opaque glass bottles. The 18 groundwater samples were analyzed for $\delta^{13}C_{\%_o VPBD}$ with an isotope ratio mass spectrometer (IRMS) and for ^{14}C with an accelerator mass spectrometer (AMS) at the AGH University of Science and Technology, Kraków, Poland and in the Poznań Radiocarbon Laboratory, Poznań, Poland, respectively, in September 2015.

3.1.1 ^{14}C correction for dissolution and diffusion

10 The ^{14}C activity (A_m) was measured in the dissolved inorganic carbon (DIC) content of the groundwater. Uncertainties arise from geochemical and hydrodynamic processes that change the ^{14}C content in the aquifer (e.g. Bethke and Johnson, 2008; Sudicky and Frind, 1981). The dissolution of fossil "dead" (^{14}C-free) carbon dilutes the ^{14}C content in groundwater and results in lower ^{14}C concentrations (Appelo and Postma 2005). Diffusion into aquitards also reduces the ^{14}C concentration in the aquifer (Sanford 1997). Both processes reduce the ^{14}C concentration and result in an apparent groundwater age that is
15 older than the true age. Consequently, the measured ^{14}C activities were corrected for carbonate dissolution as well as aquitard diffusion prior to use in the calibration.

A modified chemical correction was applied that takes into account the effect of dissolution as described by Boaretto et al. (1998). This method was successfully used in Danish geological settings similar to those investigated in the present study (equation 2; Boaretto et al. 1998; Hinsby et al. 2001b). The initial ^{14}C activities were corrected for fossil carbon dissolution
20 (Pearson and Hanshaw 1970) assuming an atmospheric ^{14}C activity (A_0) and soil $\delta^{13}C$ concentration in the soil CO_2 of 100 pMC (percent Modern Carbon) and -25‰, a dissolved carbonate concentration of 0 pMC and 0‰. With a decay rate constant (λ) of 1.21×10^{-4} 1/a for the ^{14}C decay, the dissolution-corrected age τ_C was calculated as (e.g. Bethke and Johnson, 2008)

$$\tau_c = \frac{-1}{\lambda} ln(\frac{A_m}{A_0}) \tag{1}$$

25 with

$$A_0 = \frac{\delta^{13}C}{-25} * 100 \tag{2}$$

Subsequently, a diffusion correction was made to take into account diffusion loss into low-permeability layers (Sanford, 1997). Aquitard diffusion is sensitive to porosity, diffusion coefficient and the thicknesses of the active flow (aquifer) and stagnant (aquitard) zones (Sudicky and Frind 1981). Because of the geological complexity, the sand-to-clay ratio based on voxel
30 lithology was used to calculate the relative aquifer/aquitard (a/b = 0.72) thicknesses. Diffusion-corrected groundwater ages were calculated for three different diffusion coefficients: $1.26 \times 10^{-9} \, \text{m}^2/\text{s}$ (Jaehne et al. 1987) representing the CO_2 diffusion

in water, $1 \times 10^{-10}\,\mathrm{m^2/s}$ as an average for clay deposits (Freeze and Cherry 1979; Sanford 1997), and $2.11 \times 10^{-10}\,\mathrm{m^2/s}$ as calculated by Scharling (2011), using aquifer effective porosities (n_e) ranging from 0.16 to 0.35 and aquitard (b) thicknesses between 10 m and 50 m. Based on the ranges of variables, an average corrected age and the corresponding standard deviation were obtained for each sample (Table 1). Corrected groundwater sample age (τ_D), also referred to as the apparent age, was calculated as:

$$\tau_D = \tau_C * \left(\frac{\lambda}{\lambda + \lambda'} \right) \tag{3}$$

with

$$\lambda' = 2 * \tanh\left[\left(\frac{b}{2} \right) * \left(\frac{\lambda}{D} \right)^{\frac{1}{2}} \right] * \frac{(\lambda D)^{\frac{1}{2}}}{n_e a} \tag{4}$$

3.2 Groundwater flow model

Meyer et al. (2018a) simulated the 3D steady state regional groundwater flow using MODFLOW 2000 (Harbaugh et al. 2000). A brief description of the model set up and calibration results are presented here, further details can be found in Meyer et al. (2018a). The model was discretized horizontally by 200 m x 200 m in the west and 400 m x 200 m in the east and vertically by 5 m above 150 m b.s.l. and 10 m below 150 m b.s.l. resulting in 1.2 million active cells. The voxel geology was interpolated to the MODFLOW grid and 46 hydrogeological units were defined. No-flow boundaries were used along flow lines in the north and south, at a water divide in the east and at the bottom, where the Palaeogene clay constitutes the base of the aquifer system. At the western coast a density-corrected constant head boundary was applied (Figure 1; Guo and Langevin 2002; Post et al. 2007; Morgan et al. 2012). Distributed net recharge, averaged over the years 1991-2010 was extracted from the national water resources model (Henriksen et al. 2003) and included as a specified flux condition. Internal specified boundaries included abstraction wells with a total flux of $26 \times 10^6\,m^3/year$ (averaged over the years 2000-2010, corresponding to 4% of the total recharge), rivers and drains.

Horizontal hydraulic conductivities, one for each hydrogeological unit, two anisotropy factors (Kh/Kv), one for sand and one for clay units, as well as river and drain conductances were calibrated, using a multi-objective regularized inversion scheme (PEST; Doherty, 2016a), using head and mean stream flow observations as targets. The resulting head distribution is shown in Figure 1. Horizontal hydraulic conductivities were estimated in a range of $K_h \in [1\,\mathrm{m/d};\, 83\,\mathrm{m/d}]$ for Pleistocene sand units, $K_h \in [0.028\,\mathrm{m/d};\, 0.19\,\mathrm{m/d}]$ for Pleistocene clay units, $K_h \in [0.008\,\mathrm{m/d};\, 0.016\,\mathrm{m/d}]$ for the Maade formation, $K_h \in [16\,\mathrm{m/d};\, 46\,\mathrm{m/d}]$ for Miocene Sand and $K_h \in [0.14\,\mathrm{m/d};\, 0.23\,\mathrm{m/d}]$ for Lower Miocene Clay. The vertical anisotropy factor (K_h/K_v) was estimated to 25 and 85 for sand and clay units, respectively.

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.

Table 1. Sampling wells, uncorrected and corrected groundwater ages. Gray shade indicates samples used for calibration. Note that lower numbers of the wells indicate deeper locations (m b.s. = meter below ground surface, std = standard deviation, pMC = percent Modern Carbon).

well	DGU no.	~~filter~~ screen depth [m b.s.]	aquifer geology	measured ^{14}C [pMC]	uncorrected ^{14}C [years]	$\Delta^{13}C_m$ [‰VDPD]	age corrected for dissolution and diffusion (std)[years]
A1	166.761-1	246-252	Buried valley	46.44	6161	-13.2	344 (59)
A2	166.761-2	204-210	Buried valley	49.95	5576	-13	108 (19)
B1	166.762-1	160-166	Buried valley	49.84	5593	-13.9	293(50)
B2	166.762-2	102-108	Buried valley	51.9	5268	-13.2	46 (8)
C1	167.1545-1	306-312	Buried valley	0.48	42889	-5.9	10429 (1789)
C2	167.1545-2	273-276	Buried valley	1.03	36755	-7.7	9097 (1569)
C3	167.1545-3	215-218	Buried valley	0.16	51714	-11	15038 (2593)
C4	167.1545-4	142-149	Buried valley	33.84	8703	-13.2	1191 (205)
C5	167.1545-5	116-123	Buried valley	43.18	6746	-13.1	518 (89)
D1	159.1335-1	290-295	Miocene	1.8	32271	-7.9	7671 (1323)
D2	159.1335-2	277-282	Miocene	1.35	34582	-10.6	9229 (1591)
E1	159.1444-1	194-200	Buried valley	31.34	9320	-12	1141 (197)
E3	159.1444-3	81-87	Buried valley	40.29	7302	-12.8	642 (111)
F1	168.1378-1	372-378	Miocene	46.12	6216	-12.3	173 (30)
F2	168.1378-2	341-345	Miocene	2.85	28580	-13.3	7836 (1351)
F3	168.1378-3	208-214	Miocene	25.73	10904	-12.6	1800 (310)
G1	168.1546-1	110-120	Miocene	42.57	6860	-12.3	388 (67)
G2	168.1546-2	74-84	Pleistocene/ Miocene	45.33	6355	-12	153 (26)

3.3 Advective transport model

Advective transport simulation was performed using MODPATH (Pollock, 2012) in particle back-tracking mode. Hereby, the travel time of a particle ~~(t)~~, released in a cell, is calculated based on the MODFLOW cell-by-cell flow rates (q). The advective travel time (t) along the travel paths in 3D ($\underline{x}$) is calculated as

$$5 \quad t(\underline{x}) = \int_{\underline{x_0}}^{\underline{x}} \frac{\mathbf{n}_e(\underline{x})}{\mathbf{q}(\underline{x})} d\underline{x} \tag{5}$$
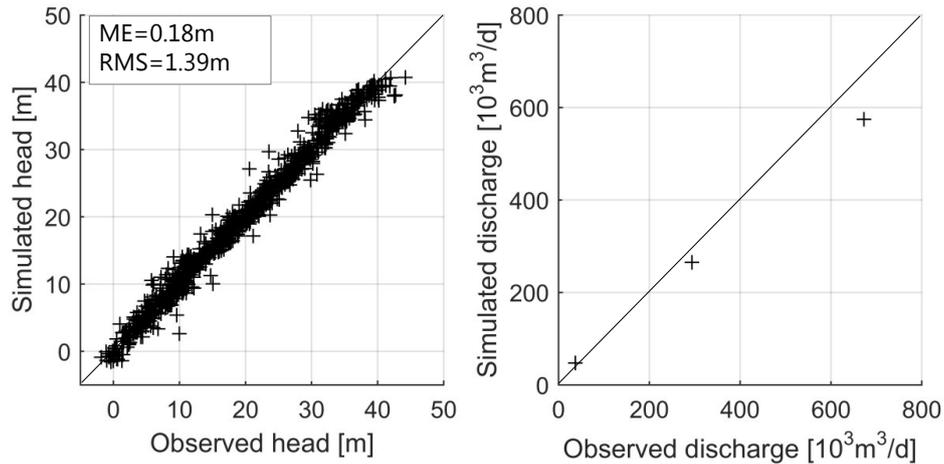
Figure 3. Calibration results of steady-state groundwater flow model that forms the basis for the advective transport model (modified after Meyer et al. (2018a). Left: simulated versus observed hydraulic head; right: simulated versus observed stream discharge. ME=mean error, RMS=root mean square.

In addition to the input data required by MODFLOW to generate the flow solution, MODPATH requires a value for effective porosity (n_e) to calculate the seepage velocity.

The groundwater age can be seen as the backward integration of travel times along the travel path back to its recharge location. Hence, the simulated groundwater age is a function of the ratio of flux to effective porosity and the travel distance. In this study,

5 the total flux is controlled by prescribed recharge and heterogeneous distribution of hydrogeological parameters (e.g. hydraulic conductivity, porosity).

In order to ensure stability (Konikow et al., 2008), 1000 particles were distributed evenly in the cell of the well screen and their average simulated particle age was compared with apparent groundwater ages (derived from equation 3).

The corrected ^{14}C ages were used as targets in the objective function (see below) of the simulated average travel time during

10 calibration. ~~According to Sanford (2011), neglecting hydrodynamic dispersion in advective transport simulations on a regional scale is a reasonable approach when old-age tracers, such as ^{14}C, are used as dispersion might not be crucial for these tracers. On the other hand, diffusion into stagnant zones can create a significant loss in old-age tracer concentration which was taken into account by correcting the ^{14}C (paragraph 3.1.1) before calibration.~~ 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

15 kinematic ages by perturbing the measurement location within the cell reflecting the mixing of waters from different origins. The ^{14}C ages have also been diffusion-corrected (paragraph 3.1.1) so that dilution or mixing due to loss of ^{14}C into the stagnant zones have been accounted for.

3.3.1 Calibrating porosity

The flow solution of the calibrated flow model (Meyer et al., 2018a) constitute the base for the 3D advective transport model. Depending on the depositional environment and clay/sand content, effective porosities of seven units corresponding to two Pleistocene sand, two Pleistocene clay, one Miocene sand and two Miocene clay units, were estimated using regularized (Tikhonov) inversion with PEST (Tikhonov and Arsenin, 1977; Doherty, 2016). As the calibration approach is similar to the one of Meyer et al. (2018a) only additional characteristics are described in the following. Average corrected ^{14}C groundwater ages from 11 samples with a ^{14}C activity higher than 5 pMC (Table 1) were used as calibration targets. ^{14}C activity lower than 5 pMC were not used as it was assumed that the boundary conditions of the flow model (e.g. sea level, recharge, head gradients) were not representative for pre-Holocene conditions. Moreover, the data from well F1 was excluded from calibration as an age inversion with F2 was observed here (Table 1), probably due to local heterogeneity or contamination of water with higher ^{14}C concentration, which is not possible to reproduce by the model. The average uncertainty of apparent ages was estimated to about 102 years. This value was based on the average of the standard deviation of the diffusion correction for the selected 11 samples and was used for weighting of the individual ages.

When Tikhonov regularization is applied, a regularized objective function (Φ_r) is added to the measurement objective function (Φ_m) in form of the weighted least-squares of the residuals of preferred parameter values and parameter estimates. Within the limits of the user-defined objective function (PHIMLIM) and the acceptable objective function (PHIMACCEPT), the weight of the regularized objective function (μ) increases and the parameter estimates are directed towards the preferred values. Calibration settings such as initial and preferred values and final parameter estimates are shown in Table 2. Values for PHIM-LIM and PHIMACCEPT were set to 60 and 100, respectively. The total objective function (Φ_{tot}), minimized by PEST is then the sum of the measurement objective function (Φ_m) and the regularized objective function (Φ_r)

$$\Phi_{tot} = \Phi_m + \mu^2 \Phi_r \tag{6}$$

with

$$\Phi_m = \sum \left(\omega_a \left(a_{obs} - a_{sim} \right) \right)^2 \tag{7}$$

where a_{obs} and a_{sim} are observed and simulated groundwater ages, respectively, and the weight ω_a is the inverse of the standard deviation of the observed age. The calibration is evaluated based on the mean error (ME) and the root mean square (RMS) between apparent (corrected ^{14}C ages) and advective groundwater ages. Parameter identifiability (Doherty and Hunt, 2009) is used to investigate to what extent the effective porosities were constrained through model calibration. Identifiability close to one means that the information content of the observations used during calibration can constrain the parameter. Parameters with an identifiability close to zero cannot be constrained.

3.4 Direct age

To visualize the mean groundwater age pattern in the regional 3D aquifer system, direct simulation of mean groundwater age was performed with MT3DMS (standard finite difference solver with upstream weighting) chemical reaction package using

a zeroth-order production term (Goode, 1996; Bethke and Johnson, 2008). Hereby, mean groundwater age is simulated in analogy to solute transport as an "age mass" (Bethke and Johnson, 2008). For each elapsed time unit (day) the water "age mass" increases by one day in each cell. Increase or decrease of ages is a results of diffusion, dispersion and advection (Bethke and Johnson, 2008). The transient advection-dispersion equation of solute transport of "age mass" in three dimensions and

5 with varying density and porosity is given by Goode (1996)

$$\frac{\partial a n_e \rho}{\partial t} = n_e \rho - \nabla a \rho \mathbf{q} + \nabla n_e \rho \mathbf{D} * \nabla a + F \tag{8}$$

where F is an internal net source of mass age, $\mathbf{q}$ the Darcy flux (m/d), a the mean age (d), n_e the effective porosity, ρ the density of water (kg/m^3) and $\mathbf{D}$ the dispersion tensor (m^2/d), including molecular diffusion and hydrodynamic dispersion. The initial concentration of the "age mass" was set to zero, while a constant age of zero was assigned to the recharge boundary and the

10 constant head boundary at the coast. Steady state conditions were evaluated based on the change in mass storage in a 40000 year simulation. The age mass storage (m) in the whole model was calculated for each time step as the sum of mass in each cell (m_i). The latter was calculated by multiplying the cell dimensions ($\Delta z, \Delta x, \Delta y$) with porosity ($n_e$) and age ($a_s$)

$$m = \sum m_i \tag{9}$$

with

15 $$m_i = \Delta z * \Delta x * \Delta y * n_e * a_s \tag{10}$$

The percentage change in mass storage (Δm_t) per time step (Δt) was calculated as

$$\frac{\Delta m_t}{\Delta t} = \left(\frac{m_t - m_{t-1}}{m_{t-1}} \right) * 100 \tag{11}$$

The integral of the change in mass storage over time was used to define quasi-steady state conditions. This was reached when

$$\int_{t1}^{t} \Delta m(t) dt \geq 0.95 * \int_{t1}^{\infty} \Delta m(t) dt \tag{12}$$

20 Dispersion experiments were carried out for longitudinal dispersivity α_L values of 0 m, 5 m, 20 m, 50 m, 500 m, while the horizontal transversal α_{TH} and vertical transversal α_{TV} dispersivities were specified to 10% and 1% of α_L, respectively. A diffusion coefficient of $1x10^{-9} m^2/s$ was used to account for self-diffusion of the water molecule at about 10 °C (Harris and Woolf 1980).

3.5 Capture zones

25 Well capture zones are used in water management to define areas of groundwater protection, where human actions, such as agricultural use, are restricted. Simulated by the uniform effective porosity and distributed effective porosities model, the capture zones of one existing well (Abild, abstraction rate $27 \, m^3/d$) located in a buried valley and one virtual well (AW, abstraction rate $280 \, m^3/d$) located in a Miocene sand aquifer were evaluated and compared for different back tracking times

using 100 particles per well. ~~No~~ 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. Hence, the capture zone areas at different times were compared.

4   Results

4.1 ^{14}C corrections

Figure 4 shows the corrected and uncorrected ^{14}C ages over depth. Except for well F, ages increase with depth at each multi-screen location. Otherwise, no clear trend between age and depth can be identified on the regional scale. Uncorrected ages range from 5000 years to 50000 years (Table 1). After correction, all ages decrease and the relative difference between the corrected ages increase, now within a range from 46 years to 15000 years. Hence, it is expected that the oldest water recharged the groundwater at the end of the last glacial period. The majority of the samples represent younger waters with 12 out of 18 samples being less than 2000 years.
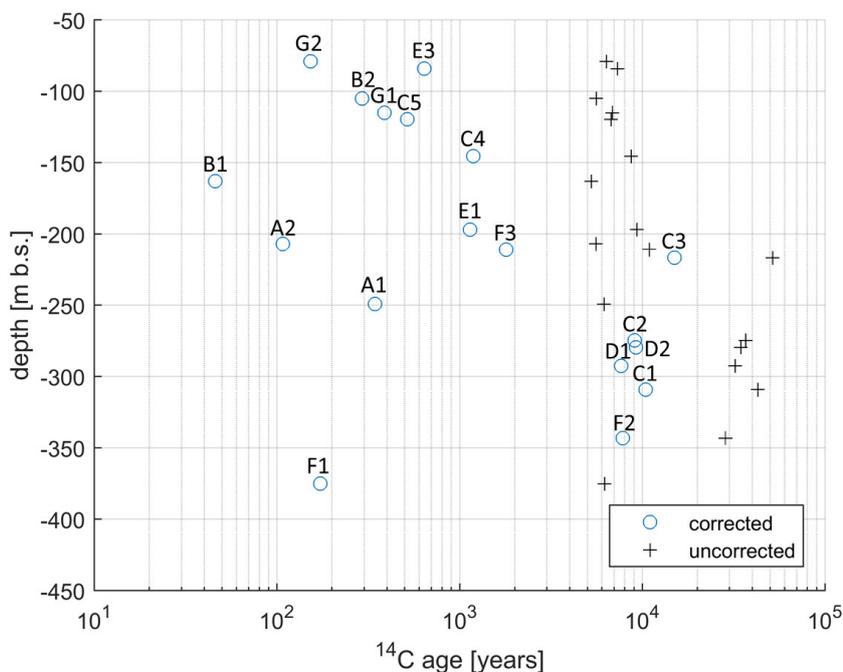
Figure 4. Apparent groundwater ^{14}C ages as a function of groundwater sampling depth: black crosses indicate ages without correction for dissolution and diffusion, blue circles show ages with correction. Labels indicate well location and ~~filter~~ screen number (cf. Figure 1 and Table 1).

4.2 Calibration results

The match between the average of simulated groundwater ages (particle tracking with MODPATH) and corrected ^{14}C ages is shown in Figure 5a.

Results from the distributed effective porosity model were compared to those from a uniform effective porosity model with

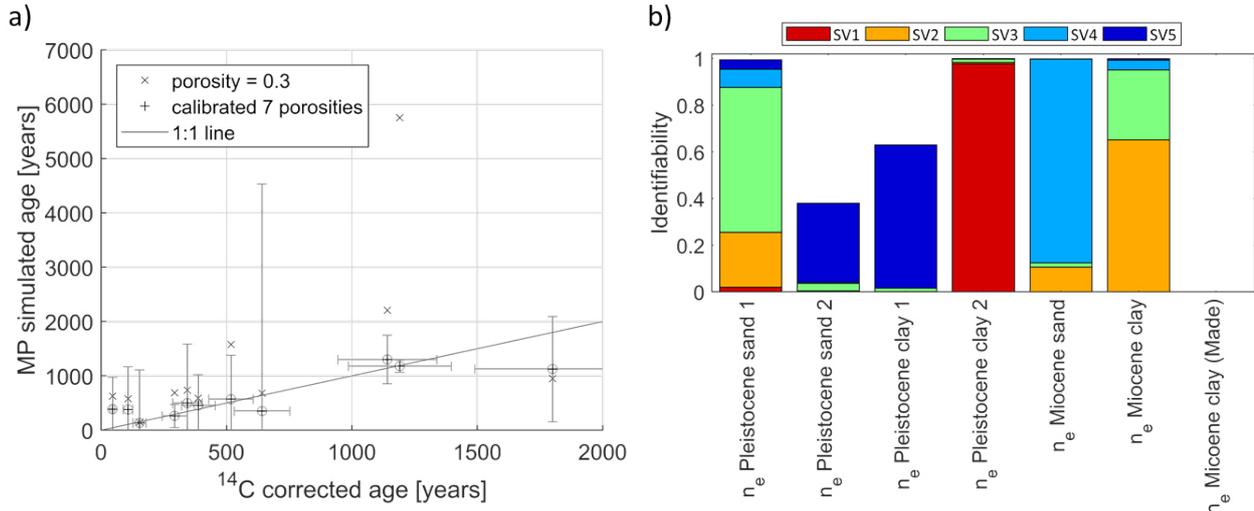**Figure 5.** Calibration results: a) red crosses 'x' show apparent ages simulated with MODPATH (MP) and a porosity of 0.3 as often used in porous media models and blue crosses '+' are MP ages simulated based on the 7 calibrated porosities (Table 2); standard deviations based on MP and correction terms (see section 3.1.1) are shown. b) parameter identifiability of effective porosities (warmer colors correspond to singular values (SV) of a lower index, cooler color to SV of higher index) of the different geological formations; the identifiability of the Maade porosity is close to zero.

5 an effective porosity of 0.3 which is a typical textbook value for porous media (Holting and Coldewey, 2013; Anderson et al., 2015) and often used in groundwater modelling studies (e.g. Sonnenborg et al., 2016). The calibrated distributed effective porosity model is able to match all the observations reasonably. This is not the case for the single effective porosity model where especially one sample is poorly simulated with an estimate of more than 5500 years whereas the corresponding observation only reach about 1200 years. The ME and RMS of the calibrated distributed effective porosity model were -2.3 years

10 and 267 years, respectively, which correspond to a reduction in ME of 99% and RMS of 82% compared to the single effective porosity model. Considering the uncertainties involved in estimation of apparent age, see uncertainty estimates in Table 1, column to the right, the match is found acceptable. Comparing the average uncertainty on apparent ages used for calibration of 102 years with the achieved RMS of 267 years indicate that no overfitting occurred and mismatches can be a result of small scale heterogeneity below grid resolution, errors in the model structure or uncertainties of parameters..

15

Table 2. Calibration settings and results: parameters with initial, preferred and estimated values for effective porosity.

parameter (n_e)	Initial/preferred value	estimated value	% of cells	objective function	
Pleistocene sand 1	0.3	0.130	24.4	PHIMLIM	60
Pleistocene sand 2	0.3	0.263	2.5	PHIMACCEPT	100
Pleistocene clay 1	0.1	0.085	11.6	ϕ_m achieved	74
Pleistocene clay 2	0.05	0.043	4.8		
Miocene sand	0.3	0.450	15.1		
Miocene clay	0.1	0.102	22.8		
Miocene clay	0.05	0.049	18.8		
(Maade formation)					

The estimated effective porosities of the seven hydrogeological units are listed in Table 2. Realistic values are found for all parameters and the values of the sand units are generally higher than those of the clay units. However, the effective porosity estimate of 0.13 for Pleistocene sand 1 is relatively low. This may be explained by the fact that this unit does not represent sand exclusively everywhere. The Pleistocene deposits in the area are highly heterogeneous (Jørgensen et al., 2015) and it is

5 therefore difficult to identify units exclusively composed of sand, partly due to the difficulties in using AEM data to guide the distinction between sand and clay at a relatively small scale. Hence, Pleistocene sand 1 may to some extent represent a mixture of sand and clay. The relatively small effective porosities for clay units might be due to compaction as a result of glacial loading in the course of several glacial periods during the Pleistocene. Additionally, uncertainties in the estimates of hydraulic conductivity from Meyer et al. (2018a) will translate into errors in seepage flux and hence ages. Uncertainties and errors in

10 hydraulic conductivity may therefore be partly compensated by estimates of effective porosity that are somewhat different from the expected value.

The parameter identifiability (Figure 5b) shows that the corrected ^{14}C ages may constrain four out of seven estimated effective porosities, i.e. of Pleistocene sand 1, Pleistocene clay 2, Miocene sand and Miocene clay. The warmer colors (red-yellow) indicate that the parameter is less influenced by measurement noise (Doherty, 2015, Figure 5b). Where the parameter identifiability

15 is relatively low (< 0.8), i.e. for effective porosities of Pleistocene sand 2, Pleistocene clay 1 and Miocene clay (Maade), the estimated parameter value is more constrained by the regularization and hence stays close to the preferred value (Table 2). The low identifiability is a result of the distribution (or density) of observations compared to the particle travel paths. Figure 6 shows the pathlines of particle back-tracking (for better visualization only one path line is shown per well screen). As mentioned above, only ^{14}C observations with an activity higher than 5 pMC (Table 1) were used, which excludes results from well

20 screens C1, C2, C3, D1, D2, F1 and F2. The recharge area is mostly located to the east (Figure 6). The Maade formation is more dominant towards the west while it is patchy in the east. Consequently, it does not affect the particle tracking as much in the east. Only effective porosities of geological units through which particles actually travel are well informed by the observations. The low-permeability Maade unit acts as an obstacle to the travel paths and since the particles circumvent the Maade

14

formation the actual value of porosity has no impact on the age. The Maade unit significantly affects the age distribution due to its influence on travel paths, but no sensitivity to the porosity of the unit is found.

Pleistocene sand 2 represents less than 5% of the total amount of cells (Table 2) and it occurs mostly in the west. Pleistocene clay 1 is mostly shallow, patchy and located far away from the well locations. Hence, the impact of these two geological units on the particle tracks is also relatively small and results in low identifiability (Figure 5b).
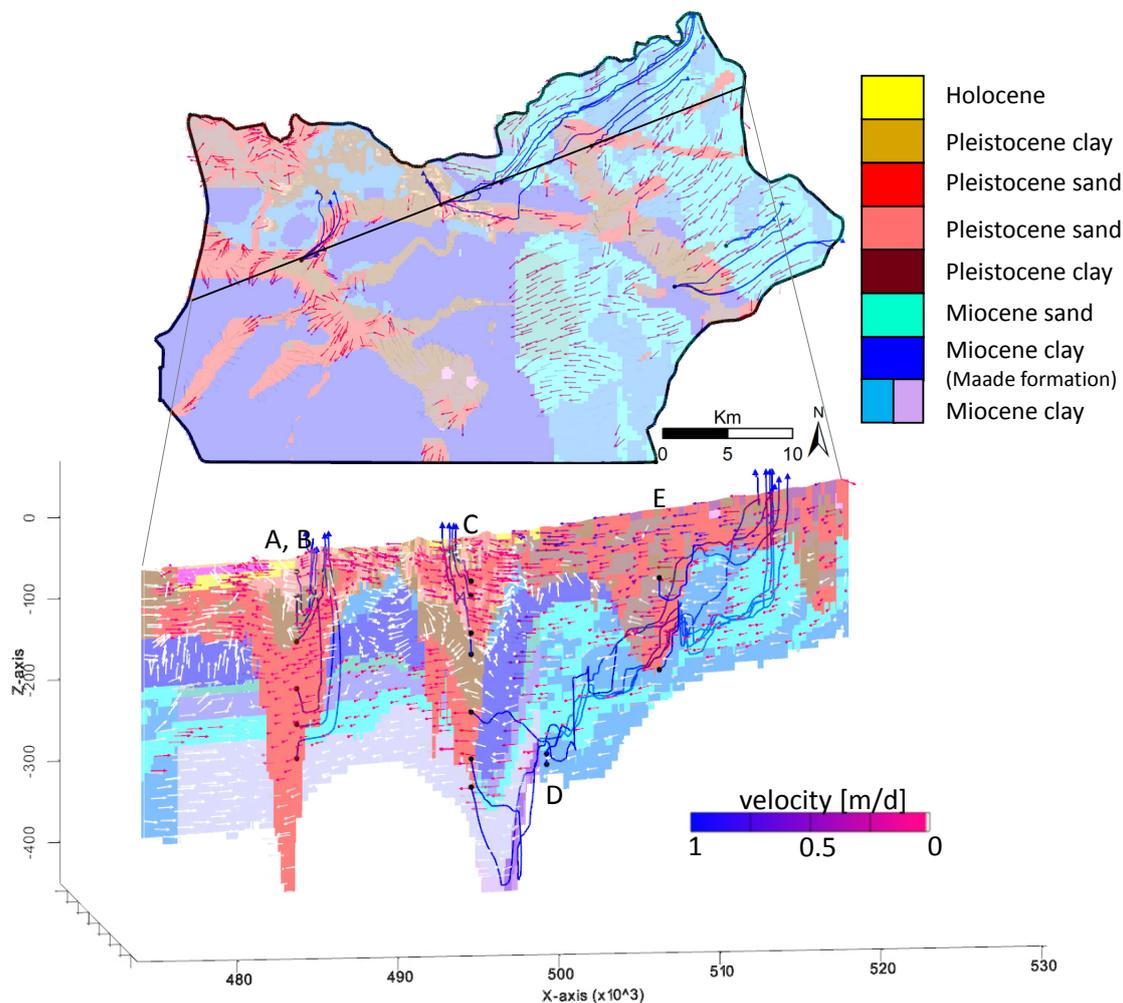
Figure 6. Horizontal geological cross-section at an elevation of -100 m a.s.l. and a SW-NE cross-section through sampling well locations (A-E). Bluish (cold) colors represent pre-Pleistocene sediments (dark blue = clay, light blue = sand), while warm colors represent Pleistocene deposits (red = sand, brown = clay). Also shown are MODPATH back-tracking lines (1 per cell) and groundwater flow velocity vectors.

5

4.3 Advective age distribution at observation wells

Figure 7 shows the simulated advective age distribution at the sampling locations (A-G, Figure 1).

15

Table 3. Results of the analysis of particle age distributions and path lengths. Gray shade indicates well screens used for calibration..

	A	B	C	D	E
well	mean particle age [years]	std particle age [years]	median particle age [years]	mean path length [km]	std path length [km]
A1	536	72	503	7.50	0.22
A2	392	16	387	6.17	0.36
B1	400	71	367	6.28	0.41
B2	272	31	277	3.69	0.61
C1	7232	2814	6503	26.68	1.64
C2	3654	2816	2818	27.52	0.94
C3	2640	608	2542	27.83	0.86
C4	1038	45	1036	3.24	0.07
C5	542	116	512	3.19	0.17
D1	14122	7563	13479	22.12	1.05
D2	5028	4498	3064	21.94	0.97
E1	1306	1508	908	13.56	0.82
E3	404	448	300	11.50	1.30
F1	6649	1405	6394	17.01	0.47
F2	2950	1584	2768	16.34	0.78
F3	1129	60	1120	14.83	0.57
G1	470	258	514	6.82	0.62
G2	135	17	130	6.02	0.46

The results show a wide variety of mean particle ages (Table 3) and the shape of the age distributions is very different (Figure 7). The well screens with mean particle ages less than 1000 years (except for C4 and F3 which have slightly higher mean particle ages) show particle age distributions that are mostly narrow and unimodal (except E3 and G1), which is also reflected in a small standard deviation (smaller than 20% of the mean age, except E3 and G1), see column B in Table 3. The particle age distributions of older waters with a mean particle age significantly larger than 1000 years (Table 3) tend to have broader and/or multi-modal shapes (Figures 7c,d) and large standard deviations (Table 3, column B).

The mean distance that particles travel from their recharge points to the sampling well (Table 3, column D) ranges between 3 km and 28 km. The younger waters (mean particle age <1000 years) show path lengths less than 10 km (except at well location E3 and F3), Figure 8, while most of the older waters travel more than 20 km. However, the relation between path length and travel time is far from linear. At some well locations (e.g. well locations A2, B2, C4, C5) the relation between path length and travel time forms a few distinct small clouds without much spread, indicating that the particles follow alternative large-scale preferential flow paths. At other locations a larger and more diffusive spread is found, either in travel times (e.g.
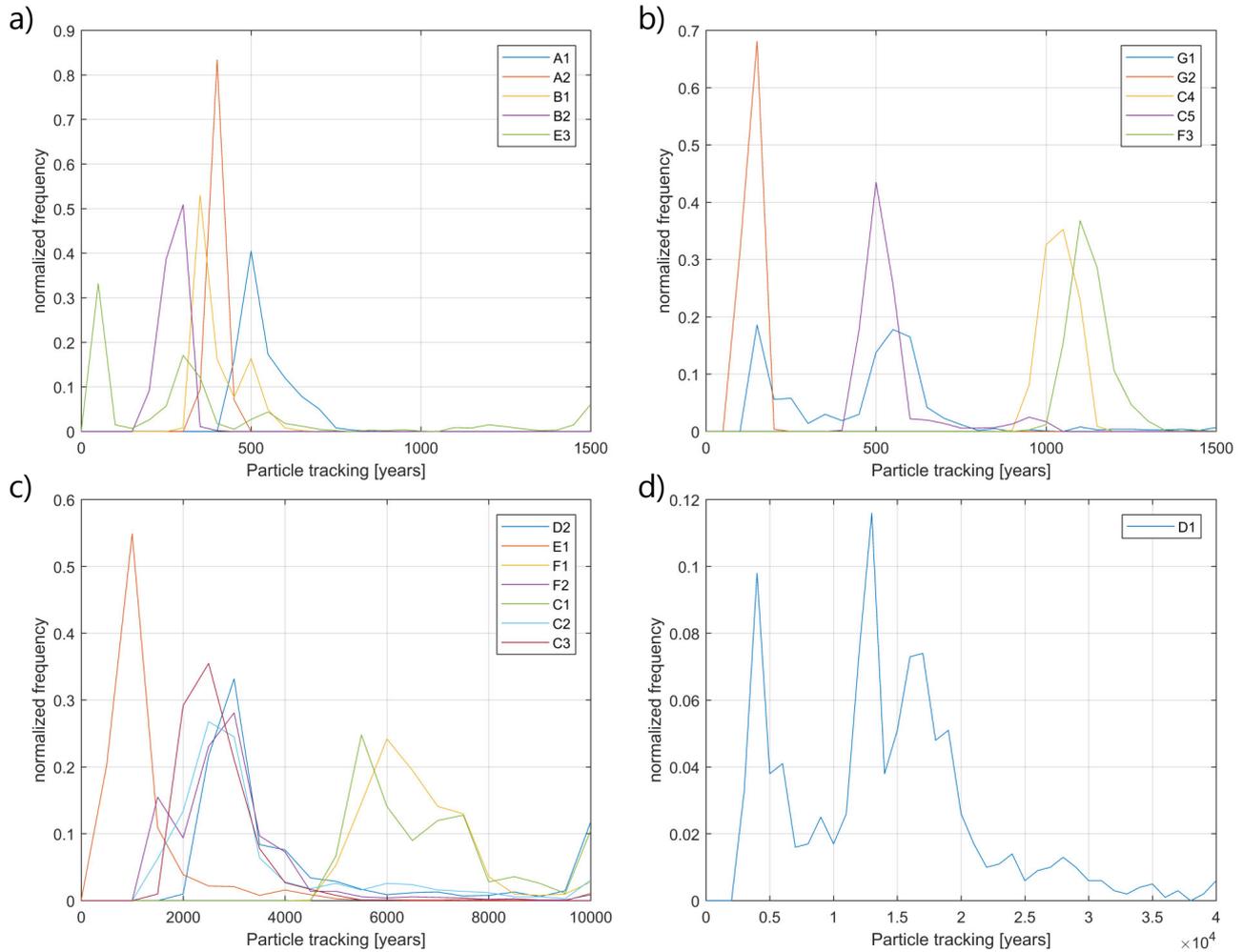
16

Figure 7. 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.

well locations C1, C2, C3, D1, E1) or path lengths (e.g. well locations F3, G2, E3). The large spread in travel times indicates that some particles travel slowly through clay units of various thicknesses. The large spread in path lengths originates from long and quick or short and slow travel paths through or around clay units and reflects the geological heterogeneity.

4.4 Regional age distribution based on direct age simulation

5 Figure 9 shows the ME and RMS of the direct mean age and the apparent age (corrected ^{14}C) for different α_L values (α_{TH} and α_{TV} are tied to α_L, see section 3.4). Minimum ME and RMS values are achieved for longitudinal dispersivities $\alpha_L <5$ m.
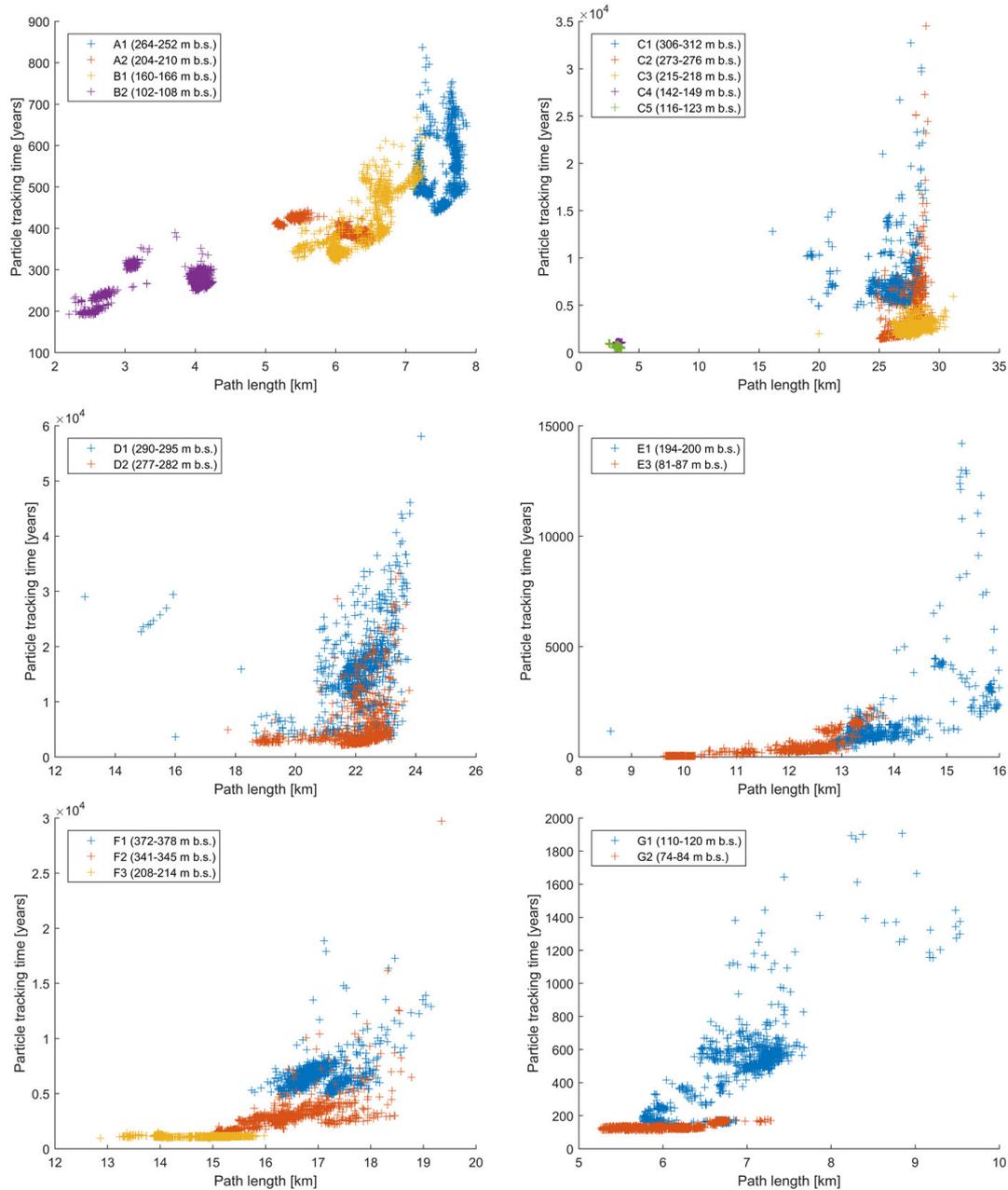
Figure 8. Particle tracking time over path length for the different well locations (cf. Figure 1.; the screen depth is indicated in parentheses).

For lower α_L the effect on ME and RMS is insignificant. ~~as numerical dispersion is expected to dominate at this scale. With higher α_L values ME and RMS increase significantly.~~Other regional-scale studies (e.g. Sonnenborg et al., 2016) have used longitudinal dispersivity in the magnitude of tens of meters or more to account for geological heterogeneities at formation

18
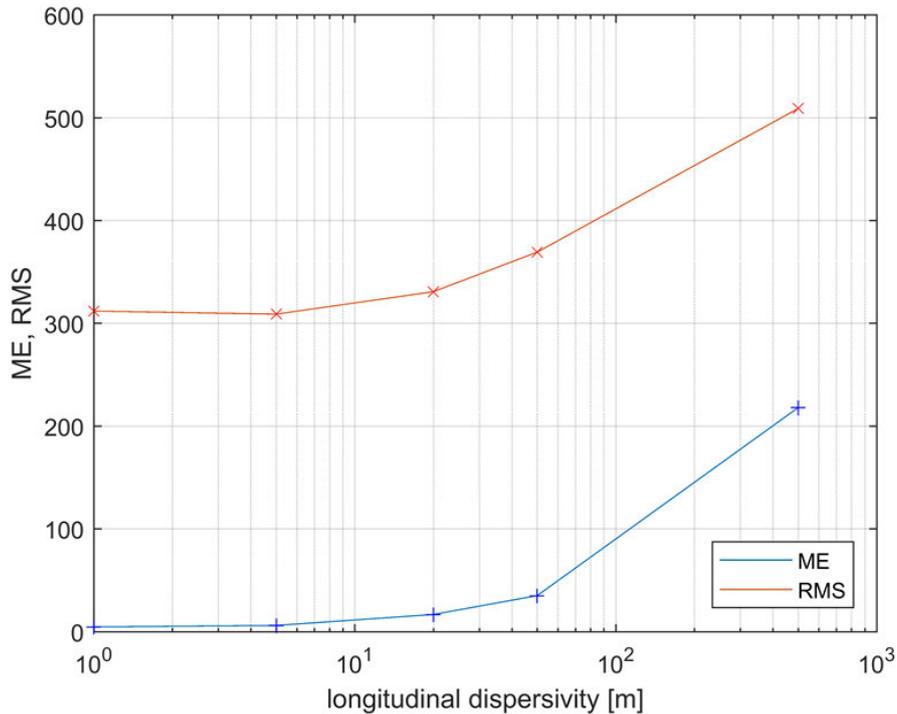
Figure 9. Mean error (ME) and root mean square (RMS) between corrected ^{14}C (shaded in gray in Table 1) and directly simulated ages as a function of longitudinal dispersivity.

scale. The very detailed voxel geological model ~~that~~ resolves heterogeneities at a scale of ~~200~~ 200 m x 200 m. Hence, it is assumed that mixing at scales larger than 200 m is accounted for by the geological model. Therefore, the dispersivity should only describe the heterogeneity at flow scale of several hundred of meters which justifies the use of a relatively small α_L. ~~Hence, the dispersivity only describes the effect of heterogeneity at the grid scale, .~~ In accordance with Gelhar et al. (1992) ~~this~~

5   ~~results~~ flow scales of hundreds of meters result in α_L ~~with a magnitude of~~ 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 2 m were applied in similar sized models (e.g. Oude Essink et al., 2010; Pauw et al., 2012). Similar to Weissmann et al. (2002) and LaBolle and Fogg (2001) the simulations showed little sensitivity to local scale dispersivity because at the modelling scale of tens of kilometers, dispersion is dominated by facies-scale heterogeneity which is captured by the detailed, highly resolved geological model. On the grid scale of 200 m

10 to 400 m and with the standard difference solver for the advection-dispersion equation a substantial numerical dispersion is expected. Choosing the TVD or MOC solver scheme for the advection-dispersion equation would have been more accurate in terms of less numerical dispersion, but would have required excessive running times which made it impractical to use in this study. Since there is no sensitivity for lower α_L (numerical dispersion dominates at this scale), ~~macrodispersivity of was used~~

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.~~

The directly simulated mean age distribution on a regional scale (Figure 10) shows a general age evolution from young water in the recharge area in the east towards older water in the west (Figure 10 b, e, f). Young water also enters the system through the coastal boundary in the west (Figure 10 b, e, f). The age distribution is strongly affected by the heterogeneity in flow and transport through the aquifers geology and is therefore in good agreement with the interpretation of the flow system by Meyer et al. (2018a). Two main aquifers are present on a regional scale: a shallow Pleistocene sand aquifer and a deep Miocene sand aquifer, separated by the Maade formation and locally connected through buried valleys (conceptual model in Figure 2, Figure 10 g, h). The regional mean age distribution also reflects this system. Younger waters dominate the shallow Pleistocene aquifers (Figure 10 a, e, f), where the flow regime can be described as mostly local and intermediate (cf. Tóth, 1963). The separating Maade formation with its increasing thickness towards the west (Figure 10 d) acts as a stagnant zone where groundwater age increases (Figure 10 c). The underlying Miocene sand shows the mean age evolution from young water in the recharge areas in the east to older water towards the discharge zones in the west (Figure 10 b, e, f). Here the flow regime is dominated by regional flow (cf. Tóth 1963). Special features are the buried valleys where downward flow of young waters, upwelling of old waters and mixing occurs (Figure 10 e, f, g, h). At the coastal boundary in the west young water enters the system and due to the density-corrected head boundary a wedge is formed with young waters in the wedge and old water accumulating in the transition zone (Figure 10 e, f). The two cross-sections e) and f) (Figure 10) differ in their geological connection to the sea-boundary (compare geological sections g) and h) in 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 waters 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. Another feature is the human land use change including an extensive drainage network with drain elevations below the sea level in the marsh area. There, old groundwater is forced upward, partly through buried valleys, before it could discharge into the sea.

4.4.1 Direct simulated mean age distribution in geological units

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 and climatic changes (glacial cover, sea level, ect.). Therefore, the steady-state assumption is a notable simplification, which is further discussed in section 5.1.
In Figure 11 the normalized direct age distributions are shown for a) the whole model, b) the Pleistocene aquifer, c) the Maade clay formation that acts as an aquitard, and d) the Miocene sand aquifer (compare the geological setting with conceptual model in Figure 2). The directly simulated mean groundwater ages for the whole model, the Pleistocene sand, the Maade formation and the Miocene sand were determined by a moment analysis (Levenspiel and Sater 1966) as 2574 years, 1009 years, 3883 years,
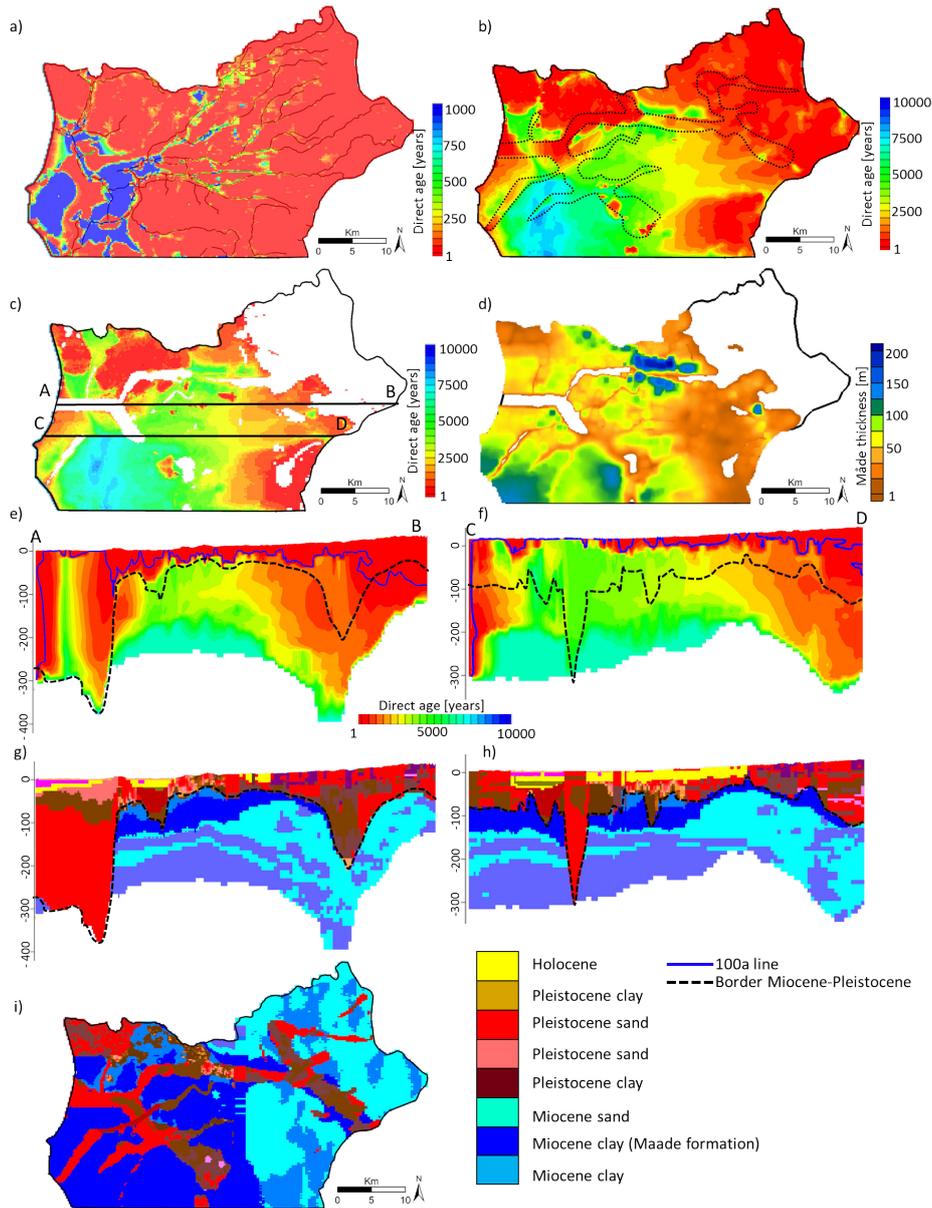
Figure 10. Directly simulated ~~mean~~ ages ~~and velocity vectors~~ presented at: a) horizontal section at layer 2, also showing river network; b) horizontal section at a depth of 100 m a.s.l. (buried valleys indicated with dotted lines); c) horizontal section at the top of the Maade formation; d) extent and thickness of the Maade formation; e) cross-sections A-B and f) C-D; Pleistocene-Miocene boundary indicated with dashed lines (buried valleys), 100 year lines; g) and h) geological cross-section and i) horizontal geological section, main geological units indicated (a detailed geological description is given in Meyer et al. 2018a). Notice that the color scheme in a) is different in order to better resolve younger ages close to the surface.
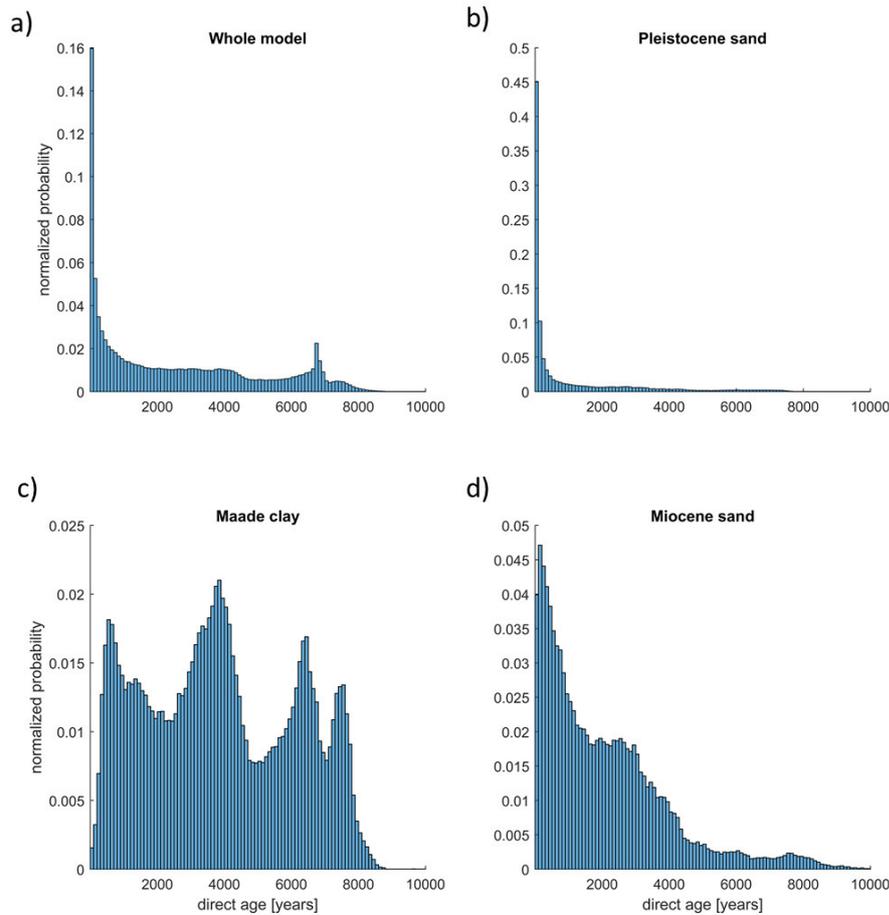
Figure 11. Frequency distributions (bin size = 100 years) of directly simulated groundwater ages in a) the whole model, b) the shallow Pleistocene aquifer, c) the separating Miocene clay (Maade formation) and d) the deep Miocene aquifer.

and 2087 years, respectively. The shape of the age distribution in these units varies significantly. The Pleistocene sand shows a unimodal distribution with one peak at ≈100 years and a tail (Figure 11b). The age distribution is governed by recharge of young water and discharge through rivers and drains, which are fed by the upwelling older groundwater (Figure 10a). The age distribution in the Maade formation is multi-modal with five peaks at about 600 years, 1400 years, 3900 years, 6500 years and 7600 years (Figure 11c). Comparison of Figures 10c and 10d reveals a positive relation between age and thickness of the Maade formation. The age distribution in the underlying Miocene sand has one peak at 200 years followed by a plateau between 1600 years and 3100 years and a small peak at 7800 years (Figure 11d). This distribution is controlled by the overlying and separating Maade formation in the west and the interlayering with Miocene clay.

5

4.4.2   Advective and directly simulated ages

The comparison of the advective ages with the direct simulated ages at the sampling well locations shows a good match for advective ages with a small variance and worsens when the variance increases (Figure 12). Older ages are generally associated with larger variances. Where the mismatch between advective and direct ages is large, the direct simulated mean ages are
5   consistently lower than mean ages derived from particle back tracking (see discussion below) because of diffusion into clay units. However, most of them lie within one standard deviation; but please observe that the standard deviation spans several thousands of years at some locations, where particle travel time distributions show a multi-modal shape.
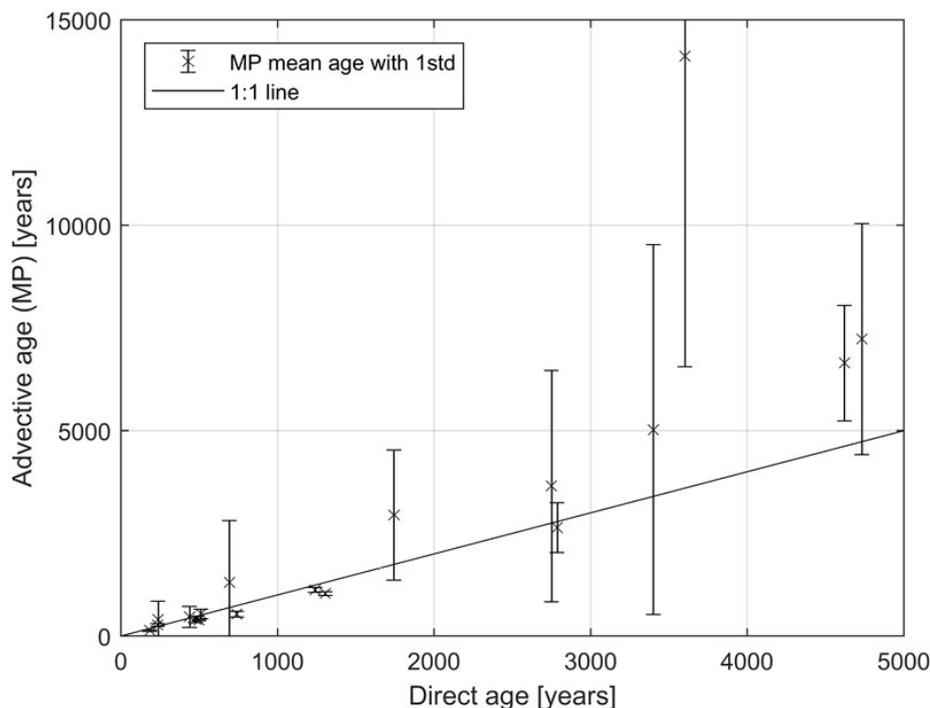
Figure 12. Mean advective age (MODPATH (MP) particle backtracking) compared to directly simulated mean groundwater age at sampling well locations; error bars on advective age represent 1 standard deviation.

4.4.3   Capture zones: effect of porosity

Figure 11 shows the capture zones at the Abild well for 1500 years and 2000 years and for the virtual well (AW) for 1000,
10   2000 1000 years, 2000 years and 3000 years for a constant effective porosity of 0.3 (solid line) and the calibrated distributed effective porosities model (dashed line), respectively. The capture zones of the two models vary both in extent and shape. The areas of the capture zone differ by up to 50%. Interestingly, it is not always the same effective porosity model that has the

smaller capture zone, but it changes due to the heterogeneity in the geological model and the assigned effective porosities. However, the results illustrate the importance of reliable estimates of effective porosity when delineating the capture zone of an abstractions well.
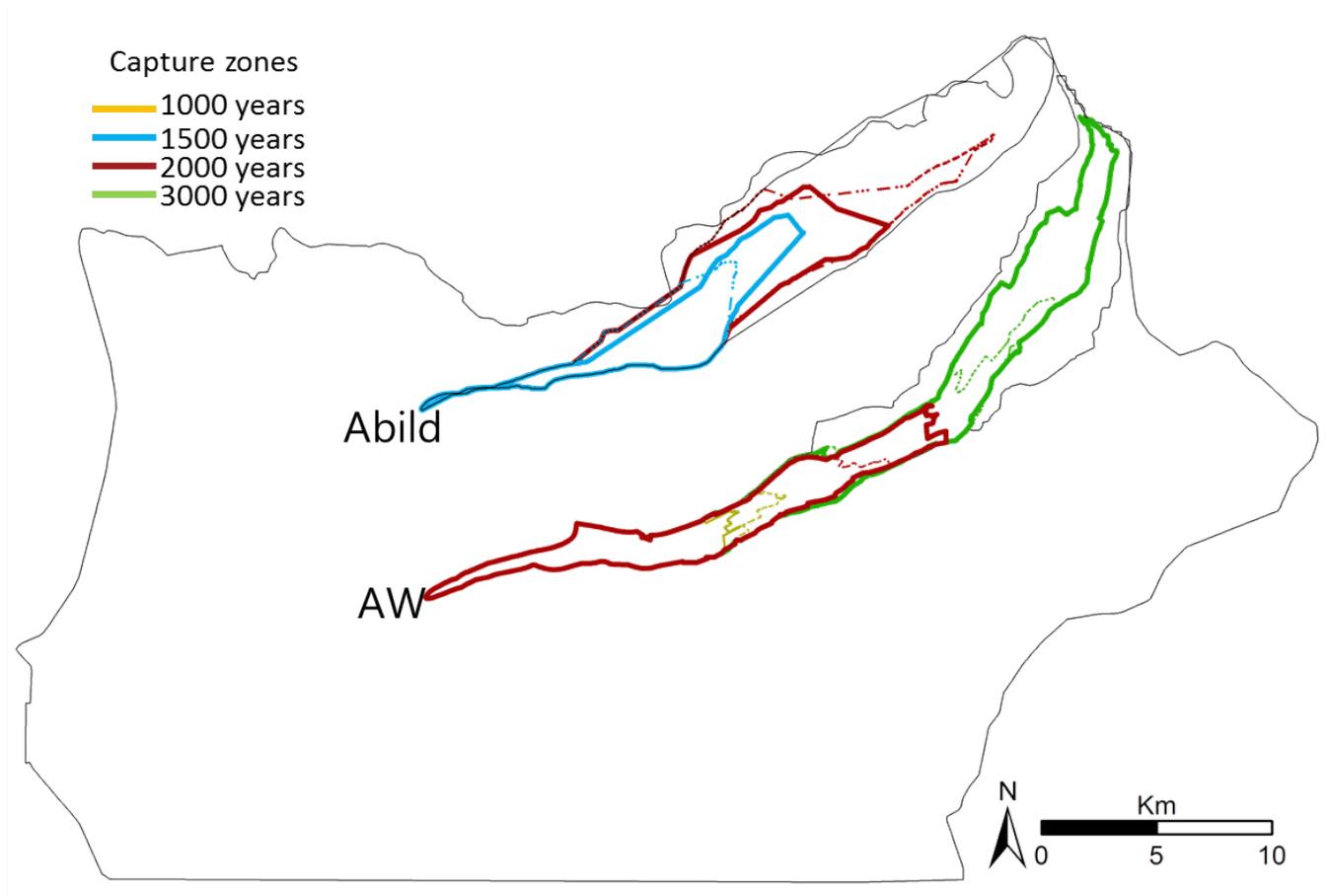

Figure 13. Capture zones at a well in Abild and a virtual well (AW) with a comparison of capture zones for a model with homogeneous porosity of 0.3 in all geological units (solid lines) and one with seven different porosities (dashed lines).

5 Discussion

^{14}C observations were used to constrain the estimation of effective porosities of a large-scale coastal aquifer system using an approach similar to Konikow et al. (2008), Weissmann et al. (2002) and Starn et al. (2014). Advective transport modelling and direct age simulations were applied to gain insight into the regional age structure of this highly heterogeneous geological system. In the following, limitations, uncertainties and simplifications of the model structure, estimated parameters and result-

ing interpretations are discussed. A detailed description of the age distribution is provided to highlight the relevant physical processes and their interactions.

5.1 Uncertainties

5.1.1 Boundary conditions

5 Uncertainties in model results originate partly from simplifications in boundary conditions and geological heterogeneities that are not resolved at the grid scale. Groundwater recharge, drain levels, well abstractions and sea levels were assumed constant over time for practical reasons and to reduce computational time. However, Karlsson et al. (2014) showed that recharge has changed significantly in Denmark during the last centuries. Changes in recharge could result in different age patterns (cf. Goderniaux et al. 2013). Similarly, sea level changes that were disregarded in this study would have an effect on the

10 groundwater age distribution in the coastal areas .(Delsman et al., 2014). Prescribing a vertical coastal age boundary of zero years is another simplification that neglects the vertical mixing and dispersion, which would result in an increase of age with depth (Post et al. 2013). However, since these physical processes were difficult to quantify, estimating age at this boundary would be highly uncertain. Thus, a constant age of zero years was applied.

The area close to the coast is not only affected by changing sea levels during the past thousands of years, but also by saltwater

15 intrusion. In this study, the density effects on flow were accounted for in a simplified way by using a density-corrected constant head boundary at the coast. Both, sea level changes and density effects, would also have affected the age distribution. The impact on age calculations due to density effects would be largest close to the coast. However, most of the groundwater samples used for age estimations were collected several tens of kilometers inland and are therefore expected to be affected to a minor extent. To quantify the impact of boundary conditions and saltwater intrusion on the particle tracking, the differences of particle

20 travel path lengths for a 200 year period, investigated based on the present model and a preliminary density-driven model (SEAWAT) accounting for non-stationary and density effects (similar to the one presented in Meyer, 2018c) are computed. The relative differences are below 10% (except at location A and B). Also, the uncertainties introduced by simplifying the density boundary effects are likely less important compared to other uncertainties associated, e.g., with estimating the groundwater age by the procedures for correcting ^{14}C activities. A solution would, of course, be to use a fully density-driven model such

25 as SEAWAT as in Meyer (2018c) or Delsman (2014). But, the very long computer run times for these kinds of models and the need of several thousands of model runs during calibration made it infeasible to use a variable-density flow model.

5.1.2 Apparent age as calibration target

Uncertainties in the use of ^{14}C as a groundwater dating tool and as calibration target arise at different levels. First, sampling of well screens with a length of 6 -10 m 6 m -10 m would encompass a range of groundwater ages as a result of mixing of

30 groundwater of different ages. Hereby younger waters, corresponding to DIC with a higher ^{14}C content, would dominate older ages (Park et al. 2002). The ^{14}C content is measured in the DIC of the groundwater. In order to obtain a reliable age estimate, the origin of DIC in groundwater is important. For the different processes that can affect the DIC and change its ^{14}C content

(e.g dissolution, precipitation, isotopic exchange) a variety of correction models exists (see overview of correction models in IAEA 2013). For the investigated system, corrections for carbonate dissolution and diffusion were applied, but it cannot be ruled out that also other chemical processes might have changed the ^{14}C content over the past thousands of years. The ^{14}C correction for diffusion into stagnant zones is sensitive to aquifer porosity, aquitard thickness and diffusion constant. The

5　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. Hence, average values of diffusion corrections were applied with parameters varying in ranges realistic for an aquifer system at this scale. However, in reality a groundwater particle would have been exposed to a variety of aquifer/aquitard thicknesses and porosities along its flow path implying smaller or larger diffusion. The correction results show that both carbonate dissolution and diffusion into stagnant

10　zones reduce the apparent groundwater age considerably, both at a similar magnitude as observed by Scharling (2011) and Hinsby et al. (2001).

As mentioned in the introduction, the apparent age (or radiometric age) is not equal to the mean particle-based kinematic age. This introduces additional, but unknown uncertain. Ideally, one could develop an advection-dispersion equation for the second moment and solve for the variance of ages (Varni and Carrera, 1998) and use that together with the directly simulated mean

15　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 ^{14}C (i.e. uncertainty on analyses, and corrections for effects of geochemical and physical processes).

Finally, the calibration of effective porosity using an advective transport model relies on a calibrated 3D flow solution that already bears uncertainties with respect to structure and parameters, as addressed by Meyer et al. (2018a). The number and

20　position of the released particles contribute to the uncertainty especially in heterogeneous systems as pointed out by Konikow et al. (2008) and Varni and Carrera (1998). The use of a high number of particles – here 1000 particles were distributed in one cell – generally reduces the uncertainty and enhances stability of the solution. The arithmetic mean of the 1000 released particles evenly distributed in the sampling cells resulted in estimates of effective porosities in the range of 0.13 to 0.45 for sand and 0.043 to 0.1 for clay units, which is significantly different to porosities of 0.25 or 0.30 that are often used in porous media

25　(e.g. Sonnenborg et al., 2016). The reliability of the estimated effective porosities was assessed through the identifiability that depends on the observation density (see section 4.2) and is high for four out of the seven estimated porosities.

5.1.3　~~Mean age~~Commensurability

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

30　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 ~~the direct age corresponds to the flux-averaged mean, the particle tracking age is resident-averaged (Varni and Carrera, 1998). Hence, the age distribution of the 1000 simulated particles, especially when it~~

~~is broad and multi-modal, shifts the mean age towards older ages. By using the harmonic mean of travel times of particles back-tracked from one cell (Konikow et al., 2008) more weight is given to younger ages which would more closely correspond to a flux-weighted mean. This approach improves the comparison (Figure 12; red stars), especially at wells, where the variances are large. Nonetheless, this approach is empirical and do generally not guarantee a better result. Hence, there are still some~~

5 ~~mismatches that~~ 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.

10 ## 5.2 Flow system and age distribution interpretation

5.2.1 Advective age distribution

The analysis of the advective age and travel distance distributions (Figures 7 and 8, Table 3) revealed a larger variance of ages for waters with a higher mean age. Following the pathlines of wells with younger waters (e.g. Figure 6, well locations A, B, C4, C5 and G), recharge areas are more proximal (path length <10 km, Figure 8, Table 3). Consequently, the particles pass

15 through fewer hydrogeological units and hence the flow path is less influenced by heterogeneous geology, which results in a smaller variance in ages and path lengths (Figure 8, Table 3). Particles travelling to well locations C1, C2, C3, D, E and F (e.g. Figure 6) have to travel through a variety of hydrogeological units, characterized by different hydraulic conductivities and effective porosities, hence showing a broader age distribution and larger variance as well as longer travel distances. Their broad and multi-modal age distributions reflect the up-gradient heterogeneity in fluxes, related to hydraulic conductivity and

20 effective porosity. This behaviour is in accordance with conclusions by Weissmann et al. (2002) who investigated groundwater ages in a heterogeneous 3D alluvial aquifer based on particle tracking and CFC-derived ages.

5.2.2 Regional age pattern

The regional age pattern derived from direct age simulation is consistent with the findings of Meyer et al. (2018a) about the flow system. The two-aquifer system is separated by a confining aquitard in the west. The shallow aquifer system consisting

25 of glaciotectonically disturbed Pleistocene sands mixed with clays is dominated by local and intermediate flow regimes and contains water of younger ages. The confining aquitard (Maade formation) shows older waters and a positive relation between ages and aquitard thickness what agrees with Bethke and Johnson (2008). In the deep Miocene sand aquifer that is interbedded with Miocene clay, regional flow regimes dominate and groundwater ages vary from young waters in the recharge areas in the east, where the overlying confining aquitard does not exist, to very old waters (up to 10000 years) in the west. The confining

30 Miocene aquitard (Maade formation) influences the age distribution pattern in the underlying Miocene sand in two ways. First, it limits deeper groundwater to seep upward and mix with the younger waters in the shallow aquifer. Secondly, the age flux from the aquitard to the aquifer shows a positive correlation with the ratio between aquitard thickness and aquifer thickness

27

(Bethke and Johnson, 2008).

At the buried valleys, groundwater exchange and hence age mixing occurs. Upwelling of the older groundwater from the deeper aquifer happens preferentially through these buried valleys. The dense drainage network in the west close to the coast

5   acts as a regional sink, with younger groundwater flowing horizontally and older water vertically and discharging to the drains. At the coastal boundary in the west, where a constant concentration of an "age mass" zero was assigned to the density-corrected constant head boundary, an age wedge characterized by waters of contrasting ages is established as a result of intruding young ocean water that meets old waters in the transition zone. This agrees with the findings by Post et al. (2013) based on simulation of synthetic groundwater age patterns in coastal aquifers using density-driven flow.

10   The results of our study differ significantly from findings by Sonnenborg et al. (2016) who investigated a regional aquifer system with a similar geological setting located a few hundred kilometers north of the present study area. Their direct simulation of groundwater ages shows a pattern of much younger water than here, rarely exceeding 700 years even in the deepest aquifers, while in our study ages exceeding 10000 years occur. The discrepancies may arise from differences in the geological models. In the area of Sonnenborg et al. (2016) the thickness of the Miocene sand units decreases towards west and disappears

15   before reaching the west coast. Sonnenborg et al. (2016) conclude that rivers control the age distribution even in deep aquifers. Based on particle tracking they found that the flow regimes were dominated by local and intermediate flow (cf. Tóth, 1963) with flow lengths not exceeding 15 km. In contrast, in the study presented here, the Miocene sand extends to the coast and probably beyond. While the age pattern in the shallow aquifers is controlled by rivers and drains (similarly to Sonnenborg et al., 2016), the age pattern in the deep aquifers is dominated by the extend and thickness of the Maade formation, the Marsh

20   area as a location of preferred discharge and the occurrence of buried valleys as locations of groundwater exchange, especially upwelling of old groundwater. Particle path lengths reach up to 30 km and regional flow dominates in the Miocene aquifer.

5.3   Perspectives of using spatial and temporal groundwater age distributions in groundwater quantity and quality assessment and management

The groundwater age distribution in aquifers is closely related to the distribution of physical (e.g. hydraulic conductivity and

25   porosity) and chemical parameters (e.g. concentrations of contaminants and natural geogenic elements) of the aquifers and aquitards. Hence, tracer and model estimated groundwater age distributions provide important information for the assessment of the hydraulic properties of the subsurface as demonstrated in this study, and as an indicator of groundwater quality and vulnerability (Hinsby et al., 2001a; Sonnenborg et al., 2016) including contaminant migration (Hinsby et al., 2001a), contents of harmful geogenic elements such as Arsenic and Molybdenum (Edmunds and Smedley, 2000; Smedley and Kinniburgh,

30   2002, 2017) and the risk of saltwater intrusion (MacDonald et al., 2016; Larsen et al., 2017; Meyer et al., 2018b). Groundwater age distributions in time and space are therefore important information for groundwater status assessment and the development of proper water management strategies that consider and protect both water resources quality and quantity (MacDonald et al., 2016). Water quality issues are often related to human activities such as contamination or overabstraction (MacDonald et al., 2016) and are typically found in waters younger than 100 years to depth of about 100 m (Seiler and Lindner, 1995;

Hinsby et al., 2001a) although deep subsurface activities may threaten deeper and older resources (Harkness et al., 2017). Deeper and older water is generally not contaminated or affected by human activities, but the impact of natural processes and contents of dissolved trace elements increases with depth and transport times (Edmunds and Smedley, 2000). Similarly, the risk of salt water intrusion from fossil seawater in old marine sediments increase with depth in inland aquifers and reduce the amount of available high quality groundwater resources (MacDonald et al. 2016; Larsen et al. 2017; Meyer et al. 2018b). Furthermore, old groundwater resources which are only slowly replenished are more vulnerable to over-exploitation, which lead to declining water tables, increasing hydraulic gradients and long-term non-steady state conditions that change the regional flow pattern (Seiler and Lindner, 1995) and potentially result in contamination of deeper groundwater resources by shallow groundwater leaking downward. The presented modelling results show that the Miocene sand aquifer is protected by the overlying Maade formation over a wide area. The aquifer bears old waters ($>$100 a, cf. Figures 10e,f) of high quality (Hinsby and Rasmussen, 2008), especially in the east and the central part of the area, as the risk of seawater intrusion increase towards the west. However, caution should be shown as the shallow and the deep aquifers are naturally connected through buried valleys, where groundwater exchange occurs in both direction (Meyer et al., 2018a). In these geological features, young and possibly contaminated water can be found to greater depth (Seifert et al., 2008, Figure 10e). Moreover, deep, old waters are vulnerable to contamination by modern pollutants as a result of the construction of wells with long screens, connecting different aquifers separated by aquitards (Seiler and Lindner, 1995; Jasechko et al., 2017, Figure 2).

6 Conclusions

The originality of this study comes from a 3D multi-layer coastal regional advective transport model, where heterogeneities are resolved on a grid scale. The distributed effective porosity field was found by parameter estimation based on apparent ages determined from ^{14}C activities, corrected for dissolution and diffusion. Based on regularized inversion seven effective porosities were estimated. Four of these were found to have high identifiability indicating that they are well constrained by the age data. The remaining three have moderate to low identifiability implying that they are less or poorly constrained by the data. In the latter case, parameter estimates close to the preferred values were obtained because of the use of Tikhonov regularization. By using a distributed effective porosity field, it was possible to match the observed age data significantly better than if effective porosity was assumed to be homogeneous and represented by a single value.

The advective age distributions at the well locations show a wide range of ages from few hundreds to several thousand years. Younger waters show narrower unimodal age distribution with small variances while older waters have wide age distributions, often multi-modal with large variances. The variances in age distribution reflect the spatial heterogeneity encountered by the groundwater when travelling from the recharge location to the sampling point.

The estimated effective porosity field was subsequently applied in a direct age simulation that provided insight into the 3D groundwater age pattern in a regional multi-layered aquifer system and the probable advance of modern potentially contaminated groundwater. Large areas in the shallow Pleistocene aquifer is dominated by young recharging groundwater ($< 200\,a$)

while older water is upwelling into rivers and drains in the marsh area. Hence, the upper aquifer is prone to contamination. In large areas the deeper Miocene aquifer is separated and protected by the Maade formation bearing old water, whereas young and possibly contaminated water is located in the recharge area in the East and in the buried valleys where the shallow and deep aquifer systems are shortcut.

5

The study clearly demonstrate the governing effect of the highly complex geological architecture of the aquifer system on the age pattern. Even though there are multiple uncertainties and assumptions related to groundwater age and its use in calibration, the results demonstrate that it is possible to estimate transport parameters that contain valuable information for assessment of groundwater quantity and quality issues. This can be used in groundwater management problems in general, as demonstrated

10 in an example of capture zone delineation where a heterogeneous distributed effective porosity field resulted in a 50% change in the capture zone area compared to the case of homogeneous effective porosity. The adopted approach is easy to implement even in large-scale models where auto-calibration of transport parameters using models based on the advection-dispersion equation might be restricted by computer run time.

References

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

Appelo, C.A.J., Postma, D., 2005. Geochemistry, Groundwater and Pollution. Balkema Publishers, Amsterdam.

Bethke, C.M., Johnson, T.M., 2008. Groundwater Age and Groundwater Age Dating. Annu. Rev. Earth Planet. Sci. 36, 121-152. https://doi.org/10.1146/annurev.earth.36.031207. 124210

Bethke, C.M., Johnson, T.M., 2002. Paradox of groundwater age. Geology 30, 107-110. https://doi.org/10.1130/0091-7613(2002)030<0107:POGA>2.0.CO

Boaretto, E., Thorling, L., Sveinbjörnsdóttir, Á.E., Yechieli, Y., Heinemeier, J., 1998. Study of the effect of fossil organic carbon on 14C in groundwater from Hvinningdal, Denmark, in: Mook, W.G., Plicht, J. van der (Eds.), Proceedings of the 16th International 14C Conference. pp. 915-920.

Bohlke, J.K., Denver, J.M., 1995. Combined use of ground- water dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, atlantic coastal Plain, Maryland. Water Resour. Res. 31, 2319-2339. https://doi.org/10.1029/95WR01584

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

Castro, M.C., Goblet, P., 2005. Calculation of ground water ages-a comparative analysis. Groundwater 43, 368-380. https://doi.org/10.1111/j.1745-6584.2005.0046.x

Castro, M.C., Goblet, P., 2003. Calibration of regional groundwater flow models:Working toward a better understanding of site-specific systems. Water Resour. Res. 39, 1172. https://doi.org/10.1029/2002WR001653

Cook, P.G., Herczeg, A.L., 2000. Environmental tracers in subsurface hydrology. Springer Science and Business. https://doi.org/10.1017/CBO9781107415324.004

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

Delsman, J.R., Hu-a-ng, K.R.M., Vos, P.C., de Louw, P.G.B., Oude Essink, G.H.P., Stuyfzand, P.J., Bierkens, M.F.P., 2014. Paleo-modeling of coastal saltwater intrusion during the Holocene: an application to the Netherlands. Hydrol. Earth Syst. Sci. 18, 3891–3905. https://doi.org/10.5194/hess-18-3891-2014

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

Doherty, J., 2016. Model-Independent Parameter Estimation I. Watermark Numer. Comput. 366.

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

Eberts, S.M., Böhlke, J.K., Kauffman, L.J., Jurgens, B.C., 2012. Comparison of particle-tracking and lumped-parameter age-distribution models for evaluating vulnerability of production wells to contamination. Hydrogeol. J. 20, 263-282. https://doi.org/10.1007/ s10040-011-0810-6

Edmunds, W.M., Smedley, P.L., 2000. Residence time indicators in groundwater: The East Midlands Triassic sandstone aquifer. Appl. Geochemistry 15, 737-752. https://doi.org/ 10.1016/S0883-2927(99)00079-7

Engesgaard, P., Molson, J., 1998. Direct simulation of ground water age in the Rabis Creek Aquifer, Denmark. Ground Water. https://doi.org/10.1111/j.1745-6584.1998.tb02831.x

5   Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall.

Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A Critical Review of Data on Field-Scale Dispersin 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

10   Gleeson, T., Befus, K.M., Jasechko, S., Luijendijk, E., Cardenas, M.B., 2015. The global volume and distribution of modern groundwater. Nat. Geosci. 9, 161-167. https://doi.org/10.1038/ ngeo2590

Goderniaux, P., Davy, P., Bresciani, E., De Dreuzy, J.R., Le Borgne, T., 2013. Partitioning a regional groundwater flow system into shallow local and deep regional flow compartments. Water Resour. Res. 49, 2274-2286. https://doi.org/10.1002/wrcr.20186

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

15   Guo, W., Langevin, C.D., 2002. User's Guide to SEAWAT: A Computer Program For Simulation of Three-Dimensional Variable-Density Ground-Water Flow. Tallahassee.

Hansen, B., Dalgaard, T., Thorling, L., Sørensen, B., Erlandsen, M., 2012. Regional analysis of groundwater nitrate concentrations and trends in Denmark in regard to agricultural influence. Biogeosciences 9, 3277-3286. https://doi.org/10.5194/bg-9-3277-2012

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW-2000 , The U. S. Geological Survey modular ground-water
20   model-user guide to modularization concepts and the ground-water flow process. USGS Open-File Rep. 00-92.

Harkness, J.S., Darrah, T.H., Warner, N.R., Whyte, C.J., Moore, M.T., Millot, R., Kloppmann, W., Jackson, R.B., Vengosh, A., 2017. The geochemistry of naturally occurring methane and saline groundwater in an area of unconventional shale gas development. Geochim. Cosmochim. Acta 208, 302-334. https://doi.org/10.1016/j.gca.2017.03.039

Harris, K.R., Woolf, L.A., 1980. Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water. J.
25   Chem. Soc. Faraday Trans. 1 76, 377. https://doi.org/10.1039/f19807600377

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.

Henriksen, H.J., Troldborg, L., Nyegaard, P., Sonnenborg, T.O., Refsgaard, J.C., Madsen, B., 2003. Methodology for construction, calibration and validation of a national hydrological model for Denmark. J. Hydrol. 280, 52-71. https://doi.org/10.1016/S0022-1694(03)00186-0

30   Hill, M.C., Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: with analysis of data, sensitivities, predictions and uncertainties. Wiley.

Hinsby, K., Edmunds, W.M., Loosli, H.H., Manzano, M., Condesso De Melo, M.T., Barbecot, F., 2001. The modern water interface: recognition, protection and development — advance of modern waters in European aquifer systems. Geol. Soc. London, Spec. Publ. 189, 271-288. https://doi.org/10.1144/gsl.sp.2001.189.01.16

35   Hinsby, K., Harrar, W.G., Nyegaard, P., Konradi, P.B., Rasmussen, E.S., Bidstrup, T., Gregersen, U., Boaretto, E., 2001. The Ribe Formation in western Denmark – Holocene and Pleistocene groundwaters in a coastal Miocene sand aquifer. Geol. Soc. London, Spec. Publ. 189, 29-48. https://doi.org/10.1144/GSL.SP.2001.189.01.04

Hinsby, K., Purtschert, R., Edmunds, W.M., 2007. Groundwater age and quality, in: Quevauviller, P. (Ed.), Groundwater Science and Policy. Royal Society of Chemistry, London.

Hinsby, K., Rasmussen, E.S., 2008. The Miocene sand aquifers, Jutland, Denmark, in: Edmunds, W.M., Shand, P. (Eds.), Natural Groundwater Quality. Blackwell.

5   Hölting, B., Coldewey, W.G., 2013. Hydrogeologie, 8th ed. Springer-Verlag, Berin, Heidelberg. https://doi.org/10.1007/978-3-8274-2354-2

Høyer, A.-S., Vignoli, G., Hansen, T.M., Vu, L.T., Keefer, D.A., Jørgensen, F., 2016. 2017. Multiple-point statistical simulation for hydrogeological models: 3D training image development and conditioning strategies. Hydrol. Earth Syst. Sci. Discuss. 21, 6069–6089. https://doi.org/10.5194/hess-2016-567

IAEA, 2013. Isotope Methods for Dating Old Groundwater. Internation Atomic Agency, Vienna.

10   Jaehne, B., Heinz, G., Dietrich, W., 1987. Measurement of the diffusion coefficients of sparingly soluble gases in water. J. Geophys. Res. 92, 10767-10776.

Jasechko, S., Perrone, D., Befus, K.M., Cardenas, M.B., Ferguson, G., Gleeson, T., Luijendijk, E., McDonnell, J.J., Taylor, R.G., Wada, Y., Kirchner, J.W., 2017. Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. Nat. Geosci. 10, 425-429. https://doi.org/10.1038/ngeo2943

15   Jørgensen, F., Høyer, A.-S., Sandersen, P.B.E., He, X., Foged, N., 2015. Combining 3D geological modelling techniques to address variations in geology, data type and density - An example from Southern Denmark. Comput. Geosci. 81, 53-63. https://doi.org/10.1016/j.ca geo.2015. 04.010

Karlsson, I.B., Sonnenborg, T.O., Jensen, K.H., Refsgaard, J.C., 2014. Historical trends in precipitation and stream discharge at the Skjern River catchment, Denmark. Hydrol. Earth Syst. Sci. 18, 595-610. https://doi.org/10.5194/hess-18-595-2014

20   Kazemi, G.A., Lehr, J.H., Perrochet, P., 2006. Groundwater Age. Wiley - Interscience.

Konikow, L.F., Hornberger, G.Z., Putnam, L.D., Shapiro, A.M., Zinn, B.A., 2008. The use of groundwater age as a calibration target, in: Proceedings of ModelCare: Calibration and Reliability in Groundwater Modelling: Credibility of Modelling. IAHS, pp. 250-256.

LaBolle, E.M., Fogg, G.E., 2001. Role of Molecular Diffusion in Contaminant Migration and Recovery in an Alluvial Aquifer System. Transp. Porous Media 42, 155–179. https://doi.org/10.1023/A:1006772716244

25   Larsen, F., Tran, L.V., Van Hoang, H., Tran, L.T., Christiansen, A.V., Pham, N.Q., 2017. Groundwater salinity influenced by Holocene seawater trapped in incised valleys in the Red River delta plain. Nat. Geosci. 10, 376-381. https://doi.org/10.1038/ngeo2938

Levenspiel, O., Sater, V.E., 1966. Two-phase flow in packed beds. I EC Fundam. 5, 86-92.

MacDonald, A.M., Bonsor, H.C., Ahmed, K.M., Burgess, W.G., Basharat, M., Calow, R.C., Dixit, A., Foster, S.S.D., Gopal, K., Lapworth, D.J., Lark, R.M., Moench, M., Mukherjee, A., Rao, M.S., Shamsudduha, M., Smith, L., Taylor, R.G., Tucker, J., van Steenbergen, F.,

30   Yadav, S.K., 2016. Groundwater quality and depletion in the Indo-Gangetic Basin mapped from in situ observations. Nat. Geosci. 9, 762-766. https://doi.org/10.1038/ngeo2791

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.

Manning, A.H., Solomon, D.K., Thiros, S.A., 2005. 3 H3H/3 He 3He Age Data in Assessing the Susceptibility of Wells to Contamination.

35   Ground Water 43, 353-367. https://doi.org/10.1111/ j.1745-6584.2005.0028.x

McCallum, J.L., Cook, P.G., Simmons, C.T., 2015. Limitations of the Use of Environmental Tracers to Infer Groundwater Age. Groundwater 53, 56-70. https://doi.org/10.1111/gwat. 12237

McCallum, J.L., Cook, P.G., Simmons, C.T., Werner, A.D., 2014. Bias of Apparent Tracer Ages in Heterogeneous Environments. Groundwater 52, 239-250. https://doi.org/10.1111/ gwat.12052

McMahon, P.B., Carney, C.P., Poeter, E.P., Peterson, S.M., 2010. Use of geochemical, isotopic, and age tracer data to develop models of groundwater flow for the purpose of water management, northern High Plains aquifer, USA. Appl. Geochemistry 25, 910-922. https://doi.org/10.1016/j.apgeochem.2010.04.001

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

Meyer, R., Engesgaard, P., Sonnenborg, T.O., 2018b. Origin and dynamics of saltwater intrusions in regional aquifers; combining 3D saltwater modelling with geophysical and geochemical data. Water Resour. Res., submitted.

Meyer, R., 2018c. Large scale hydrogeological modelling of a low-lying complex coastal aquifer system. University of Copenhagen. 184p.

Molson, J.W., Frind, E.O., 2012. On the use of mean groundwater age, life expectancy and capture probability for defining aquifer vulnerability and time-of-travel zones for source water protection. J. Contam. Hydrol. 127, 76-87. https://doi.org/10.1016/j.jconhyd.2011.06.001

Morgan, L.K., Werner, A.D., Simmons, C.T., 2012. On the interpretation of coastal aquifer water level trends and water balances?: A precautionary note. J. Hydrol. 470-471, 280-288. https://doi.org/10.1016/j.jhydrol.2012.09.001

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

Park, J., Bethke, C.M., Torgersen, T., Johnson, T.M., 2002. Transport modeling applied to the interpretation of groundwater 36 Cl age. Water Resour. Res. 38, 1-15. https://doi.org/10.10 29/2001WR000399

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

Pearson, F.J., Hanshaw, B.B., 1970. Sources of dissolved carbonate species in groundwater and their effects on Carbon-14 dating. Isot. Hydrol.

Pollock, D.W., 2012. User Guide for MODPATH Version 6 - A Particle-Tracking Model for MODFLOW. U.S. Geol. Surv. Tech. Methods 6-A41.

Post, V.E.A., Kooi, H., Simmons, C., 2007. Using hydraulic head measurements in variable-density ground water flow analyses. Ground Water 45, 664-71. https://doi.org/10.1111/j.1745-6584.2007.00339.x

Post, V.E.A., Vandenbohede, A., Werner, A.D., Maimun, Teubner, M.D., 2013. Groundwater ages in coastal aquifers. Adv. Water Resour. 57, 1-11. https://doi.org/10.1016/j.advwatres. 2013.03.011

Rasmussen, E.S., Dybkjær, K., Piasecki, S., 2010. Lithostratigraphy of the Upper Oligocene-Miocene succession of Denmark. Geol. Surv. Denmark Greenl. Bull. 22.

Salmon, S.U., Prommer, H., Park, J., Meredith, K.T., Turner, J. V., McCallum, J.L., 2015. A general reactive transport modeling framework for simulating and interpreting groundwater 14C age and delta 13C. Water Resour. Res. 51, 359–376. https://doi.org/10.1002/2014WR015779

34

Sanford, W., 2011. Calibration of models using groundwater age. Hydrogeol. J. 19, 13-16. https://doi.org/10.1007/s10040-010-0637-6

Sanford, W.E., 1997. Correcting for Diffusion in Carbon-14 Dating of Ground Water. Ground Water 35, 357-361.

Sanford, W.E., Plummer, L.N., Busenber, G.C., Nelms, D.L., Schlosser, P., 2017. Using dual-domain advective-transport simulation to reconcile multiple-tracer ages and estimate dual-porosity transport parameters. Water Resour. Res. 53, 5002-5016. https://doi.org/10.1002/2016WR019469.Received

Sanford, W.E., Plummer, L.N., McAda, D.P., Bexfield, L.M., Anderholm, S.K., 2004. Hydrochemical tracers in the middle Rio Grande Basin, USA: 2. Calibration of a groundwater-flow model. Hydrogeol. J. 12, 389-407. https://doi.org/10.1007/s10040-004-0326-4

Scharling, P.B., 2011. Hydrogeological modelling and multiple environmental tracer analysis of the deep Miocene aquifers within the Skjern and Varde river catchment. University of Copenhagen.

Seifert, D., Sonnenborg, T.O., Scharling, P., Hinsby, K., 2008. Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability. Hydrogeol. J. 16, 659-674. https://doi.org/10.1007/s10040-007-0252-3 Seiler, K., Lindner, W., 1995. Near-surface and deep groundwaters. J. Hydrol. 165, 33-44. https://doi.org/10.1016/0022-1694(95)92765-6

Smedley, P.L., Kinniburgh, D.G., 2017. Molybdenum in natural waters: A review of occurrence, distributions and controls. Appl. Geochemistry 84, 387-432. https://doi.org/10.10 16/j.apgeo chem.2017.05.008

Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behaviour and distribution of arsenic in natural waters. Appl. Geochemistry 17, 517-568. https://doi.org/10.10 16/S0883-2927(02)00018-5

Sonnenborg, T.O., Scharling, P.B., Hinsby, K., Rasmussen, E.S., Engesgaard, P., 2016. Aquifer Vulnerability Assessment Based on Sequence Stratigraphic and 39 Ar Transport Modeling. Groundwater 54, 214-230. https://doi.org/10.1111/gwat.12345

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

Stroeven, A.P., H́attestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B.W., Harbor, J.M., Jansen, J.D., Olsen, L., Caffee, M.W., Fink, D., Lundqvist, J., Rosqvist, G.C., Strömberg, B., Jansson, K.N., 2016. Deglaciation of Fennoscandia. Quat. Sci. Rev. 147, 91-121. https://doi.org/10.1016/j.quascirev.2015.09.016

Sudicky, E.A., Frind, E.O., 1981. Carbon 14 dating of groundwater in confined aquifers: Implications of aquitard diffusion. Water Resour. Res. 17, 1060-1064. https://doi.org/10.1029/ WR017i004p01060

Tikhonov, A.N., Arsenin, V.Y., 1977. Solutions of ill-posed problems. Winston, Washington, DC.

Tóth, J., 1963. A theoretical analysis of groundwater flow in small drainage basins. J. Geophys. Res. 68, 4795-4812. https://doi.org/10.1029/JZ068i016p04795

Troldborg, L., Jensen, K.H., Engesgaard, P., Refsgaard, J.C., Hinsby, K., 2008. Using Environmental Tracers in Modeling Flow in a Complex Shallow Aquifer System. J. Hydrol. Eng. 13, 199-204. https://doi.org/10.1061/(ASCE)1084-0699(2008)13

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

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

Weissmann, G.S., Zhang, Y., LaBolle, E.M., Fogg, G.E., 2002. Dispersion of groundwater age in an alluvial aquifer system. Water Resour. Res. 38, 16.1-16.13. https://doi.org/10.1029/ 2001WR000907

Wood, C., Cook, P.G., Harrington, G.A., Knapton, A., 2017. Constraining spatial variability in recharge and discharge in an arid environment through modeling carbon-14 with improved boundary conditions. Water Resour. Res. 53, 142-157. https://doi.org/10.1002/2015WR018424

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

5