Responses to Referee Review 1

We thank the referee reviewer for his comprehensive and insightful comments. Our responses to the reviewers' comments are given below. The original comments from referee reviewer 1 were marked with blue color, and our response in black.

Significance The research is significant, and fits in the scope of several recent papers on TTDs, SAS functions and using distributed models to calculate these. More knowledge on the effect of input and parameter uncertainty on TTDs is very welcome. It is interesting to see a study in which TTDs are calculated using several parameter variations in a distributed groundwater model.

General comments: However, the grammar and language of the paper is not up to publication standard. Because errors were many, I have not focussed on this in the current review. For example, Line 1 (page 2) needs 'the' before 'Travel/transit time' and needs 'a description' instead of 'an description'. This continues throughout the manuscript and needs significant work. The manuscript can be shortened and more to the point as it contains quite some repetition.

Overall, the paper lacks sufficient in-depth discussion and conclusions. Several observations are made, but no process-based explanations are given. In addition, several important recently published papers were overlooked.

A considerable amount of work is required both on language and content, but the current manuscript offers a good basis for this.

Response: Thank you very much for your overall assessment to our manuscript as well as for your insightful suggestions. We have revised the manuscript carefully following your suggestions. The discussion and conclusions were restructured and modified accordingly. Several recent publications were also included in the revised version.

Specific Comments: P1L1 refers to Page 1 Line 1.

P2L7: Suggested reference Wang et al., 2016 STOTEN

Response: We have added the reference accordingly. Please check Page 2, Line 7 in the revised manuscript.

P3L15: 'threaten', what is meant by this?

Response: We have changed the word 'threaten' to 'hamper'.

P3L16-17: 'The combination of expert knowledge and parameterization is generally recommended in hydrogeological modelling.' This sentence can be removed as it does not add much.

Response: Changed as proposed.

P3L24-35: This is a list of earlier research. But what does it add? What are the conclusions/implications for the current study?

Response: This is a list of past studies that are very relevant to our key focuses: the factor that controls the shape and scale of predicted travel time distributions. We have revised the text, shorted and restructured this paragraph. Please check Page 3, Line 19-29 in the revised manuscript.

P4: An important assumption in the paper is steady-state groundwater flow. However it is unclear if the mHM model is steady state as well. What exactly is modelled by which model? What is meant by 'terrestrial hydrological processes'? Was the mHM model only used to compute realistic values for recharge in the OGS model?

Response: In this study, mHM and OGS are one-way coupled, because our focus is the influence of recharges (values and their spatial pattern) and hydro-geological properties (Ks values) on the resulting TTDs.

The mHM is not run under the steady state condition, rather it is a dynamic model which is run on a daily time step for a time-span of 60 years (1951-2010). We used the long-term averaged recharge values based on the mHM runs and use them to force the OGS groundwater model which is run under the steady-state condition.

Terrestrial hydrological processes means that mHM-OGS only calculates land surface and subsurface hydrological processes (e.g. discharge, groundwater recharge, ET, soil moisture, groundwater flow and transport). However, the atmospheric hydrological cycle cannot be modeled using the mHM-OGS framework.

Yes, mHM is only used to compute realistic values for recharge in the OGS model in this study.

In the revised manuscript, we added the above details to the Methodology section.

P5L7 & Equation 2: What about horizontal groundwater flow? Was only vertical flow modelled?

Response: The groundwater module based on the OGS model account for the three-dimensional groundwater flow, and the flow path lines can be visualized in Figure 6.

P5L10-11: 'The functionality: : : by Part et al. (2008)' is not needed.

Response: We deleted this sentence as proposed.

Section 2.2: is it needed to explain the RWPT in such detail? Especially since it is explained in the referred papers. This distracts from the current study.

Response: Thank you for this suggestion. We have moved most of the content describing RWPT from section 2.2 to the appendix A.

P6L1: Suggest adding 'analytical' to the paragraph header.

Response: We have changed the title into 'Travel Time Distribution and Analytical StorAge Selection function'.

P6L6-7: Repetition of P4L32-33.

Response: We deleted the redundant sentences accordingly.

P6L8: 'output flux (Q1, Q2, etc)': in a steady state system Q would not vary.

Response: Agreed. In the following equations where the steady-state assumptions are introduced, all the applied fluxes are expressed in their time-invariant form.

P6L9: Define 'T' and 't'.

Response: T is the residence time of the oldest water parcel in storage (S_T), and t denotes the chorological time. They are two basic variables in the master equation (ME) of the TTDs. We followed the reviewer's comment and added the explanation of these two terms in the revised manuscript.

Eq10: First introduce SAS functions. Also refer to Figure 9a here.

Response: Changed as proposed (Page 12, Line 8). The Eq. 4 is updated accordingly and the texts in the following paragraph are also revised.

P6L21: Define 'RTD', not mentioned earlier.

Response: Changed as proposed (Page 12, Line 12). The full name "Residence Time Distribution (RTD)" were added into the revised manuscript at its first appearance.

P6L26: Maybe add some references to TTD literature where Exponential TTDs are used.

Response: Thank you for this suggestion. We added several references related to exponential TTDs usages in earlier studies (Page 12, Line 14).

P6-7: Unclear how Equations 13 and 14 follow from Eq9, 10, 13. These equations possibly need check (or references).

Response: We have double-checked these equations. We have also added some references to these equations.

Generally section 2.3 is quite hard to follow. Which equations are needed for the current study?

Response: Eq. 12 (Eq. 6 in the revised manuscript) is the equation for the exponential TTD calculation. Eq. 14 (Eq. 10 in the revised manuscript) is used for calculating the SAS function.

P7L11: 'As indicated in Bayes equation': which equation is this?

Response: In probability theory and statistics, Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. We use the term "Bayes' theorem" to replace the "Bayes equation" (Page 13, Line 8).

P7L23: Suggest shifting order of paragraphs. Start with Site Description. Then Numerical Model & Model setup. Then analytical model?

Response: Thank you for this helpful suggestion. We changed the order of paragraphs as suggested.

P7L27: '164 m': above Mean Sea Level?

Response: Yes, this means 164 m above mean sea level (a.m.s.l.). Please check Page 4, Line 15 in the revised manuscript.

P8L5: '5 in km, 4 in ku,: : :': these terms have not been introduced here yet. Move to later, after P8L16.

Response: Thank you for this suggestion. We have revised the texts accordingly (Page 5, Line 1-2).

P8L19: This recharge is the recharge calculated by the mHM model right? This remains unclear here. Same for P8L26, these are from the mHM model?

Response: Yes, the recharge is calculated by the mHM. Accordingly, we changed this sentence as "The gridded recharges estimated by mHM are interpolated and then assigned to each grid node on the upper surface of the OGS mesh using a bilinear interpolation approach." Please check Page 6, Line 27-28 and Page 7, Line 7-8.

P8L29-30: spatially distributed conductivity fields?

Response: Yes, we use the spatially distributed conductivity fields.

P9Figure1: Show the location in Germany of the study area. Indicate which layers are aquifers and which aquitards. In the legend, give the full names. At this moment the hydrogeology is not clear from this Figure.

Response: We have now revised the plot with you suggestions. It includes now the details on the geographical location (within Germany) and the full name of aquifers and aquitards. Please check the updated Figure 1.

P9L5: 'the model': which model?

Response: Here we refer to the OGS model.

P10Table1: Possibly use Hydraulic conductivity in m/d. Also, it would help to use more significant numbers, at this moment it's hard to compare the values as the differences are hidden in the small superscripts.

Response: Thank you for this observation. Since we have used the standard unit of hydraulic conductivity throughout the manuscript, it would be tricky to use another unit system only in Table 1. Therefore we kept the unit of Ks as m/s in this Table 1.

P10L2-3: Repetition of P8L29-30.

Response: We have deleted one of those texts accordingly.

P10L5-6: All parameters sets gave a good fit? What was the definition of 'compatible'?

Response: All parameters sets show a reasonable fit as indicated in Figure 5. Compatible is just another way to express that the model simulation results have a good fit with the observations. We use the expression "conditioned on" to replace "compatible" in the revised manuscript.

P10L7 injected on top of the land surface or groundwater table (top of groundwater model).

Response: Thank you. We mean here on top of the groundwater table. This approach is similar to the approach used by Danesh-Yazdi et al., 2016 [1]. Please check Page 8, Line 14.

P10L11: Repetition of P10L7-8.

Response: We deleted one of the sentences as proposed.

P11L1-2: Present how much recharge (mm) each particle represent.

Response: Each particle tracer represents a volumetric recharge rate of around 700 m^3/year. We added this information at the corresponding location in the manuscript (Page 8, Line 16-17).

P11L12-13: Isn't the porosity 0.2 in all model runs? Or was this varied as well?

Response: This value was used for all model runs. We base this information on a prior study by Kohlhepp et al., 2017 [2] who found that the porosity in the study site is quite homogeneous in space.

P11L15: This is repetition

Response: Deleted as proposed.

P12Table2: What is the Composite parameter sensitivity? What does this table show More this table to the Discussion where it is referred to.

Response: The PEST algorithm calculates the sensitivity of model outputs with respect to each parameter corresponding to all observations (with the latter being defined as per user-assigned weights), namely the "composite sensitivity". The composite sensitivity of parameter i is defined as

$$csp_i = \frac{\left[\mathbf{J}^t\mathbf{Q}\mathbf{T}\right]_{ii}^{1/2}}{n}$$

where, J denotes the Jacobian matrix that includes the sensitivities of all predictions to all model parameters, Q is the weight matrix, n is the number of observations with non-zero weights. Please check Page 22, Line 25 – Page 23, Line 1.

P12L18-20: unclear what was exactly done.

Response: We mean to say that the aquifer system is very heterogeneous, and the parameters used in this study are regionalized parameters which are representative for the equivalent homogeneous media. In the revised manuscript, the corresponding sentences do not appear.

P13L3-P14L2: Move to Methods.

Response: Since the concerned part was not directly relevant to the central idea of the manuscript, we decided to move it to the appendix B.

P14L4: Sensitive to the hydraulic conductivity of the Middle muschelkalk?

Response: We mean here that the hydraulic conductivity of Middle Muschelkalk (mm) is highly sensitive to groundwater head observations. We changed the expression in the revised manuscript accordingly (Page 23, Line 3-5).

P14L11-12: You can add that this is 'because the conductivity increases with increasing recharge and keeping the same groundwater head'.

Response: Thanks. We have now added this in the revised version (Page 11, Line 1-2).

P14L15: Whether a RMSE of < 4.6m is sufficient depends on the mean/variation present. For instance, in a flat area this would not be sufficient. What is the mean groundwater level?

Response: The mean groundwater level at this point is about 235 m, and the standard deviation is about 56 m. So we think the error of 4.6 m is within a reasonable bound.

P15Figure6: This Figure is unclear and does not add much to the paper. Consider removing.

Response: This figure was included to provide the reader with a general idea on the flow pathlines and travel times in the study area. We hope some of the readers may find the information given n this figure interesting (we also consulted other publications [1][3] for similar sort of a figure).

P15L2: What is R5K1?

Response: We number the recharge realizations from R1 to R8 from the lowest recharge to the highest recharge realizations, respectively. For each recharge realizations, we conducted the model runs with 50 hydraulic conductivity fields that were numbered from K1 to K50. Accordingly, R5K1 represents the combination of the first hydraulic conductivity realization with the fifth recharge realization. We have clarified this in the revised manuscript (Page 13, Line 21-23).

P15L3: Do the deep low-permeable geological layers act as aquifers in other scenarios? Same for P16L9.

Response: The Muschelkalk formation has been generally considered as aquifer. However, the complex fine-scale, thin-bedded aquifer-aquitard succession makes it difficult to model. The new bore log data showed that the deep low-permeable geological layers (mo2, mm2, mu2) can be aquitard [2]. In this study, they are therefore considered as aquitards for the groundwater simulations. For small number of simulations, mo2 and mm2 are considered as aquifers when the resulting hydraulic conductivities are high.

P15L5: '400 hydraulic conductivity fields': Shouldn't this be 50?

Response: Thank you for pointing out this mistake. We chave hanged the sentence accordingly (Page 14, Line 1-2).

P15L6: Refer to Equation (12?)

Response: Changed as proposed (Page 14, Line 2)..

P15L11: Why is this not surprising based on Eq. 16?

Response: Because Eq. 16 indicates an (inversely) linear dependency between the recharge (J) and the mean travel time. This is coherent with the trend shown in this Figure.

P15L15-16: How is the analytical exponential TTD fitted? Which parameters?

Response: Here we fit the effective storage (S) based on TTDs of the theoretical exponential fit and the detailed GW model. We have revised the corresponding texts to better reflect this part.

P16L23: Figure 9c does not exist.

Response: Sorry for this error. We deleted this sentence.

P16L24-25: The difference between the SAS-functions under different recharge realizations is moderate. But there still is a difference, how do you explain this difference?

Generally it is assumed that SAS functions only react to internal changes (changing groundwater flow paths).

Response: The conductivity fields are also different for different recharge realizations (Figure 4). Therefore, the difference in SAS functions is indeed introduced by internal changes, i.e., variation in hydraulic conductivity realizations. The variation in Ks will lead to a variation of flow paths, which in our case appears to be only moderate because the resulting Ks fields are constrained by the groundwater head observations. We revised the manuscript accordingly (Page 15, Line 19-25).

P17Figure7: Add to the legends what the black line represents. Give the panels clear titles: 'R1, R2'. Currently isn't unclear that the panels are the results from the different recharge scenarios.

Response: Changed as proposed. Please check the updated Figure 7.

P17L2-3: Repetition of P16.

Response: Deleted as proposed.

P17L3-4: The system does not change to a preference of old water. But there is still significant variation (more or less preference for younger water). How do you explain this? Is it not logical to see changes when the hydraulic conductivity of different layers is changed? I would hypothesize that groundwater flow becomes more shallow or deeper as a result, leading to changes in the TTD and SAS functions.

Response: Thank you for this observation. The concerned variation may be caused due to the spatial distribution and velocity of flow pathlines that are controlled by different hydraulic conductivity. For example, a more permeable shallow aquifer layer will gather more flow pathlines in this layer, forming preferential flow pathways, and thus introduce a stronger preference for young water. Particularly, a significant variation in hydraulic conductivities in the deepest geological layer i.e., Lower Muschelkalk (mu1 and mu2), has a pronounced impact on the selection for old water. With a thickness of saturated layer as 100 m, the hydraulic conductivity of the last layer controls how many water parcels can enter into this layer, and how deep the flow paths can develop. This effect can be evidenced by large differences in the SAS functions related to old ages and a relatively smaller difference in those related to young ages (Figure 9b). Please check Page 15, Line 19-25.

P18Figure8b: Why use 1/J? Not J? A lower MTT with higher recharge is obvious, as more water is passing through the system in the same time (same conductivity). Showing the inverse makes this confusing.

Response: We change 1/J to J as proposed. Please check the updated Figure 8b.

P19Figure10: The inset in Figure 10a is unclear. Just give the numbers for the MTTs.

Response: Changed as proposed. Please check the updated Figure 10.

P19: Section 4.4 is very interesting and deserved more space in the paper. What is the effect of spatial changes in flow paths on TTDs and SAS functions?

Response: We have now a detailed discussion related to this analysis as follows:

The difference in TTDs and SAS functions is not induced by the variability in internal hydraulic properties since the two simulations share the same hydraulic conductivity field. Rather it is mainly induced by different spatial distributions for flow paths of particle tracers. The spatially distributed recharge simulated by mHM indicates that the upstream mountainous area has higher recharge rates compared to those in the lowland plain. By construct the uniform recharge neglects this spatial non-uniformity. This difference results in: (a) under uniform recharge scenario, more particle tracers enter the system from locations near the streams at lowland plain, indicating more particle tracers are transported in local flow system rather than in regional flow system [5], and (b) higher recharge rates at lowland plain accelerate the particles' movement in this area and shorten their travel times. As such, local particle flow paths within the shallow aquifer layer at lowland plain (e.g., Middle Keuper) are activated, leading to a stronger preference for selecting local flow paths in shallow aquifer layer, and therefore a stronger preference for young ages. Our findings are in line with the observations by Kaandorp et at. [4], wherein the authors found a relatively higher preference for selection of older water in the upstream area than that in the downstream area of the study catchment.

Please check Page 18, Line 4-15 in the revised manuscript.

P20L4-7: Combine these sentences.

Response: Changed as proposed. This sentence is rewritten as: "In the idealized aquifers where groundwater flow is Dupuit-Forchheimer type, the recharge is uniform, and the aquifer is locally-homogeneous, TTD is controlled by recharge, saturated aquifer thickness and porosity, and is independent of hydraulic conductivity." Please check Page 19, Line 2-5.

P20L10 & L14: Repetition.

Response: deleted as proposed.

P20L14-17: Unclear. Need revision.

Response: We removed some redundant sentences to reflect our ideas in a clear way. Please check Page 19, Line 8-10.

P20L27 & L30: Repetition.

Response: Deleted as proposed.

P20L30: 'sensitive to the spatial pattern of recharge'. This is interesting and deserves more discussion. At the moment it's only presented as a result. But why is the TTD different? What determines this? When does the TTD shift to more younger discharge and when to older? With spatial differences in recharge the assumptions in Eq. 16 are not met.

Response: Thank you for these insightful observations. Following your questions, we formulated the texts and add them in the revised manuscript. It reads as:

The sensitivity of the TTDs and SAS functions on the spatial pattern of recharge forcings can be mainly explained by the different flow paths of particle tracers, resulting mainly from the spatially heterogeneous fields of recharge across the study catchment. For the regional groundwater system, the spatial variation of recharge determines the distribution of starting points of flow pathlines of tracer particles, For example, more particles will be injected from recharge zones which are typically located in high-elevation regions, resulting in a higher weight of flowlines starting from high-elevation regions. The pronounced spatial variability of recharge also controls the systematic (water age) preference for particles existing from the system (to river discharge) that originated from different regions, and therefore exerts a strong control on the shape of the SAS function.

In the study catchment, neglecting spatial variability of recharge results in a smaller mean travel time and a strong preference for discharging young water compared to ones taking the spatial variability of recharge. Such observations are conditioned to site-specific features of the study catchment. It is noticed only when (a) a site is located in a headwater catchment under a humid climate condition, (b) recharge in areas close to the drainage network is generally lower than that in areas far away from the drainage network, and (c) the system is under (near) natural conditions meaning that artificial drainage and pumping do not dominate the groundwater budget.

Please check Page 20, Line 11-24.

P20L35 – P21L1: Repetition.

Response: Deleted as proposed.

P21L6: 'the analytical solution using Eq. 12 may underestimate the MTT': Alway underestimate? Or can it also overestimate the MTT?

Response: Thank you for these questions. This conclusion holds when the simulated TTD has a relatively larger long-tail behavior than the exponential distribution. Such observations have been also reported for (other) real-world aquifers (Eberts et al., 2012;Kaandorp et al., 2018). Please check Page 20, Line 11-24.

P21L12: 'aggregation error' is mentioned in P2L31, and here. Without reading the referred papers, it is unclear what is meant by this. Either remove this, or give more explanation.

Response: Thank you for this suggestion. We use the expression "predictive error" to replace "aggregation error" to avoid misunderstanding. With predictive error, we mean the aggregation error caused by neglecting spatial heterogeneity of inner hydraulic properties. Please check Page 19, Line 32-33.

P21L18: But Figure 9b showed some differences, so it is sensitive? Also, is this conclusion is only valid for homogeneous recharge and conductivity?

Response: This conclusion is valid for not only homogeneous recharge and conductivity, but also for the conditions resulting from non-uniform recharge and heterogeneous conductivity scenarios. We changed this sentence and added more details about the dependency of the SAS functions on external factors: "…We find that the SAS functions are weakly dependent on the hydraulic conductivity fields in the stratigraphic aquifer system, but the overall preference for discharging young water does not change. This weak dependency can be explained by the fact that different realizations of hydraulic conductivity fields modify the spatial distribution of particle flow paths ….".

Please check Page 20, Line 5-9 in the revised manuscript.

P22L27-29: This sentence is unclear. Needs rewriting.

Response: We revised the concerned texts. Please check Page 21, Line 23-25.

P23L1: Conclusion 2 is not a conclusion. An idealized aquifer system is one of the assumptions for the analytical solution.

Response: We deleted Conclusion 2 accordingly.

P23L7: What exactly are the new possibilities? Numerical simulations were already combined with SAS functions, see e.g. these recent papers (also consider referring to these papers and using them in the introduction/discussion): Remondi, F., Kirchner, J.W., Burlando, P., Fatichi, S., 2018. Water Flux Tracking With a Distributed Hydrological Model to Quantify Controls on the Spatiotemporal Variability of Transit Time distributions. Water Resour. Res. 3081–3099. doi:10.1002/2017WR021689 Kaandorp, V.P., de Louw, P.G.B., van der Velde, Y., Broers, H.P., 2018. Transient Groundwater Travel Time Distributions and Age-Ranked Storage-Discharge Relationships of Three Lowland Catchments. Water Resour. Res. 1–18. doi:10.1029/2017WR022461 Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., Fleckenstein, J.H., 2018. Exploring the Dynamics

of Transit Times and Subsurface Mixing in a Small Agricultural Catchment. Water Resour. Res. 2317–2335. doi:10.1002/2017WR021896

Response: Thank you for providing us with these literatures. We have added them into the references. To our knowledge, none of above literatures has investigated the joint effects of (recharge) forcing and (Ks) parameters as comprehensively as we conducted in this study . Our study provides a novel-modeling framework to explore the effect of input uncertainty and parameter equifinality on TTDs and SAS functions through the combination calibration-constrained Monte Carlo parameter generation, numerical model, and SAS function framework. Please check it out at Page 21, Line 29-31.

P23L9-11: As mentioned before, this is one of the interesting observations. Consider adding more detail to this part of the study.

Response: Thank you for this suggestion. We have added more details on this part. Please check it out at Page 18, Line 4-15 and Page 20, Line 11-24.

Technical Corrections

As stated before, the paper needs significant rewriting. It contains many typing errors (e.g. P2L22 '„', P2L24 'StorAge Secletion', P2L26 'the the') which could have been found using a spelling checker, spelling errors (e.g. P2L29 'thorough') and generally language is not up to publication standard.

References: We carefully corrected all the syntax errors. Besides, we polished our language with the help of native speakers.

References:

[1] Danesh-Yazdi, M., Foufoula-Georgiou, E., Karwan, D. L. and Botter, G.: Inferring changes in water cycle dynamics of intensively managed landscapes via the theory of time-variant travel time distributions, Water Resour. Res., 613–615, doi:10.1002/2016WR019091, 2016.

[2] Kohlhepp, B., Lehmann, R., Seeber, P., Küsel, K., Trumbore, S. E., & Totsche, K. U. (2017). Aquifer configuration and geostructural links control the groundwater quality in thin-bedded carbonate-siliciclastic alternations of the Hainich CZE, central Germany. *Hydrology and Earth System Sciences*, *21*(12), 6091–6116. http://doi.org/10.5194/hess-21-6091-2017

[3] Engdahl, N. B., & Maxwell, R. M. (2015). Quantifying changes in age distributions and the hydrologic balance of a high-mountain watershed from climate induced variations in recharge. *Journal of Hydrology*, *522*, 152–162. http://doi.org/10.1016/j.jhydrol.2014.12.032

[4] Kaandorp, V. P., de Louw, P. G. B., van der Velde, Y. and Broers, H. P.: Transient Groundwater Travel Time Distributions and Age-Ranked Storage-Discharge Relationships of Three Lowland Catchments, Water Resour. Res., 1–18, doi:10.1029/2017WR022461, 2018.

[5] Toth, J.: A Theoretical Analysis of Groundwater Flow in Small Drainage Basins 1 of phe low order stream and having similar t he outlet of lowest impounded body of a relatively, J. Geophys. Res., 68(16), 4795–4812, doi:10.1029/JZ068i016p04795, 1963.

Responses to Referee Review 2

We thank the referee reviewer Prof. Dr. Erwin Zehe for his comprehensive and insightful comments. Our responses to the reviewers' comments are given below. The original comments from referee reviewer 1 were marked with blue color, and our response in black.

Summary:

The proposed study explores controls on residence and travel time distributions in a forward coupled model exercise, using a coupled version of the mHm and OpenGeoSys models. Study area is the Naegelstaedt catchment in Germany. The authors explore 8 different recharge scenarios from the mHm which serve as input to the ground water model and which are marked by tracer to tag the path and the age recharge water when it travels through the aquifer to the stream. To this end they generate several realizations of random hydraulic conductivity fields which are constrained to fit a set of distributed head data. The authors compare their simulated travel time distributions to an exponential travel time distribution which is based on an analytical solution, which reveals stronger skewness in the simulated ones. The author do furthermore quantify the uncertainty in average travel time, shed light on the fraction of active to total storage and discuss the age selection of the catchment.

Evaluation: The proposed study has a high scientific significance and I very much like the general approach. Nevertheless, it is in the present form not acceptable, because quite a few important points need further clarification and the presentation quality is up to the standard of HESS.

Response: Thank you very much for your overall assessment to our manuscript as well as for your insightful suggestions. We have revised the manuscript carefully following your suggestions. A revised manuscript will be uploaded soon.

Major points:

- Eq. 9 (the master equation) assumes that storage components of an age tau <T are well mixed. I wonder whether this can be assumed for the selected random fields. This depends strongly on the correlation lengths and the total extent of the domain and maybe even more on the question whether preferential flow paths are present here?

Response: Thank you for the comments. Eq. 9 (the master equation) is the fundamental formula for connecting conservation of mass and water age. In general, it does not rely on any presumed mixing hypothesis (Botter et al. 2011). Nevertheless, we agree with the reviewer that the well-mixed assumption needs to be made to derive the analytical solution.

We agree with the reviewer that the random K_s fields used in this study do not guarantee a well-mixed storage. Actually, our study is designed to investigate that in a real-world catchment, how skewed is the shape of the simulated TTD compared to the well-mixed TTD, and how the waters particles with different ages are discharged into streams.

The well-mixed assumption is valid when the aquifer is homogeneous, and the drop in the water table between the maximum and minimum head is small compared with the aquifer total depth. Otherwise, the SAS function can deviate from the well-mixed scheme and take on complex shapes even in the saturated region of a homogeneous aquifer depending on the bed form (Van Der Velde et al. 2012).

Are they present? And what is the correlation length of the generated random fields, and the nugget to sill ratios? How did you assess this information and did you vary them between the realizations? Or this is uncorrelated noise?

Response: This is a misunderstanding in the K_s fields. We apologize for the unclear description of the random fields in the original manuscript. We would like to elaborate more on the hydraulic conductivity (K_s) fields used in this study. The K_s fields are not based on geostatistical interpolation. They are based on zonation, whereby parameter values are assigned as piecewise constant values to defined areas (zones) in the model domain (Anderson P. 2002). Spatial changes in parameter values occur only among zones. Delineation of zones relies on information contained in the hydrogeological investigation that identifies areas where parameters are likely to be the same. The geometric mean of expected values of a given parameter within the zone is assigned to the zone if heterogeneity is thought to be random, which means the variance and correlation length are not included in this approach (Anderson P. 2002). This zoned aquifer system indicates that water particles can go through more-permeable zones (i.e. layers with high Ks values) more easily than low-permeable zones, thus forming preferential flow pathways in more-permeable layers. To avoid this misunderstanding, we revised the manuscript accordingly. Please check out the revised manuscript.

On the basis of the points stated above, the random fields do not follow the well-mixed assumption. Alternatively speaking, the well-mixed scheme is a baseline scenario for quantifying the transport dynamics in a complex real-world catchment. The influence of spatial variability of input forcings in the systematic preference for waters with different ages is also investigated in this study, which has not been investigated in a real-world catchment before based on our knowledge.

- There might be a conceptual problem, depending on what your particles shall actually represent. In case the particles shall mark the travel path of water (not of a solute) I think they should move in a purely advective manner, which means that eq. 6-8 need to be different. There is not diffusive mixing among water molecules (as long as we neglect different isotopic compositions). Or do they mark the fraction of different water

isotopes, than this should be stated? But in this case I wonder where the dispersivity does stem from? Other tracers?

Response: Thank you again for this comment. We fully agree with the reviewer that in the case that particle tracer represents the water rather than the solute, the dispersion process should be ignored. The random walk particle tracking algorithm is capable to deal with reactive transport problems. Therefore in the original manuscript, Eq. 6-8 are written in their full form to incorporate both diffusion and advection processes, but we only consider the advection process in this study. Actually, we clarified this point already in the original manuscript: *"In this study, we focus on the predictive uncertainty within the convection process. Therefore, the molecular diffusion coefficients are universally set to 0 for all ensemble simulations."* (Page 11, Line 5-7 in the original manuscript).

The particle tracking scheme used in this study is capable to simulate both diffusion and advection processes, therefore Eq. 6-8 were written in a complete form to incorporate both processes. The velocity component in these three equations, namely V_x, V_y, and V_z, are essentially different.

We revised the manuscript according to the reviewer's comments. In the revised manuscript, these equations were moved to the appendix in order to make the structure clear and avoid misunderstanding.

- The recharge amount, the generated parameter fields and base-flow production are not independent. I see that the ks parameter field is adjusted such that the generated parameter sets match the head data (which is by the way not so difficult). But to have a consistent model the simulated base-flow production from OpenGeoSys needs to match the simulated base-flow of the mHm (which is calibrated to stream flow). A consistent match of both the head and the base-flow is crucial for credibility of the model structure and it's ability to simulate travel time distributions for the selected system.

Response: Thank you again for this important observation. We completely agree with the reviewer that the recharge, the generated parameter fields and the baseflow production are not independent. As the reviewer pointed out, both baseflow and groundwater heads should be matched to close the water budget and achieve realistic parameter values. This is also what we did in this study. Because for the steady-state system, the total amount of inflow (i.e. groundwater recharge) equals to the total amount of outflow (baseflow in this case) for the OGS groundwater model. Given that the recharge is directly taken from mHM, the baseflow is also consistent with the one estimated by mHM. The water budget is naturally closed. We addressed and discussed in more detail in our previous study (Jing et al. 2018).

Following the reviewer's advice, we also clarified this point in the revised manuscript: "*For the steady-state system and the one-way coupled model, outflow from aquifer to the streams (i.e. baseflow) proves to be consistent with the baseflow originally estimated*

by mHM, implying that the water budget in the subsurface system is essentially closed (Jing et al. 2018).” Please check it out at Page 6, Line 30 – Page 7, Line 2.

We also agree with the reviewer that the K_s fields also have an impact on the recharge. We admit this influence was not considered in this study, because the one-way coupled model is not capable to include such a two-way interaction. We fully acknowledge this limitation in our previous paper, where we also discuss some of its ramifications (Jing et al. 2018).

Technical details

- The control for contamination is in fact the Dammkoehler number, which relates residence times and degradation time scales.

Response: We fully agree that the Dammhoehler number is the relevant scaling number for reactive transport processes. However, in this study, we deal with water flow only. As a result, we do not consider it to be necessary to include this number in our discussion.

- Eq. 9: PQ (T,t) is a exceedance probability (otherwise this does not make sense).

Response: Agreed. $P_Q(T,t)$ is the exceedance probability.

- Eq. 9 What is Qj and what is N- the number of different "outlets"?

Response: Exactly. Q_j is the *j*-th outflux, and N is the total number of outlfluxes.

- I have problems with the terminology of a "StorAge selection" function (even if it is established), as the stream doesn’t do an active select water of different ages.

Response: As this term has been widely used by many other researchers, we simply apply the same name with them. Maybe this terminology is tricky, it is beyond our capability to judge whether this name is reasonable or not.

- Preferential flow does not necessarily mean that Peclet number is large, if the flow is still in the near field and mixing among the flow paths is small. There is literature evidence for this.

Response: Agreed. We also found out that this statement is not directly relevant to the main idea of this paper, so we deleted this sentence in the revised manuscript.

- Eq. 6 - 8: Z is a Gaussian random number, otherwise the coefficient in below the root is 1/6.

Response: Agreed. We changed it as proposed (Page 22, Line 8).

- Parts of the section 4.1 should be shifted into the methods section!

Response: Thank you for this observation. We changed it as proposed.

- Page 14: Figure 5 is a scatter plot of heads (simulated and observed) not of the head residuals.

Response: Changed as proposed (Page 11, Line 1-2).

- Page 6 line 5: Repetitive statement on the TTD of the soil?

Response: We deleted the repetitive statement as proposed by the reviewer.

- Not sure what is meant with "backward travel time distribution"?

Response: The backward travel time distribution complies with the problem of how a sample of water taken at a time t is the result of transport processes that involve inputs generated from all previous times (Benettin, Rinaldo, et al. 2015).

TTDs can be interpreted in two different ways, depending on whether they track ages forward or backward in time. In ‘‘forward’’ tracking, one selects a given particle injection at a fixed time t_i and follows the subsequent exit times. In ‘‘backward’’ tracking, instead, one focuses on a given exit time t_{ex}, considers the particles that leave the system at t_{ex} and then tracks their various entrance times backward in time (Benettin, Kirchner, et al. 2015). In this study, we only use backward distributions.

- Page 8 line 20: how are they interpolated?

Response: We use a bilinear interpolation approach. Following the reviewer's suggestion, we modified the manuscript as follows: *"The gridded recharges estimated by mHM are interpolated and then assigned to each grid nodes on the upper surface of OGS mesh using a bilinear interpolation approach."* Please check Page 6, Line 26-28.

- Figure 1: Caption is not self-explaining: what is mo, mu, mm etc?

Table 1: Please explain km and ku.

Response: Thank you again for this observation. Mo, mu, and mm stand for three geological zones -- Upper Muschelkalk, Middle Muschelkalk, and Lower Muschelkalk, respectively. Km and ku stand for the Middle Keuper and Lower Keuper. We added the full name of these geological zones into the Figure 1 and Table 1 as you suggested.

-What is the estimation variance of the mean you calculated (based on the standard deviation and the sample size), might be nice to add this to Figure 8.

Response: Thank you for this suggestion. Following this suggestion, we added the variance of the mean travel time (MTT) into the Figure 8 b). Please check it out in the updated manuscript.

- I think the paper would greatly benefit from a thorough proof reading.

Response: We did a thorough proofreading already together with a native speaker. Please check out the revised manuscript, which will be uploaded soon to HESS.

References:

Anderson P., W.W. and H.J., 2002. *Applied Groundwater Modeling*, Available at: http://www.sciencedirect.com/science/article/pii/B9780080886947500092.

Benettin, P., Kirchner, J.W., et al., 2015. Modeling chloride transport using travel time distributions at Plynlimon, Wales. *Water Resources Research*, pp.3259–3276.

Benettin, P., Rinaldo, A. & Botter, G., 2015. Tracking residence times in hydrological systems: Forward and backward formulations. *Hydrological Processes*, 29(25), pp.5203–5213.

Botter, G., Bertuzzo, E. & Rinaldo, A., 2011. Catchment residence and travel time distributions: The master equation. *Geophysical Research Letters*, 38(11), pp.1–6.

Jing, M. et al., 2018. Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwater model OpenGeoSys (OGS). *Geoscientific Model Development*, 11(5), pp.1989–2007. Available at: https://www.geosci-model-dev.net/11/1989/2018/.

Van Der Velde, Y. et al., 2012. Quantifying catchment-scale mixing and its effect on time-varying travel time distributions. *Water Resources Research*, 48(6), pp.1–13.

List of main changes made in the manuscript

1. A thorough proofreading has been made together with a native English speaker.
2. Figure 1, 7, 8 and 10 were modified according to the reviewers' suggestions.
3. The descriptions on RWPT method were moved from "Methodology and Materials" to "Appendix A".
4. The order of paragraphs has been shifted as suggested by reviewer 1. The updated manuscript starts with Site Description, then Numerical Model and Model setup.
5. Some redundant sentences were removed to shorten the length of the manuscript.
6. As suggested by reviewer 1, the discussion on the sensitivity of TTDs on spatial pattern of forcings has been enhanced (see section 5.3 in the revised manuscript).
7. The paragraphs about StorAge Selection function have been double-checked and several references were added.
8. Conclusion 2 in the original manuscript was deleted.

Influence of input and parameter uncertainty on the prediction of catchment-scale groundwater travel time distributions

Miao Jing[1,2], Falk Heße[1], Rohini Kumar[1], Olaf Kolditz[3,4], Thomas Kalbacher[3], and Sabine Attinger[1,5]

[1]Department of Computational Hydrosystems, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
[2]Institute of Geosciences, Friedrich Schiller University Jena, Burgweg 11, 07749 Jena, Germany
[3]Department of Environmental Informatics, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
[4]Applied Environmental Systems Analysis, Technische Universität Dresden, Dresden, Germany
[5]Institute of Earth and Environmental Sciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

Correspondence: Miao Jing (miao.jing@ufz.de); Falk Heße (falk.hesse@ufz.de)

Abstract. Groundwater travel time distributions (TTDs) provide a robust description of the subsurface mixing behavior and hydrological response of a subsurface system. Lagrangian particle tracking is often used to derive the groundwater TTDs~~, but such approach may suffer from~~ . The reliability of this approach is subjected to the uncertainty of external forcings, internal hydraulic properties, and the interplay between them. Here, we evaluate the uncertainty of catchment groundwater TTDs in an agricultural catchment using a 3-D groundwater model with an overall focus on revealing the relationship between external forcing, internal hydraulic ~~property~~properties, and TTD predictions. ~~A stratigraphic aquifer model is applied to represent the spatial structure of the aquifer. Several~~ Eight recharge realizations are sampled from a high-resolution dataset of land surface fluxes and states. ~~Constrained to expert knowledge and groundwater head observations, many realizations of~~ Calibration-constrained hydraulic conductivity fields (K_s fields) are stochastically generated using the null-space Monte Carlo (NSMC) method for each recharge realization. The random walk particle tracking (RWPT) method is used to track the pathways of particles and compute travel times. Moreover, an analytical model under the random sampling (RS) assumption is fitted against the numerical solutions, serving as a reference for the mixing behavior of the model domain. The StorAge Selection (SAS) function is used to interpret the results in terms of quantifying the systematic preference for discharging young/old water. The simulation results reveal the ~~dominant role of recharge in controlling the TTD predictions~~primary effect of recharge on the predicted mean travel time (MTT). The different realizations of calibration-constrained ~~hydraulic conductivity~~ K_s fields moderately magnify or attenuate the predicted ~~mean travel time (MTT), provided that most parameters can be well constrained to the data~~MTTs. The analytical ~~solution under a random sampling assumption~~ model does not properly replicate the numerical solution, and it underestimates the mean travel time. ~~The SAS functions of ensemble simulations~~ Simulated SAS functions indicate an overall preference for young water ~~in the saturated zone~~ for all realizations. The spatial pattern of recharge ~~also has a strong impact on~~ controls the shape and breadth of simulated TTDs and SAS functions by changing the spatial distribution of particles' pathways. In conclusion, overlooking the spatial nonuniformity and uncertainty of input (forcing) ~~uncertainty~~ will result in biased travel time predictions~~and may underestimate the overall uncertainty in groundwater TTDs. We emphasize the~~

dominant role of recharge estimation (including the spatial pattern) in the TTD prediction, the . We also highlight the worth of reliable observations in reducing predictive uncertainty , and the good interpretability of SAS function functions in terms of understanding catchment transport processes.

1 Introduction

Travel/transit time distribution (TTDdistributions (TTDs) of groundwater provides an provide a description of how aquifers store and release water and pollutants under external forcing conditions. It , which has significant implications for interdisciplinary environmental studies. For example, remarkable time-lags of the reaction of streamflow to with outer forcings and considerable amounts of "old water" (i.e., water with an age of decades or longer) in streamflow have been observed in many studies (Howden et al., 2010; Stewart et al., 2012). Moreover, the legacy nitrogen in groundwater storage may dominate the annual nitrogen loads in agricultural basins (Van Meter et al., 2016; Van Meter et al., 2017)(Wang et al., 2016; Van Meter et al., 2016; Van Meter et . Groundwater TTDs offer important insights on into the vulnerability of aquifers to pollution spreadingand is , and they are critically important for the environmental assessment of non-point source non-point-source agricultural contamination (Böhlke and Denver, 1995; Böhlke, 2002; Molnat and Gascuel-Odoux, 2002; Eberts et al., 2012). TTDs shed light on the quantification of the long-term influence of the agricultural contamination, which is crucial for the water quality water quality and sustainability.

The accurate quantification of groundwater travel time at a regional scale is extremely challenging. A primary difficulty is that the complex geometric, topographic, meteorologic, and hydraulic properties of hydrologic systems control the flow and mixing processes , and therefore define the unique shape of the travel time distribution (TTD) (Leray et al., 2016; Hale and McDonnell, 2016; Engdahl et al., 2016). The other difficulty is that the groundwater system is intricately and tightly coupled to the land surface hydrologic processes. The fundamental characteristics and the coupled nature determine the response of a catchment to outer forcings such as anthropogenic climate change, artificial abstraction, and agricultural and chemical contamination (Tetzlaff et al., 2014; van der Velde et al., 2015; Heße et al., 2017).

Groundwater TTDs cannot be measured directly. They need to be inferred from observations and measurements using lumped analytical models, or they can be directly simulated using fully-distributed numerical models. The analytical models (i.e., The techniques for determining groundwater TTDs can be categorized into two groups: geochemical approaches and numerical modeling approaches (McCallum et al., 2014). In geochemical approaches, the lumped parameter models) have been widely are often used to interpret the catchment-scale measurements (e.g., discharge, environmental tracer)observation of an environmental tracer concentration. Environmental tracer datasets include those representing can be divided into those representing the concentration distribution of young water (e.g., ^{3}H, , $SF_6$$SF_6$, ^{85}K, and CFCs) and those representing the concentration distribution of old water (e.g., ^{36}Cl, ^{4}He, ^{39}Ar, and ^{14}C). Among the analytical modelsAdditionally, the analytical StorAge Seeletion Selection (SAS) function is a cutting-edge tool to characterize for characterizing transport processes in lumped, time-varying hydrologic systems at the hillslope/catchment scale (Botter et al., 2011; Rinaldo et al., 2011; Van Der Velde et al., 2012; Harman, 2015; Danesh-Yazdi et al., 2018). This framework provides a clear distinction between the travel

time (the ~~the~~ time spent by a water parcel or a solute from its entrance to the control volume ~~till~~ until its exit) and the residence time (the age of the water parcel or the solute ~~exists~~ existing in the control volume at a particular time). The SAS function has been successfully applied to interpret environmental tracer data ~~thorough~~ through some assumptions of the mixing mechanism (Benettin et al., 2015, 2017). However, analytical approaches ~~falls~~ fall short in representing the dispersion of transport ~~process~~ processes caused by catchment heterogeneity. Strong heterogeneity leads to significant aggregation error of mean travel times (~~MTT~~MTTs) when using analytical models to interpret the tracer data (Kirchner, 2016; Stewart et al., 2016).

In contrast to such an analytical approach, ~~physically-based~~ physically based numerical models can explicitly describe the geometry, topography, and geological structures, and they can represent the flow paths of individual water particles. ~~Physically-based~~ Physically based numerical models are structurally complex and computationally expensive ~~,~~ and often have more parameters ~~compared to lumped parameter model. They~~ than lumped parameter models do. These models can be classified as Eulerian ~~approach or Lagrangian approach~~ approaches or Lagrangian approaches (Leray et al., 2016). The Eulerian approach directly solves the partial differential equations (PDEs) ~~deriving~~ derived from mass conservation with "age mass" as the primary variable (Goode, 1996; Ginn, 2000; Engdahl et al., 2016). The Lagrangian approach, including ~~Smoothed Particle Hydrodynamics~~ the smoothed particle hydrodynamics (SPH) approach and ~~Random Walk Particle Tracking~~ the random walk particle tracking (RWPT) approach, is numerically robust and less restrictive on time-step size in solving advection-dominated problems (Tompson and Gelhar, 1990). Consequently, Lagrangian methods are more promising in simulating complex real-world transport ~~process~~ processes, as they avoid spurious mixing error in grid-fixed Eulerian methods (Benson et al., 2017). Therefore, the Lagrangian approach has been widely used to simulate large-scale reactive transport and biogeochemical problems (Park et al., 2008; de Rooij et al., 2013; Selle et al., 2013).

A reliable application of groundwater transport modeling is ~~subjected to many uncertainty sources~~ subject to many sources of uncertainty, including measurement, model structural, and parameter uncertainty (Beven, 1993). Specifically, the reliability of model prediction suffers from the uncertainty of external forcings, the uncertainty of internal hydraulic characteristics, and the interplay between them (Ajami et al., 2007). The spatially sparse measurements of recharge lead to a biased characterization of ~~spatio-temporal~~ spatiotemporal patterns of recharge (Healy and Scanlon, 2010; Cheng et al., 2017). On the other hand, the spatial scarcity of hydrogeological data always ~~threaten~~ hampers the right characterization of aquifer properties such as porosity and permeability, thus allowing a range of various realistic parameter values. The ~~combination of expert knowledge and parameterization is generally recommended in hydrogeological modeling. Hydrologic models, no matter surface models, groundwater models, or integrated surface/subsurface models, are typically calibrated against single target (e.g., catchment discharge, tracer data, or groundwater heads) to achieve a set of plausible parameters by minimizing the residuals between observation and simulation. The~~ best-fit parameter may suffer from a fitting error caused by overparameterization and equifinality (Schoups et al., 2008). Such biased parameters cause uncertain predictions because parameter error may compensate for model structural ~~defect~~ defects (Doherty, 2015). Accordingly, predictive uncertainty can be hardly assessed in a precise way.

~~The uncertainty of groundwater recharge and hydrogeological configuration lead to a biased~~ Biased characterization of the hydrodynamic system ~~, and further lead to problematic TTD prediction~~ and oversimplified assumptions will lead to a problematic prediction of TTDs. Many past studies offer insights into the influence of recharge and hydrogeological configu-

ration on the prediction of TTDs. For example, ~~La Venue et al. (1989) evaluated the groundwater travel time uncertainty using the sensitivity derivatives.~~ some research studies have been devoted to the development of analytical solutions for the idealized catchment (or aquifer) under some essential assumptions and simplifications (Neuman; Haitjema, 1995; Engdahl et al., 2016; Leray et al., 2 . Among them, Haitjema (1995) derived an analytical solution in an idealized groundwatershed under steady-state conditions and the Dupuit-Forchheimer assumption ~~,~~ and found that the groundwater mean travel time ~~seems~~ appears to be only dependent on recharge, saturated aquifer thickness, and porosity, provided that the hydraulic conductivity is locally homogeneous. ~~Fiori and Russo (2008) established a 3-D numerical model to study the TTDs at a hillslope and its dependence on realistic features of the study area, and~~ Basu et al. (2012) evaluated analytical, GIS, and numerical approaches for the prediction of groundwater TTDs and found that the ~~TTDs are weakly dependent on the heterogeneity of permeability. Selle et al. (2013) analyzed the sensitivity of groundwater travel times on recharge and discharge by defining four different recharge and discharge scenarios. Leray et al. (2016) presented the theoretical background generating travel time distributions under steady-state condition, and reviewed analytical solutions conditioned by both simple and complicated assumptions. The water flow and transport may be dominated by advection where the mixing is minimal (i.e., Peclet number is high), such as~~ simulated TTDs show a moderate difference. Many recent studies have reported the dependency of transient TTDs on the ~~preferential flow. Diffusion is an important driving force within the majority of catchments because local-scale heterogeneities of velocity field cause mass dispersion at the catchment scale (Berne et al., 2005; Bear, 2013)~~temporal variability of input forcings (Benettin et al., 2015; Yang et al., 20 , but the dependency of TTDs (as well as SAS functions) on the spatial variability of input forcings has rarely been studied.

Although studies on catchment-scale groundwater TTDs are ~~plenty, the~~ numerous, comprehensive uncertainty analysis ~~of TTD predictions aiming~~ that aims to unveil the different roles of external forcing and internal ~~hydraulic characteristics using both~~ hydrostratigraphic structure using both a numerical model and SAS functions is scarce. In this regard, two important questions are the following: (1) How does the uncertainty of recharge (including its spatial nonuniformity) and hydraulic conductivities affect the TTD predictions in a mesoscale agricultural catchment, provided that the model is constrained to reality and groundwater head observations? (2) How does the uncertainty of inputs (forcings) and parameters influence the prediction of systematic preference for young/old water?

In this paper, we aim to answer these questions through a detailed (uncertainty) analysis of an example application in a mesoscale ~~catchment. For~~ real-world catchment. In doing so, we ~~established~~ establish a detailed groundwater model ~~in a mesoscale agricultural catchment~~ coupled to a random walk particle tracking system for predicting groundwater TTDs. The groundwater model OpenGeoSys (OGS) is used to simulate the groundwater flow, while the input forcing is fed by the mesoscale hydrologic model (mHM) via the coupling interface mHM-OGS (Jing et al., 2018). The numerical model follows the steady-state assumption of groundwater flow systems. This assumption is made because at the regional scale, the groundwater flow process has a much ~~bigger~~ larger time scale than that of the high-frequency oscillation of recharge, which essentially dampens the effect of recharge oscillation (Leray et al., 2016). ~~A combination of different input forcing realizations as groundwater recharge and parameter realizations as hydraulic conductivity fields that are all compatible with the observations are applied in this study~~An ensemble of simulations using multiple recharge realizations and multiple equifinal hydraulic conductivity (K_s) fields is established. An analytical model is used as a reference for unveiling the mixing mechanism of the system. The StorAge

Selection function is also used to interpret the simulation results of the numerical model, with an overall aim to quantify the predictive uncertainty of systematic preference for young/old water.

2 Site Description

The candidate site in this paper is the Nägelstedt catchment, located in central Germany (see Figure 1). With an area of approximately 850 km^2, the Nägelstedt catchment is a headwater catchment of the Unstrut river. The terrain elevation of this area varies from 164 m to 516 m a.m.s.l. (above mean sea level). It is a subcatchment of the Unstrut basin, one of the most intensively used agricultural regions in Germany. About 88% of the land in this site is marked as arable land, which is significantly higher than the average level of Thuringian (Wechsung et al., 2008). The agricultural nitrogen input has varied over the years and locations, from 5 - 24 kg/ha in the soils of the lowlands to 2 - 30 kg/ha in the feeding area (Wechsung et al., 2008). The mean annual precipitation is approximately 660 mm.

The dominating sediment in the study area is the Muschelkalk (Middle Triassic). The Muschelkalk has an overall thickness of about 220 m, and it has been divided into three subgroups according to mineral composition: Upper Muschelkalk (mo), Middle Muschelkalk (mm), and Lower Muschelkalk (mu). The Upper Muschelkalk (mo) is mainly composed of limestone, marlstone and claystone, and it forms fractured aquifers (**?**Jochen et al., 2014; Kohlhepp et al., 2017). The Middle Muschelkalk (mm) deposits are composed of evaporites, including dolomit marlstone, gypsum, dolomit limestone and eroded salt layers. The Lower Muschelkalk (mu) is composed of massive limestone (**?**McCann, 2008; Jochen et al., 2014). The Muschelkalk formation consists of limestone sediments, which may form fractured and karst aquifers. A recent study demonstrated that karstification and the development of conduits are limited at the base of the Upper Muschelkalk at the Hainich critical zone in the Nägelstedt catchment (Kohlhepp et al., 2017). In the middle of the study area, Keuper deposits, including Middle Keuper (km) and Lower Keuper (ku), overlay the Muschelkalk formation. The Lower Keuper formation forms the low permeable aquitard on top of the Upper Muschelkalk aquifer (**?**Kohlhepp et al., 2017). Lithologically, the Muschelkalk aquifer system is a "layer-cake" aquifer system that contains interbedded marlstone aquitards (Aigner, 1982; Merz, 1987; Kohlhepp et al., 2017) .

Eighteen monitoring wells distributed in this area are used to calibrate the model (Figure 1a, in which the well W0 is abandoned in this study due to the proximity to the outer edge). The geological layers to which the wells belong are listed as follows: 5 wells in Middle Keuper (km), 4 in Lower Keuper (ku), 6 in Upper Muschelkalk (mo), 2 in Middle Muschelkalk (mm), and 1 in alluvium.

3 Methodology and Materials

3.1 Numerical Model

We use the coupled model mHM-OGS, proposed by Jing et al. (2018), to simulate terrestrial hydrological processes. This coupled model was developed for extending the predictive capability of mHM from land surface processes to the subsurface

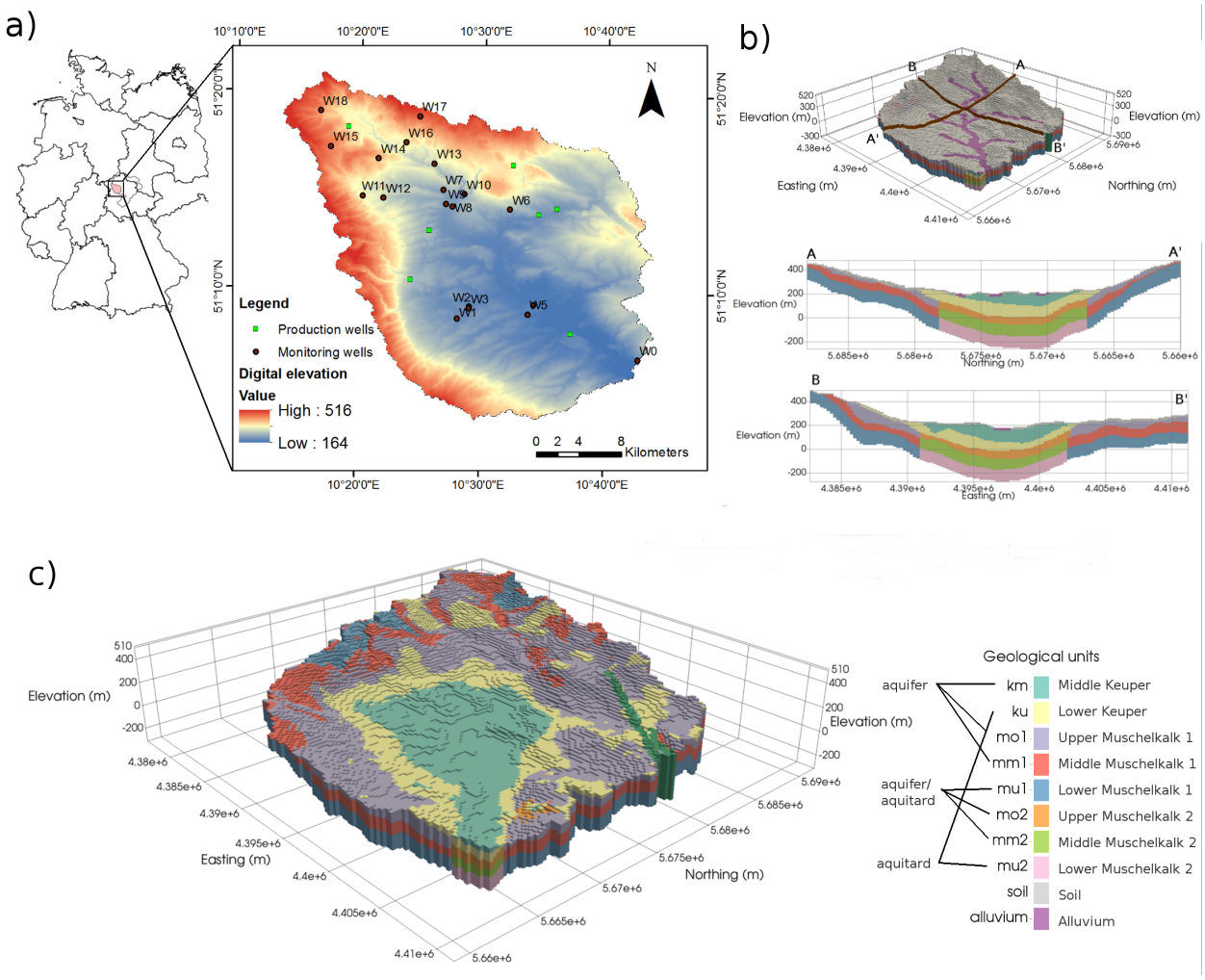

Figure 1. The Nägelstedt catchment used as the test catchment for this study (Jing et al., 2018). a) An overview of the Nägelstedt catchment and the locations of the monitoring wells used in this study. b) 3-D view highlighting the arrangement of alluvium and soil and cross-sectional view of the study area. c) 3-D view highlighting the zonation of the sedimentary aquifer-aquitard system. Note that the Muschelkalk layers (mo, mm and mu) are divided into more permeable subunits (mo1, mm1 and mu1) and less permeable subunits (mo2, mm2, and mu2).

flow and transport processes. Specifically, ~~the mesoscale Hydrologic Model (mHM)~~ mHM (Samaniego et al., 2010; Kumar et al., 2013) is used to partition water budget components, while the porous media simulator ~~OpenGeoSys (OGS)~~ OGS (Kolditz et al., 2012) is used to compute groundwater flow and transport processes by using mHM-generated recharge as driving forces. ~~For details of~~ For details on the coupled model mHM-OGS, please refer to Jing et al. (2018).

The catchment water storage is conceptually partitioned into soil zone storage and deep groundwater storage~~. The~~ ; the two corresponding components are computed by mHM and OGS, respectively. The soil zone dynamics of TTD has been well

studied using mHM in a previous work (Heße et al., 2017). Hence, in this paper, we perform explicit forward modeling of the saturated-zone TTD through a 3-D OGS groundwater model by using the mHM-generated recharge as the external forcing.

In this study, we focus on the travel times in the saturated zone. Saturated groundwater flow is characterized by the continuity equation and Darcy's law:

$$S\frac{\partial \psi_p}{\partial t} = -\nabla \cdot \boldsymbol{q} + q_s \tag{1}$$

$$\boldsymbol{q} = -K_s \nabla(\psi_p - z) \tag{2}$$

where S is the specific storage coefficient in confined aquifers, or the specific yield in unconfined aquifers [1/L]~~,~~; ψ_p is the pressure head in the porous medium [L]~~,~~; t is the time [T]~~,~~; $\boldsymbol{q}$ is the specific discharge or the Darcy velocity [LT^{-1}]~~,~~; q_s is the volumetric source/sink term [T^{-1}]~~,~~; K_s is the saturated hydraulic conductivity tensor [LT^{-1}]~~,~~; and z is the vertical coordinate [L].

3.2 ~~Random Walk Particle Tracking~~

We use the ~~Random Walk Particle Tracking (RWPT)~~ RWPT method to track the particle movement. The RWPT method is embedded in the source codes of OGS (Kolditz et al., 2012; Park et al., 2008). ~~The functionality of RWPT method of OGS has been validated by Park et al. (2008). RWPT method is derived~~ Derived from stochastic physics~~. The basic idea of this method is to assume that advection process is deterministic and~~ , RWPT is under the assumption that the advection process is deterministic and the diffusion-dispersion ~~process is stochastic. RWPT solves a diffusion equation at local Lagrangian coordinates rather than the classical advection-diffusion equation, which can be expressed as:~~

$$\mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + \mathbf{v}(\mathbf{x}(t_{i-1}))\Delta t + Z\sqrt{2\mathbf{D}(\mathbf{x}(t_{i-1})\Delta t}$$

~~where x denotes the coordinates of the particle location, Δt denotes the time step size, and Z denotes a random number with the mean being zero and variance being unity.~~ process is stochastic. The theoretical background of the RWPT method is described in detail in Appendix A.

~~The velocity v in Eq. (A1) is replaced by $\mathbf{v}_i^*$ to keep consistency with the classical advection-dispersion equation (Kinzelbach, 1986) . The expressions of $\mathbf{v}_i^*$ and the hydrodynamic dispersion tensor $\mathbf{D}_{ij}$ are:~~

$$\mathbf{v}_i^* = \mathbf{v}_i + \sum_{i=1}^{3} \frac{\partial \mathbf{D}_{ij}}{\partial x_{ij}}$$

$$\mathbf{D}_{ij} = \alpha_T |\mathbf{v}| \delta_{ij} + (\alpha_L - \alpha_T)\frac{\mathbf{v}_i \mathbf{v}_j}{|\mathbf{v}|} + \mathbf{D}_{ij}^d$$

~~where δ_{ij} denotes the Kronecker symbol, α_L denotes the longitudinal dispersion length, α_T denotes the transverse dispersion length, $\mathbf{D}_{ij}^d$ denotes the tensor of molecular diffusion coefficient and $\mathbf{v}_i$ denotes the mean pore velocity component at the ith direction.~~

The stochastic governing equation of 3-D RWPT can therefore be expressed as:

$$x_{t+\Delta t} = x_t + \left(V_x(x_t,y_t,z_t,t) + \frac{\partial D_{xx}}{\partial x} + \frac{\partial D_{xy}}{\partial y} + \frac{\partial D_{xz}}{\partial z}\right)\Delta t + \sqrt{2D_{xx}\Delta t}Z_1 + \sqrt{2D_{xy}\Delta t}Z_2 + \sqrt{2D_{xz}\Delta t}Z_3$$

$$y_{t+\Delta t} = y_t + \left(V_y(x_t,y_t,z_t,t) + \frac{\partial D_{yx}}{\partial x} + \frac{\partial D_{yy}}{\partial y} + \frac{\partial D_{yz}}{\partial z}\right)\Delta t + \sqrt{2D_{yx}\Delta t}Z_1 + \sqrt{2D_{yy}\Delta t}Z_2 + \sqrt{2D_{yz}\Delta t}Z_3$$

$$z_{t+\Delta t} = z_t + \left(V_z(x_t,y_t,z_t,t) + \frac{\partial D_{zx}}{\partial x} + \frac{\partial D_{zy}}{\partial y} + \frac{\partial D_{zz}}{\partial z}\right)\Delta t + \sqrt{2D_{zx}\Delta t}Z_1 + \sqrt{2D_{zy}\Delta t}Z_2 + \sqrt{2D_{zz}\Delta t}Z_3$$

where x, y, z are the spatial coordinates of particle, Δt is the time step, Z_i is a random number with a mean of zero and a unit variance.

3.2 Travel Time Distribution and StorAge Selection function

The travel time is defined as the time spent by a moving element (e.g., either a water particle or a solute) in a control volume of a hydrologic system. In principle, the control volume can be defined at arbitrary spatial scales (i.e., from molecular scale to regional scale). We followed the conceptualization of hydrologic systems from Botter et al. (2011) and Benettin et al. (2017), and partitioned the subsurface water storage into two conceptual storages: the shallow soil zone storage and the deeper groundwater storage. The TTD of soil zone storage has been comprehensively studied by Heße et al. (2017). Therefore, we focuses on the travel time of the groundwater storage in this study.

Considering a hydrologic system where the input flux (J) and the output flux ($Q_1, Q_2, ..., Q_n$) are known, each parcel of water within the system is tagged using its current age τ. The age-ranked storage $S_T = S_T(T,t)$ is defined as the mass of water in the system with age $\tau < T$. The backward form of the Master Equation (ME) for TTD in a control volume can be expressed as follows (Botter et al., 2011; Van Der Velde et al., 2012; Harman, 2015):

$$\frac{\partial S_T}{\partial t} = J(t) - \sum_{j=1}^{n} Q_j(t)\overleftarrow{P}_{Q_j}(T,t) - \frac{\partial S_T}{\partial T}$$

with boundary condition $S_T(0,t) = 0$, where $\overleftarrow{P}_{Q_j}(T,t)$ is the cdf of backward travel time distribution of output flux Q_j, $J(t)$ is the input flux at time t, and $Q_j(t)$ is the output flux at time t. Specifically in this study, J is the groundwater recharge, and Q is composed of two components: the stream baseflow and the abstraction at production wells. The streams and production wells are considered as the only sources of output fluxes.

On the basis of Eq. (3), the backward travel time distribution $\overleftarrow{P}_Q(T,t)$ can be calculated from the SAS function by the mapping from T to S_T:

$$\overleftarrow{P}_Q(T,t) = \Omega_Q(S_T,t)$$

for $S_T = S_T(T,t)$. Ω_Q is the cumulative form of StorAge Selection (SAS) function.

In case the age distribution of each outflow is uniformly selected from all water storages with various ages, the outflux TTDs turn into a random sample (RS) of the storage RTD. The uniform SAS function become $\Omega_Q(S_T,t) = S_T(T,t)/S(t)$. Eq. (4)

in this case has the analytical solution:

$$p_S(T,t) = \overleftarrow{p}_Q(T,t) = \frac{J(t-T)}{S(t)} \exp\Big[- \int_{t-T}^{t} \frac{Q(\tau)}{S(\tau)} d\tau \Big]$$

where $p_S(T,t)$ is the pdf of residence time distribution, $S(t)$ is the storage at time t. Specifically in the case of steady-state hydrodynamic system (e.g. $J = Q$), Eq. (5) is further simplified into a exponential form:

$$p_S(T) = \overleftarrow{p}_Q(T) = \frac{J}{S} \exp\Big(- \frac{J}{S} T \Big)$$

Eq. (6) is the analytical solution of backward TTD under the RS assumption . The Eq. (3) under steady-state condition can be further simplified as:

$$\frac{\partial S_T}{\partial T} = Q(1 - \Omega_Q(S_T))$$

By combining Eq. (4) and Eq. (9), the age-ranked storage S_T can be calculated directly under the steady-state assumption:

$$S_T(T) = Q\Big(T - \int_0^T \overleftarrow{P}_Q(\tau) d\tau \Big)$$

In the idealized saturated groundwater aquifer, Eq. 6 is equivalent to the analytical solution derived by Haitjema (1995) . based on the Dupuit–Forcheimer's assumption, Haitjema (1995) derived an formula about the frequency distribution of residence time:

$$p_s(T) = \frac{1}{\overline{T}} \exp\Big(- \frac{T}{\overline{T}} \Big)$$

$$\overline{T} = \frac{nH}{J}$$

provided that nH/J is constant over the entire domain, the recharge is spatially uniform, and the aquifer is locally homogeneous, where n is the porosity -, H is the saturated aquifer thickness L, and $\overline{T}$ is the weighted mean travel time in the aquifer.

3.2 Predictive Uncertainty of TTDs

The theoretical framework of predictive uncertainty in this paper is based on Doherty (2015). As indicated in Bayes equation, the parameters of a model retain uncertainty given that they have been adjusted to best-fit values achieved during calibration. Nevertheless, the uncertainty of parameters is subject to constraints. One of the constraints resides in the fixed adjustable range of parameters, in which expert knowledge must be respected. Another constraint is exerted by the parameterization process.

While the computationally-expensive Bayesian approach offers a complete theoretical framework for predictive uncertainty evaluation, practical modeling efforts are often based on model calibration and a following analysis of error or uncertainty in post-calibration predictions (Doherty, 2015). Ideally, the best-fit parameters achieved through calibration can reduce the

predictive error to a minimum, with the minimum predictive error being the inherent uncertainty. However, the best-fit parameter is always biased from the true parameter because the essentially imperfection of model may facilitate or hamper the achievement of the minimum. Therefore, the motivation of uncertainty analysis in this study is to quantify and minimize the predictive uncertainty of travel time distributions, given that the parameters are plausible and the model can well reproduce the groundwater heads.

4 Site Description and Model Setup

3.1 Site Description

The candidate site in this paper is the Nägelstedt catchment located in central Germany (see Figure 1). With an area of approximately 850 km^2, the Nägelstedt catchment is a headwater catchment of Unstrut river. The terrain elevation of this area varies from as low as 164 m at the outlet, to as high as 516 m at the eastern mountainous area. It is a sub-catchment of Unstrut basin - one of the most intensively-used agricultural regions in Germany. About 88% of the land in this site are marked as arable land, which is significantly higher than the average level of Thuringian (Wechsung et al., 2008). The agricultural nitrogen input vary over the years and locations between 5 - 24 kg/ha on the soils in the lowlands to 2 - 30 kg/ha in the feeding area (Wechsung et al., 2008). The mean annual precipitation is about 660 mm. 18 monitoring wells distributed in this area are used to calibrate the model (Figure 1a, in which the well W0 is abandoned in this study due to the proximity to outer edge). The geological layers that the wells belong to are listed as follows: 5 in km, 4 in ku, 6 in mo, 2 in mm, and 1 in alluvium. No well is located in the geological unit mu.

The Nägelstedt catchment used as the test catchment for this study.a) The overview of Nägelstedt catchment and the locations of monitoring wells used in this study. b) Three-dimensional view highlighting the arrangement of alluvium and soil and cross-sectional view of the study area. c) Three-dimensional view highlighting the zonation of sedimentary aquifer-aquitard system. Note that the Muschelkalk layers (mo, mm and mu) are divided into more permeable sub-units (mo1, mm1 and mu1) and less permeable sub-units (mo2, mm2, and mu2).

The dominating basin-filling sediments in the study area is the Muschelkalk (Middle Triassic). The Muschelkalk has an overall thickness of about 220 m, and has been divided into three sub-groups according to mineral composition: Upper Muschelkalk (mo), Middle Muschelkalk (mm), and Lower Muschelkalk (mu). The Upper Muschelkalk (mo) is mainly composed of limestone, marlstone and claystone, and forms fracture aquifers (Jochen et al., 2014; Kohlhepp et al., 2017). The Middle Muschelkalk (mm) deposits are composed of evaporites, including dolomit marlstone, gypsum, dolomit limestone and eroded salt layers. The Lower Muschelkalk (mu) is composed of massive limestone (McCann, 2008; Jochen et al., 2014). The Muschelkalk formation consists of limestone sediments, which may form fractured and karst aquifer. A recent study demonstrated that karstification and development of conduits are limited at the base of Upper Muschelkalk at Hainich critical zone in Nägelstedt catchment (Kohlhepp et al., 2017). In the middle of the study area, Keuper deposits including Middle Keuper (km) and Lower Keuper (ku) overlay Muschelkalk formation.

Table 1. Adjustable ranges of the hydraulic parameters.

	Hydraulic conductivity K_s [m/s]						
	soil	alluvium	km	ku	mo1	mm1	mu1
Upper limit	9.0×10^{-4}	2.0×10^{-3}	9.0×10^{-4}	8.5×10^{-5}	8.0×10^{-4}	9.1×10^{-4}	2.0×10^{-5}
Lower limit	5.0×10^{-5}	4.5×10^{-6}	1.0×10^{-6}	9.6×10^{-7}	9.0×10^{-7}	3.1×10^{-7}	2.0×10^{-8}

3.1 Numerical Model Setup

3.1.1 Boundary conditions

The steady-state model configuration is achieved using a ~~temporally-averaged recharge~~ temporally averaged recharge of the simulated daily recharges over a long period (~~1955-2005). On the upper surface of the mesh, the values of spatially-distributed recharges~~ 1955–2005). The gridded recharges estimated by mHM are interpolated and then assigned to each grid ~~nodes~~node on the upper surface of the OGS mesh using a bilinear interpolation approach. No-flow boundaries are assumed at the outer edges that are defined by catchment divides, except for the northwestern and northeastern edges, where fixed-head boundaries are applied (Wechsung et al., 2008). ~~In the model, the~~ The streams are assigned with fixed-head boundaries~~.~~, wherein the heads are equal to the long-term averaged water levels. For the steady-state system and the one-way coupled model, baseflow component generated by OGS proves to be consistent with the baseflow estimated by mHM, implying that the water budget in the subsurface system is essentially closed (Jing et al., 2018). Neumann boundaries are ~~used~~ prescribed for 7 drinking water production wells~~, whereby the pumping rates of these wells are taken from Wechsung et al. (2008)~~. However, the amount of pumping makes up only around 3% of the total amount of outflow, and it therefore has a marginal influence on the water budget (Wechsung et al., 2008).

3.1.2 Modeling procedures

The numerical experiment to explore the uncertainty of ~~TTD~~ TTDs is performed through the following workflow:

1. Eight ~~spatially-distributed recharge realizations (R1-R8)~~ spatially distributed recharge realizations are sampled from a high-resolution dataset of land surface fluxes for Germany, in which mHM is used to simulate land surface hydrological processes. The details of the dataset ~~as well as~~ and the sampling method ~~is~~ are described in the following section.

2. For each recharge realization, a series of equally probable realizations of ~~hydraulic conductivity fields (K1-K50)~~ K_s fields are generated using the null-space Monte Carlo (NSMC) method.

   The NSMC method takes advantage of the hybrid Tikhonov-TSVD method in the parameter estimation code PEST to produce ~~a Monte-Carlo~~ Monte Carlo realizations of parameters (Tonkin and Doherty, 2009; Doherty and Hunt, 2010). This approach is able to efficiently generate an ensemble of parameter fields that are conditioned to expert knowledge

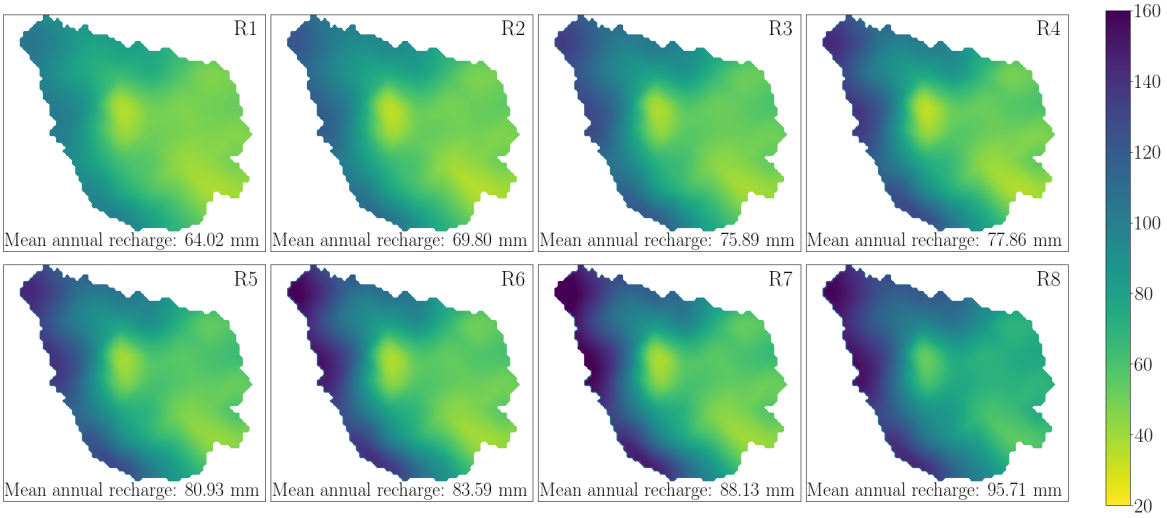

Figure 2. Recharge realizations used in this study (unit: mm). They were sampled from a high-resolution dataset of land surface fluxes for Germany (Zink et al., 2017).

and measurements. Here, the observations of groundwater levels from 18 ~~spatially-distributed~~ spatially distributed monitoring wells are used to calibrate the OGS model (the locations of monitoring wells are illustrated in Figure 1a). Before generating parameter sets, we calibrate the model to obtain the best-fit hydraulic conductivities, as well as a covariance matrix of the parameter probability distributions. On the basis of this information, ~~50 different hydraulic conductivity~~ many fully distributed K_s fields are randomly generated from a uniform distribution of hydraulic conductivity values ~~in each recharge realization using NSMC method~~ (Doherty, 2015). As ~~is~~ shown in Table 1, the range of hydraulic conductivities is predefined based on values obtained from a geological survey (Wechsung et al., 2008). As a result, a total of 400 parameter sets ~~that are all compatible with~~ conditioned on both observations and reality ~~were~~ are generated for the uncertainty analysis.

3. In each ~~recharge~~ parameter realization, a large number of particles are injected ~~on~~ through the top surface of the ~~mesh~~groundwater model. The spatial density of particles is proportional to the ~~value of~~ spatially distributed recharge rates.

   ~~In order to~~ To accurately interpret the travel time distribution, a large ~~amount~~ number of particles (e.g., ~~about~~ approximately 80 000 particles in the case study) ~~were~~ is released into the top surface of the groundwater model. The released particles serve as samples of water parcels for deriving their travel time distributions. ~~For~~ In doing so, the ~~particles are distributed on the top surface spatially weighted by the value of recharge, which means the~~ density of particles is set proportional to the recharge at the corresponding grid cell (Figure 3). Each particle tracer represents a volumetric recharge rate of around 700 m^3/year.

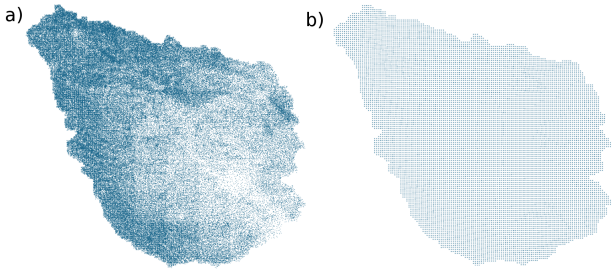

Figure 3. ~~The two~~ Two different spatial distributions of ~~emitted particles in~~ particle tracers for the ~~numerical models~~RWPT method. a) The mass-weighted distribution of particles based on the recharge estimated by mHM. This is the default spatial pattern of ~~particles~~ particle tracers in this study. b) The uniformly distributed ~~particles~~ particle tracers used in the uniform recharge scenario.

4. An ensemble of forward simulations using the RWPT method is performed ~~for each hydraulic conductivity field and recharge realization~~over all realizations of K_s fields.

   In each realization of the ensemble parameter sets, forward simulations of particle tracking are performed. In this study, we focus on the predictive uncertainty within the convection process~~, thus~~. Therefore, the molecular diffusion coefficients are universally set to 0 ~~in~~ for all ensemble simulations. The porosity of the study domain is set to 0.2 universally. Through the above procedures, the flow paths and the corresponding residence times can be fully traced in the model at random ~~time and location~~times and locations, facilitating the detailed characterization of TTDs.

In parallel to this analysis, a sensitivity analysis for the spatial variability of recharge is also performed. Two different recharge scenarios are compared for this purpose: (1) ~~R1, which is~~ the spatially distributed recharge generated by mHM ~~,~~ and (2) the ~~spatially~~ uniform recharge that ~~equals to the mean value of R1~~is equal to the spatial average of the distributed recharge. Other parameters, including the porosity and the hydraulic conductivity~~are kept~~, remain identical in these two recharge scenarios.

3.1.3 Recharge realizations

~~The recharge realizations are extracted from a~~ A high-resolution dataset of land surface ~~hydrologic fluxes over Germany (Zink et al., 2017). The dataset is derived using mesoscale Hydrologic Model (mHM)~~ fluxes and states across Germany is used for sampling recharge scenarios. This dataset was established on the basis of a daily simulation with 4 km spatial resolution using mHM for a time span of 60 years (~~1951-2010)~~ 1951–2010) (Zink et al., 2017). This dataset consists of an ensemble (100 realizations) of land surface variables, including evapotranspiration, groundwater recharge, soil moisture and discharge, with a spatial resolution of 4 km. ~~The total~~ A total of 100 realizations of land surface states are all ~~compatible with~~ conditioned on the observed daily discharge, and each of them has been derived by incorporating the uncertainty of parameterization caused by the heterogeneity of geometry, topography and geology. The modeled datasets are ~~furthermore validated against observation based~~ further validated against observation-based evapotranspiration and soil moisture data from eddy covariance stations (Heße

et al., 2017). The derived recharges show a good correspondence with the estimation from the Hydrologic Atlas of Germany (Zink et al., 2017).

Eight representative recharge realizations (R1-R8) are sampled from 100 realizations for this study to save computational time. ~~In order to~~ To enhance the representativeness of the samples, the 100 recharge realizations are sorted in an ascending order by their spatial averages. The selected recharge realizations are uniformly sampled from the sorted recharge realizations. In doing so, the maximum ~~recharge and the minimum recharge are included into~~ and minimum recharges are included in the samples such that the whole range of recharge realizations is fully covered.

3.1.4 3-D stratigraphic model

3.1.5 ~~3-D stratigraphic mesh~~

A 3-D stratigraphic mesh is established on the basis of hydrogeological characterizations elaborated in Section 2 (~~see~~ Figure 1). This mesh is based on data obtained from Thuringian State office for the Environment and Geology (TLUG) and its generation is described in **?**. The structured mesh is composed of 310 599 nodes (132 rows, 140 columns, and 82 vertical layers). The 3-D cell size of 250 m, 250 m, and 10 m in the *x*-, *y*- and *z*~~directions are~~ -directions is used in this study. Based on the German stratigraphy (Menning, 2002), the Middle Muschelkalk, the Upper Muschelkalk, the Lower Keuper, and the Middle Keuper outcrop in the Nägelstedt catchment. Accordingly, a stratigraphic aquifer system with 10 geological units is set up. The uppermost 10 m of the mesh has been separated as a soil layer, while ~~a~~ an alluvium layer consisting ~~high permeable sandy gravels~~ of high-permeability sandy gravel is set at the nodes beneath and near streams (Figure 1). Each of the Muschelkalk ~~subunits~~ layers is further divided into two categories: the more permeable parts (mo1, mm1, and mu1) ~~,~~and the less permeable parts (mo2, mm2, and mu2) (see Figure 1). For each of the Muschelkalk units, the permeability of the less permeable part is tied to the corresponding more permeable part with a factor of 0.1. The equivalent porous medium approach is applied to characterize the karst aquifer of the Upper Muschelkalk (mo). ~~We translate~~ This approach translates the parameters describing highly heterogeneous hydraulic properties at the point scale to the equivalent homogeneous medium at the regional scale to avoid adding redundant parameters ~~. Moreover, an appropriate number of parameters can effectively~~ and therefore avoid overfitting.

3.1.5 Parameter uncertainty

~~Composite parameter sensitivities of the head simulations to the parameters. (l)2-8soil alluvium km ku mo mm mu R1 6.01 1.89 0.58 1.50 9.47 20.45 0.23 R2 6.15 1.93 0.49 1.52 9.62 20.61 0.35 R3 4.05 1.78 1.38 1.91 7.20 25.84 0.82 R4 5.97 1.91 0.39 1.56 9.41 20.99 0.19 R5 7.31 1.86 0.31 1.34 10.09 19.40 0.015 R6 5.87 1.90 0.39 1.67 9.50 21.03 0.23 R7 7.77 1.93 0.57 1.93 10.53 19.07 0.41 R8 5.03 1.87 0.48 1.81 9.07 22.64 0.11 mean 6.77 1.88 0.57 1.66 9.36 21.25 0.29~

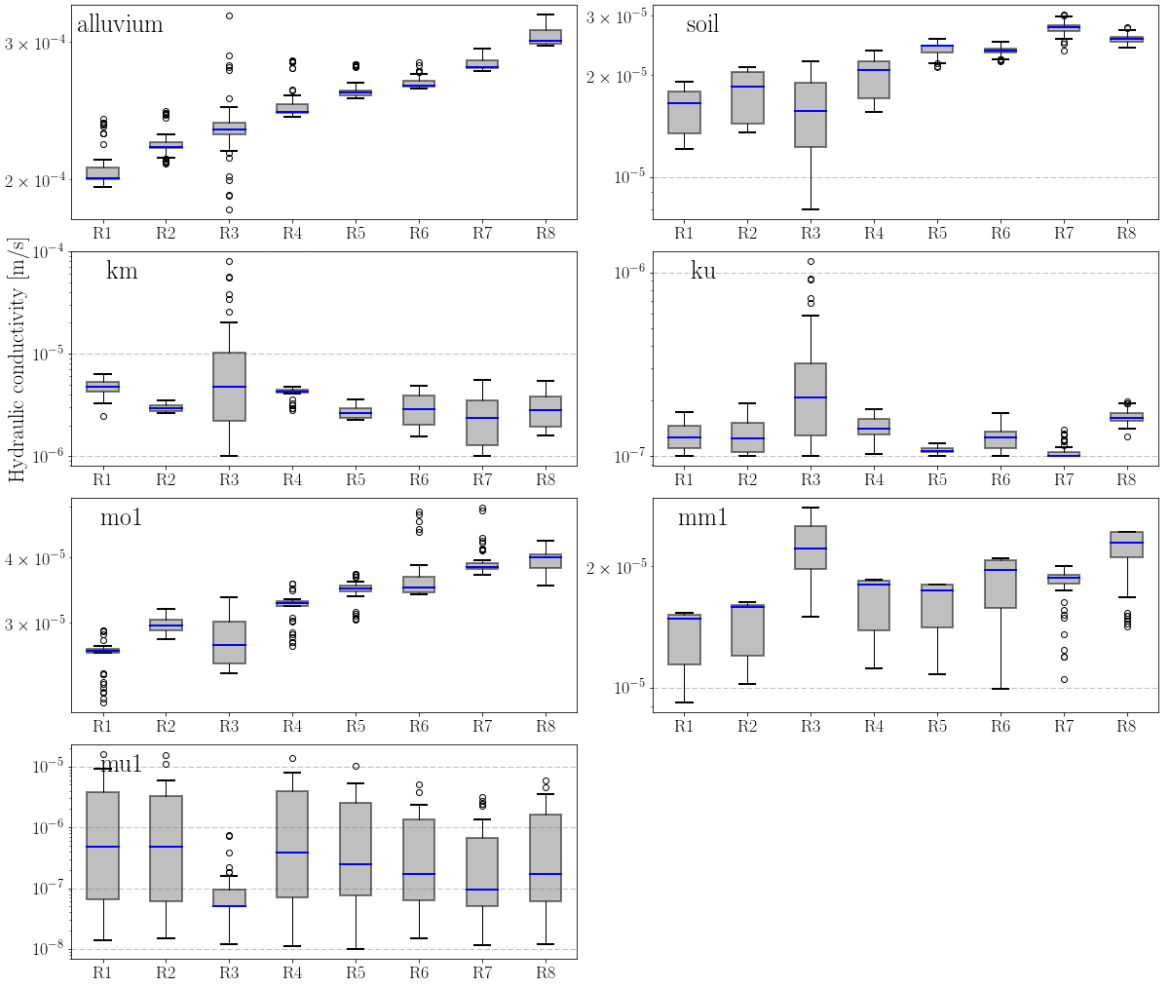

Figure 4. Box-plot of ~~stochastically-generated~~ stochastically generated hydraulic ~~conductivities of~~ conductivity (K_s) for each geological layer in 8 recharge realizations. Note that the parameters mo2, mm2, and mu2 are not shown in this figure ~~,~~ because the ~~less-permeable~~ less permeable subunits of the Muschelkalk (mo2, mm2, and mu2) are tied with the respective ~~more-permeable~~ more permeable subunits (mo1, mm1, and mu1) with a factor of 0.1.

4 ~~Results~~

3.1 ~~Parameter uncertainty~~

Multiple ~~realizations of hydraulic conductivity fields are~~ calibration-constrained K_s fields were stochastically generated for each recharge realization. ~~This parameter generation process follows a two-step procedure. The first step consists of calibrating the model in each recharge realization to get the best-fit parameters and parameter sensitivity matrices. The PEST algorithm calculates the sensitivity with respect to each parameter of all observations (with the latter weighted as per user-assigned~~

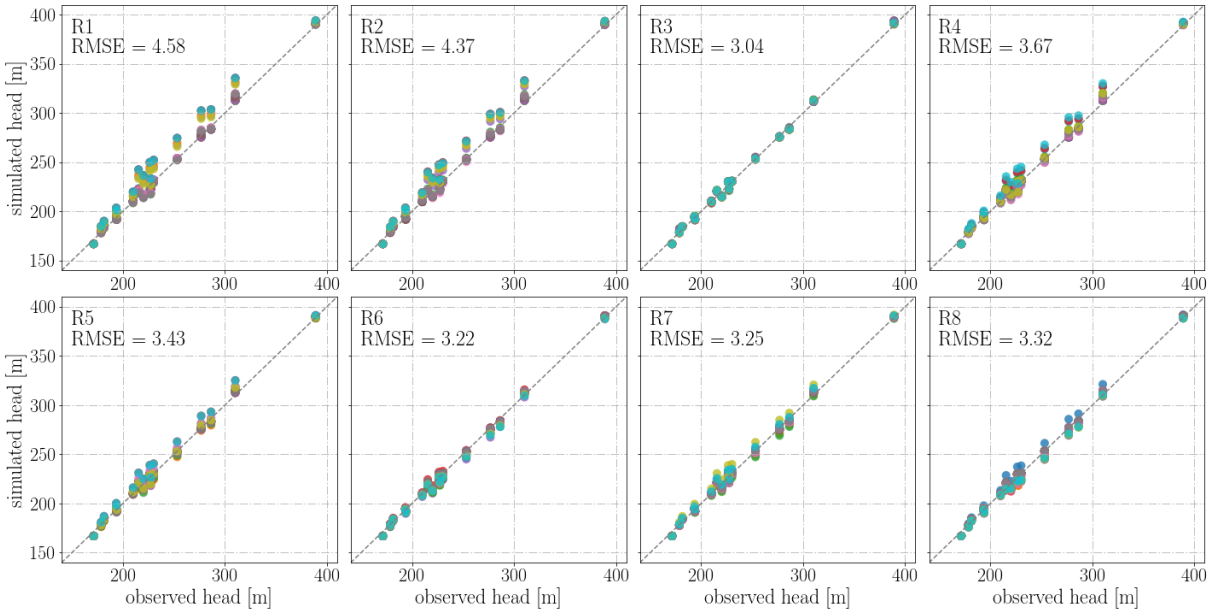

Figure 5. Observed and simulated groundwater heads for each parameter and recharge realization. The results of 400 realizations (R1K1 - R8K50) are categorized by recharge realization and shown in different panels.

~~weights), namely the "composite sensitivity" (Doherty, 2015). The composite sensitivity of parameter i is defined as $csp_i = \frac{[\mathbf{J}^t\mathbf{Q}\mathbf{J}]_{ii}^{1/2}}{n}$, where **J** denotes the Jacobian matrix that includes the sensitivities of all predictions to all model parameters, **Q** is the weight matrix, n is the number of observations with non-zero weights. In this study, all weights assigned to observations are equally set to 1. Table B1 displays the composite parameter sensitivities in each recharge realization. The mean composite parameter sensitivities of calibrations in all recharge realizations are also included in this table. The groundwater level predictions are highly sensitive to Middle Muschelkalk (mm), and insensitive to Lower Muschelkalk (mu). The sensitivity of mu, however, varies widely between different recharge realizations from the highest one in R3 (0.82) to the lowest one in R5 (0.015).~~

~~The second step is to generate multiple Monte Carlo realizations of parameters for each recharge realization that are all compatible with observations.~~ Figure 4 shows the box-plot of generated hydraulic conductivities in all realizations categorized by geological unit. The hydraulic conductivity of the Lower Muschelkalk (mu) has the highest uncertainty (10^{-8} - 10^{-5} m/s) because the ~~observations are insensitive to mu~~ conductivity of mu is insensitive to groundwater head observations, given that it is the deepest geological layer and that no monitoring well is located in this layer (Table B1). The other parameters fluctuate moderately and are constrained within one order of magnitude in most of the recharge realizations~~, despite some exceptions in R3. An ascending trend of hydraulic conductivity of~~ . Hydraulic conductivities of several permeable layers (mo, mm, alluvium, and soil) increase from R1 to R8~~can be observed in Figure 4.~~ , which is not surprising because the hydraulic conductivity increases with increasing recharge and constant groundwater head. Moreover, the hydraulic conductivities of the above layers are roughly linearly correlated to the corresponding recharge in each recharge realization. Figure 5 shows the ~~residuals of~~

simulated and observed groundwater heads ~~in~~ for all 400 realizations. All of the 400 realizations are well constrained to observations, with the ~~Root Mean Square Error~~ root mean square error (RMSE) of groundwater level residuals being lower than 4.6 m in all of the considered recharge realizations.

3.1 Theory of analytical StorAge Selection function

The travel time is defined as the time spent by a moving element (either a water particle or a solute) in a control volume of a hydrologic system. In principle, the control volume can be defined at arbitrary spatial scales (i.e., from the molecular scale to the regional scale). Considering a hydrologic system in which the input flux (J) and the output fluxes ($Q_1, Q_2, ..., Q_n$) are known, each parcel of water within the system is tagged using its current age τ. The age-ranked storage $S_T = S_T(T,t)$ is defined as the mass of water in the system with age $\tau < T$. The backward form of the master equation (ME) for TTD in a control volume can be expressed as follows (Botter et al., 2011; Van Der Velde et al., 2012; Harman, 2015):

$$\frac{\partial S_T}{\partial t} = J(t) - \sum_{j=1}^{n} Q_j(t) \overleftarrow{P}_{Q_j}(T,t) - \frac{\partial S_T}{\partial T} \tag{3}$$

with boundary condition $S_T(0,t) = 0$, where T is the residence time of the oldest water parcel in storage S_T; t is the chronological time; $\overleftarrow{P}_{Q_j}(T,t)$ is the cdf of the backward travel time distribution of output flux Q_j; $J(t)$ is the input flux at time t; and $Q_j(t)$ is the output flux at time t. Specifically in this study, J is the groundwater recharge, and Q is composed of two components: the stream baseflow and the abstraction at production wells.

The StorAge Selection (SAS) function describes the fraction of water parcels leaving the control volume at time t, which is selected from the age-ranked storage S_T. Following the above definition, the SAS function can be linked with the backward travel time distribution $\overleftarrow{P}_Q(T,t)$ (Harman, 2015):

$$\Omega_Q(S_T,t) = \overleftarrow{P}_Q(T,t) \tag{4}$$

for $S_T = S_T(T,t)$. Ω_Q is the cumulative form of the StorAge Selection (SAS) function.

Three instances of SAS functions using gamma distribution are shown in Figure 9a. In case the age distribution of each outflow is uniformly selected from all water storages with various ages, the outflux TTDs turn into a random sample of the storage residence time distribution (RTD). The random sampling (RS) is equivalent to the uniform SAS function (Figure 9a). Many past studies have also considered the random sampling as a proper description of the sampling behavior for heterogeneous catchments (Cartwright and Morgenstern, 2015; Benettin et al., 2015; Heße et al., 2017). Eq. (4) in this case has the analytical solution (Harman, 2015; Danesh-Yazdi et al., 2018):

$$p_S(T,t) = \overleftarrow{p}_Q(T,t) = \frac{J(t-T)}{S(t)} \exp\Big[- \int_{t-T}^{t} \frac{Q(\tau)}{S(\tau)} d\tau \Big] \tag{5}$$

where $p_S(T,t)$ is the pdf of the residence time distribution and $S(t)$ is the storage at time t. Specifically, in the case of a steady-state hydrodynamic system, Eq. (5) is further simplified into an exponential form:

$$\overleftarrow{p}_Q(T) = \frac{J}{S}\exp\Big(-\frac{J}{S}T\Big) \tag{6}$$

Eq. (6) is the analytical solution of backward TTD under the RS assumption.

In the idealized saturated groundwater aquifer, Eq. 6 is equivalent to the analytical solution derived by Haitjema (1995). Based on Dupuit–Forcheimer's assumption, Haitjema (1995) derived a formula about the frequency distribution of residence time:

$$p_s(T) = \frac{1}{\overline{T}}\exp\Big(-\frac{T}{\overline{T}}\Big) \tag{7}$$

$$\overline{T} = \frac{nH}{J} \tag{8}$$

provided that nH/J is constant over the entire domain, the recharge is spatially uniform, and the aquifer is locally homogeneous, where n is the porosity, H is the saturated aquifer thickness, and $\overline{T}$ is the weighted mean travel time in the aquifer.

3.2 Linking the SAS functions to the physically based numerical model

Danesh-Yazdi et al. (2018) developed an approach to link the analytical SAS functions to the fully distributed numerical model. Although differing in numerical model and particle tracking scheme, we apply the same approach to link the SAS functions and the numerical model in this study. Eq. (3) under the steady-state condition can be further simplified as:

$$\frac{\partial S_T}{\partial T} = Q(1-\Omega_Q(S_T)). \tag{9}$$

By combining Eq. (4) and Eq. (9), the age-ranked storage S_T can be calculated directly under the steady-state assumption:

$$S_T(T) = Q\Big(T - \int_0^T \overleftarrow{P}_Q(\tau)d\tau\Big). \tag{10}$$

Combining Eq. 10 with Eq. 4, the SAS function can be directly computed from the simulated backward TTD using numerical model.

3.3 Predictive uncertainty of TTDs

The theoretical framework of predictive uncertainty in this paper is based on Doherty (2015). As indicated in Bayes' theorem, the parameters of a model retain uncertainty, given that they have been adjusted to best-fit values achieved during calibration. Nevertheless, the uncertainty of parameters is subject to constraints. One of the constraints resides in the fixed adjustable range of parameters, in which expert knowledge must be respected. Another constraint is exerted by the parameterization process.

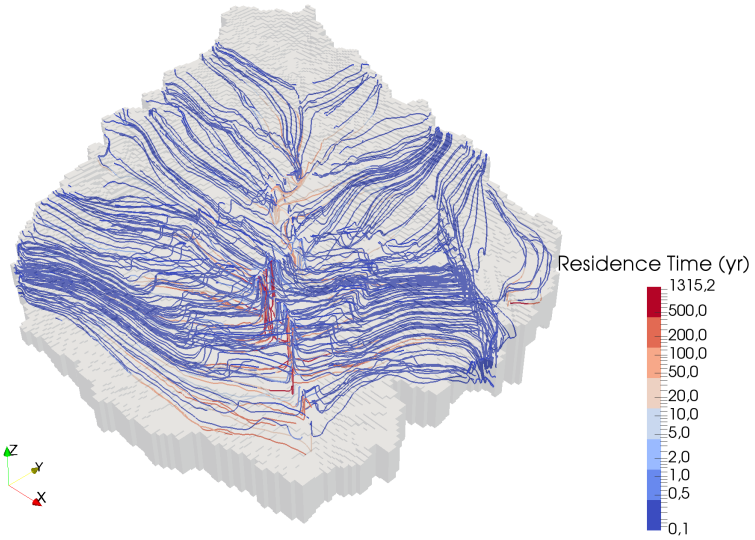

Figure 6. ~~Three-dimensional~~ 3-D view of flow pathlines of some particles in realization R5K1. Note that only a limited number of particle pathlines are displayed here.

While the computationally expensive Bayesian approach offers a complete theoretical framework for predictive uncertainty evaluation, practical modeling efforts are often based on model calibration and a subsequent analysis of error or uncertainty in post-calibration predictions (Doherty, 2015). Ideally, the best-fit parameters achieved through calibration can reduce the predictive error to a minimum, with the minimum predictive error being the inherent uncertainty. However, the best-fit parameter is always biased from the true parameter because the essential imperfection of the model may facilitate or hamper the achievement of the minimum. Therefore, the motivation of uncertainty analysis in this study is to quantify and minimize the predictive uncertainty of travel time distributions, given that the parameters are plausible and that the model can reproduce well the groundwater heads.

4 Results

For the sake of clarity, we number the recharge realizations from R1 (with the lowest recharge rate) to R8 (with the highest recharge rate). For each recharge realization, 50 K_s fields are numbered from K1 to K50. Accordingly, R1K1 represents the K_s field K1 in the recharge realization R1.

4.1 Uncertainty of TTD predictions

Flow paths of ~~some particles~~ particle tracers in a random parameter realization (R5K1) are displayed in Figure 6, serving as a visual reference for the regional groundwater flow pattern and the residence time distributions. ~~The block effect of the deep low-permeable geological layers can be observed in Figure 6, as a large portion~~ In this realization, the deep low-permeability

geological layers (e.g., mm2 and mu2) act as low-permeability aquitards. Therefore, the majority of streamlines do not enter ~~the low-permeable aquifer layers (mo2, mm2, and mu2~~these geological layers (Figure 6).

Figure 7 displays ~~the TTDs of 400 hydraulic conductivity fields within~~ both the simulated TTDs using 50 K_s fields for 8 recharge realizations (orange solid lines) ~~, as well as~~ and the reference TTDs, represented by fitted blue dash-dot curves using the exponential model ~~.~~ (Eq. 6). The ensemble average (~~$\overline{MTT}$) and~~ μ) and the coefficient of variation (~~CV~~c_v) of MTTs for each recharge realization are also calculated and shown in Figure 7. Note that if the number of parameter realizations is ~~large enough~~sufficiently large, the ensemble average of MTTs (~~$\overline{MTT}$~~μ) will converge to the simulation result using the best-fit parameters achieved through model calibration (Doherty, 2015). Noticeable variability of TTDs can be observed with respect to different recharge realizations. Generally, the ~~$\overline{MTT}$~~ μ values show a decreasing trend from 166.5 ~~yr~~ yrs in recharge realization R1 to 110.9 ~~yr~~ yrs in recharge realization R8, with only two exceptions (R3 and R6), which is not surprising ~~according to~~ based on the inversely linear dependency between recharge J and μ, indicated by Eq. 8. In each ~~of the recharge realizations~~recharge realization, the different realizations of ~~hydraulic conductivity~~ K_s fields manipulate the mean travel time. ~~The coefficient of variation (CV)~~ c_v varies from 7.81% ~~in~~ (for R5~~,~~) to 15.56% ~~in~~ (for R3), indicating a modest degree of uncertainty propagated from ~~hydraulic conductivity~~ K_s estimation to TTD prediction.

The exponential model under the RS assumption is fitted to the ensemble averaged TTD of numerical solutions (see black lines in Figure 7) using Eq. 6. As shown in Figure 7, the shape of numerically simulated TTDs significantly deviate from the exponential distribution under the RS assumption, indicating a ~~non-uniform~~ nonuniform sampling behavior of different water ages. The TTDs of numerical simulations are more right-skewed than the analytical TTDs under the RS assumption. This phenomenon reveals that the catchment TTD cannot be replicated by the single random sampling store.

Based on Eq. 8, we can approximate the "effective volume" of water involved in the transport process in the aquifer. ~~Using the ensemble averaged MTTs for each recharge realization, the~~ The effective volume of groundwater storage related to the transport process is calculated and shown in Table 2. The effective ~~volume~~ volumes of storage (S_{eff}) estimated by the numerical solutions ~~ranges~~ range from 9.8 m to 12.0 m, whereas the S_{eff} estimated by the analytical solution ~~ranges~~ range from 6.8 m to 7.5 m. The groundwater storage that contributes to the streamflow is significantly smaller than the total groundwater storage (48.3 m). This difference is because most of the released particles only exist in the upper permeable layers rather than spread evenly over the whole aquifer/aquitard system. ~~The less permeable layers (mo2, mm2, mu2 and mu1) actually act as aquitards.~~ We are aware that this is only a first-order approximation ~~as~~ because the analytical solution is only rigorously valid for the idealized homogeneous aquifer system (Haitjema, 1995).

Moreover, we assess the propagation to the MTT predictions from input and parameter uncertainty yielded by the 8 recharge realizations and the Monte Carlo realizations of hydraulic conductivities. Figure 8a depicts the distribution of MTTs of the ensemble simulations. The MTTs of the 400 realizations range from 87 ~~yr~~ to 212 ~~yr. Meanwhile~~yrs. At the same time, the ensemble average of MTTs over all realizations of recharges and ~~hydraulic conductivity fields (~~$\overline{MTT}$~~)~~ K_s fields is 135.1 ~~yr~~yrs, and the coefficient of variation is 18.93%. Figure 8b depicts the relationship between the ensemble average of MTTs and the spatially averaged recharge rates. We observe a roughly inversely proportional relationship between the ensemble average of

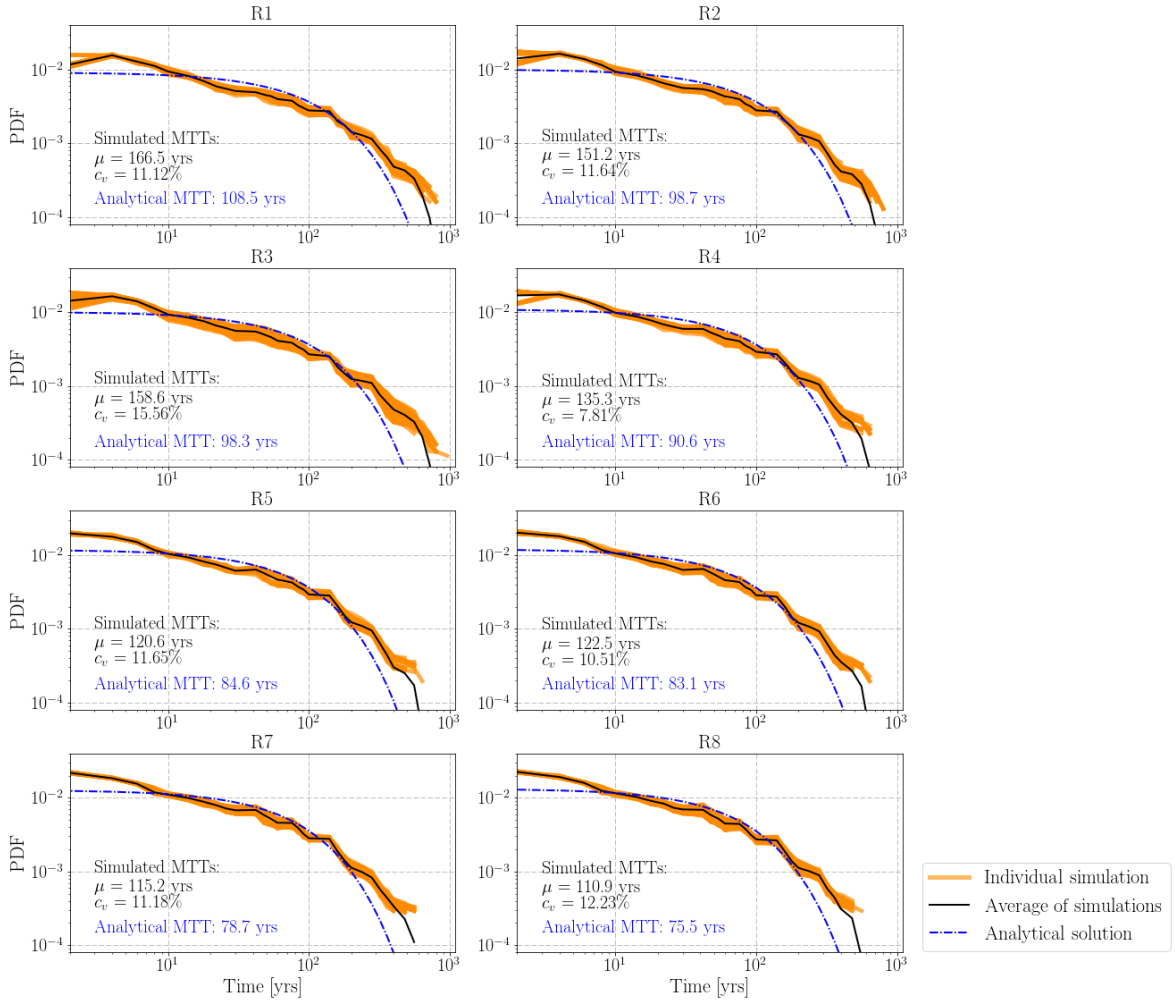

Figure 7. Travel time distributions of ensemble simulations and analytical solutions categorized by recharge realization. The orange lines show the simulated TTDs of all realizations of ~~hydraulic conductivity~~ K_s fields for each recharge realization. The black lines denote the ensemble averaged TTDs of each recharge realization. The blue dash-dot line is the fitted analytical TTD under the ~~random sampling (~~RS ~~)~~ assumption. ~~$\overline{\text{MTT}}$ denotes~~ The analytical MTT, the ~~ensemble averaged~~ mean ~~travel time~~ (~~MTT~~μ). ~~CV_{MTT} denotes~~ , and the coefficient of ~~variation~~ variance (c_v) of the ~~ensemble~~ simulated MTTs are also shown in this figure.

~~MTT~~ MTTs and the spatially averaged recharge rates (Figure 8b). The standard deviations (σ) of simulated MTTs range from 12.9 yrs (for R6) to 24.7 yrs (for R3).

4.2 Uncertainty of young/old water preference

Figure 9a provides an intuitive illustration of the relationship between the cumulative rank SAS functions and the ~~preference~~ ~~for~~ systematic preference for discharging water with different ages. Figure 9b shows the cumulative rank SAS functions of all

Table 2. Effective groundwater storages related to the transport process for each recharge realization.

	Effective volume of storage [m]								
	R1	R2	R3	R4	R5	R6	R7	R8	mean
$S_{eff\text{-}num}$ (numerical)	10.7	10.6	12.0	10.5	9.8	10.2	10.2	10.6	10.6
$S_{eff\text{-}ana}$ (analytical)	6.9	6.9	7.5	7.1	6.8	6.9	6.9	7.2	7.0

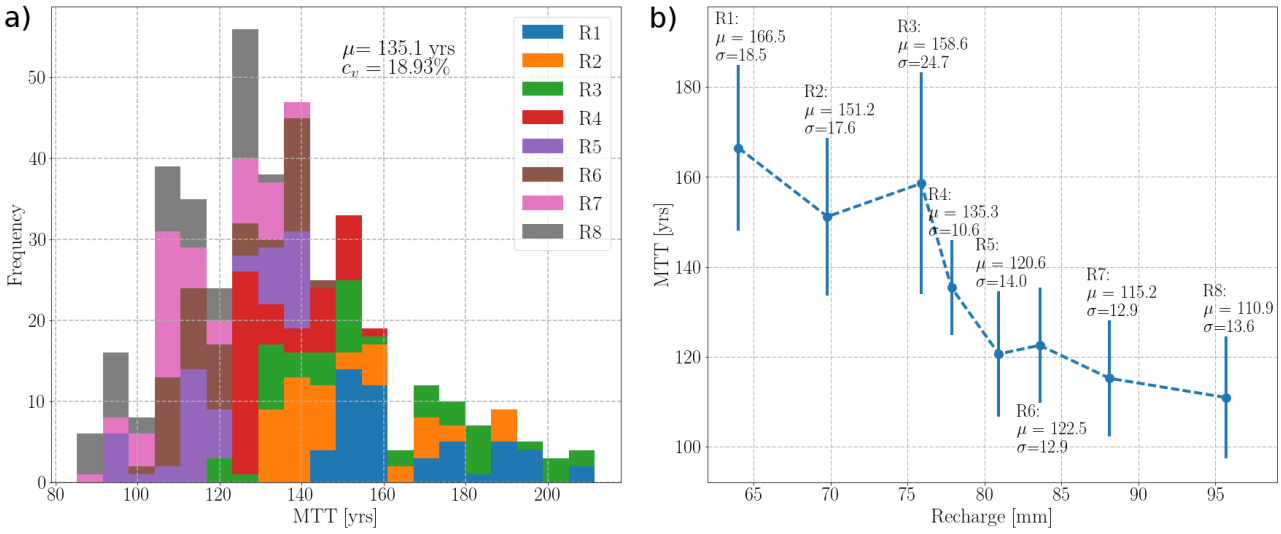

Figure 8. Uncertainty quantification: Monte Carlo simulations of MTT predictions categorized by recharge realization. Panel a) shows ~~the~~ a histogram of MTT predictions. Panel b) shows the relationship between the ensemble averaged MTT (μ) and the ~~reciprocal of~~ spatially averaged recharge~~(1/J)~~. ~~Error~~ The error bars represent the standard deviation of MTTs (σ) for each recharge realization.

ensemble simulations (obtained from 400 realizations of ~~hydraulic conductivity~~ K_s fields in 8 recharge ~~realizations~~scenarios). The figure is categorized into 8 groups by different colors and line styles, each representing a recharge realization. ~~Figure 9c depicts the ensemble averaged SAS functions for each recharge realization. The differences among SAS functions of different recharge realizations are moderate, indicating that there appears to be no systematic relationship between recharge and SAS function. Generallyspeaking~~Generally, the system has a weak preference to select younger water as discharge, despite different recharge realizations and ~~hydraulic conductivity realizations. It is also apparent that there is a certain degree of uncertainty of~~ K_s realizations. Nevertheless, a moderate variation in SAS functions for ~~different realizations of recharge and hydraulic conductivity. This uncertainty is a direct result of the propagation from recharge input and parameter estimation. Despite this degree of uncertainty, all SAS functions show a moderate tendency for sampling younger groundwater from the groundwater storage. This reveals that the different realizations of hydraulic conductivities do not change the overall sampling preference for young waterof the stratigraphic aquifer system.~~ the ensemble simulations can be observed. The variation of interest is

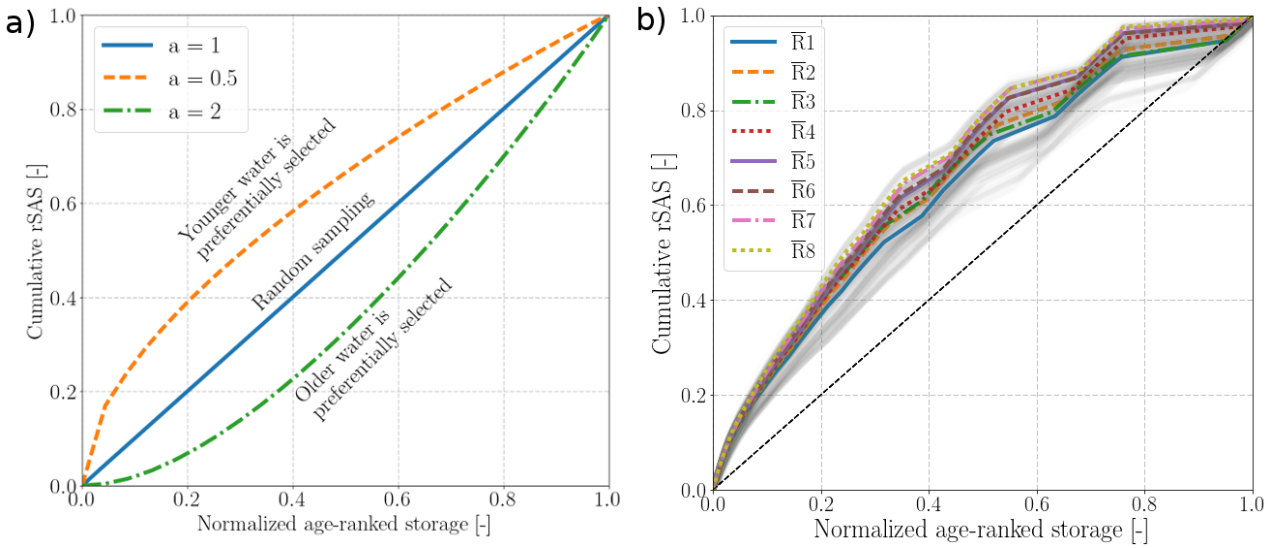

Figure 9. Cumulative rank SAS functions as a function of normalized age-ranked storage. (a) Schematic of cumulative rank SAS functions parameterized by gamma distribution with the shape parameter a = 0.5, 1, and 2. (b) Cumulative rank SAS functions of the ensemble simulations (light grey lines) and the ensemble average for each recharge realization.

introduced by the spatial distribution and velocity of flow pathlines that are controlled by different K_s fields. For example, a more permeable shallow aquifer layer will activate more flow pathways in this layer and thus will introduce a stronger preference for young water. Particularly, the significant variation in hydraulic conductivities in the deepest geological layer (e.g., Lower Muschelkalk) has a pronounced impact on the discharge of old water. With a thickness of up to 100 m, the hydraulic conductivity of this layer controls how many water parcels can enter into this layer and how deep the flow paths can develop. This effect can be evidenced by a large difference in the SAS functions related to old ages and a relatively smaller difference in those related to young ages in Figure 9b.

4.3 Sensitivity to the spatial pattern of recharge

Figure 10a depicts the sensitivity of simulated TTDs and ~~mean travel times (MTTs)~~ MTTs to the spatial distribution of recharge~~.~~ , while Figure 10b shows the sensitivity of the cumulative SAS function to the spatial pattern of recharge. The reference simulation is set up using the ~~spatially-uniform recharge that equals~~ spatially uniform recharge that is equal to the spatial average of the spatially distributed recharge, while all of the other parameters in these two simulations are ~~kept~~ held identical. The different spatial distributions of recharge have a clear effect on the shape of TTDs. It ~~seems~~ appears that the most evident difference between the TTDs of the two recharge scenarios occurs at the early period. Additionally, the simulated MTT using the uniform recharge appears to be smaller than that using the spatially distributed recharge. Figure 10b indicates that the simulation using uniform recharge has a consistently stronger preference for sampling young water than the simulation using spatially distributed recharge. Nevertheless, both ~~the two~~ scenarios show a general preference for young water. ~~This~~

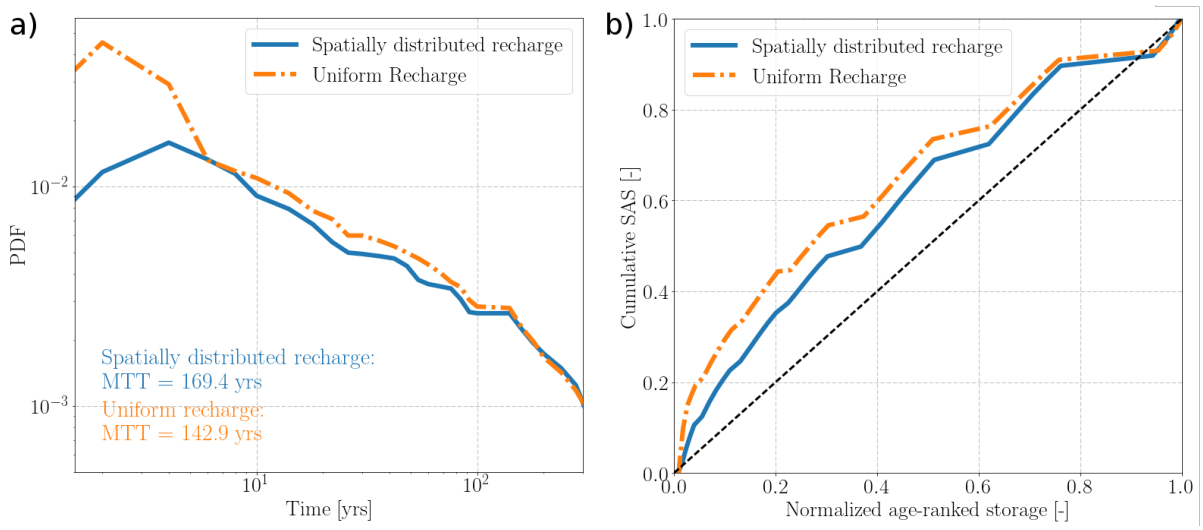

Figure 10. Sensitivity of a) TTDs and ~~mean travel time (MTT),~~ MTTs and b) SAS ~~function~~ functions to the spatial pattern of recharge.

~~phenomenon further underlines the dependency of SAS functions on the spatial pattern of recharge and the importance of reliable characterization of spatial distribution of recharge for the TTD estimation.~~

5 ~~Discussions~~

~~The purpose of this study is to quantify the uncertainty of groundwater TTDs~~ The difference in TTDs and SAS functions is not induced by the variability in internal hydraulic properties, since the two simulations share the same K_s field. Rather, it is mainly induced by the ~~uncertainty of external forcings (e. g. recharge) and internal physical characteristics (e.g. hydraulic conductivity) in a regional-scale agricultural catchment. For this purpose, a systematic uncertainty analysis considering multiple scenarios of recharge and multiple compatible estimations of hydraulic conductivity fields is performed using numerical models. Parallel to that, an analytical model has also been established, providing a reference for the effective volume of storage and the sampling behavior of the aquifer system. The SAS function is used to interpret the modeling results.~~ different flow paths of particle tracers in the two recharge scenarios. The spatially distributed recharge simulated by mHM reveals that the upstream mountainous area has higher recharge rates than those in the lowland plain. By contrast, the uniform recharge scenario neglects this spatial nonuniformity. This difference results in the following: (a) under a uniform recharge scenario, more particle tracers enter the system from locations near the streams at lowland plains (Figure 3b), indicating that more particle tracers are transported in the local flow system rather than in the regional flow system (Toth, 1963); and (b) higher recharge rates at lowland plains accelerate the particles' movement in this area and shorten their travel times. As such, local particle flow paths within the shallow aquifer layer at lowland plains are activated, leading to a stronger preference for sampling local flow paths in shallow aquifer layers and therefore a stronger preference for young ages. Our findings are in line with the observations by Kaandorp et al. (2018),

wherein the authors found a relatively higher preference for the selection of older water in the upstream area than that in the downstream area of the Springendalse Beek catchment.

5 Discussion

5.1 Uncertainty of external forcing, internal property, and TTD predictions

In the idealized aquifers ~~, TTD is controlled by recharge and independent of hydraulic conductivity (Haitjema, 1995). Specifically, given that the~~ where groundwater flow is Dupuit-Forchheimer type, the recharge is uniform, and the aquifer is ~~locally-homogeneous, the corresponding TTD has been demonstrated to be independent of hydraulic conductivity. Rather, the~~ locally homogeneous. TTD is controlled by recharge, saturated aquifer thickness~~and porosity.~~

, and porosity, and it is independent of hydraulic conductivity (Haitjema, 1995). In a real-world catchment with complex geometry and topography, stratigraphic ~~aquifer, and non-uniform~~ aquifers, and nonuniform recharge, our numerical exploration demonstrates that the groundwater TTD is dependent on both ~~the recharge and the hydraulic conductivity. Provided that the model calibration problem is well-posed (i.e., all estimable parameters can be well constrained on the basis of observations), the uncertainty of recharge will have an strong influence on the predictions of TTDs. Therefore, ignoring the recharge uncertainty will cause incorrect simulation results of TTDs and underestimate the uncertainty bounds (Ajami et al., 2007; Healy and Scanlon, 2010)~~ recharge and the hydrostratigraphic K_s field.

~~A well-posed problem indicates that all parameters are estimable based on the calibration dataset. The well-poseness of the calibration has strong impact on the predictive uncertainty of travel times. In the case study of this paper, the parameter-generation problem is not rigorously well-posed, because the sensitivities of mu and km to the calibration dataset are very low compared to other parameters (Table B1). Except for these two parameters, other parameters are well constrained to the observed groundwater heads~~ The mechanisms behind the effects of the recharge rate and the K_s field are different. Given that the spatial pattern of recharge remains the same, a higher recharge rate intensifies the fluxes for the complete range of flow pathways. This process forces water downward in the recharge zones and upward to the discharge area. As such, flow rates through all flow pathways are increased equally, and the spatial distribution of flow pathways is not changed. The corresponding SAS function is also not changed (Botter et al., 2010; Cartwright and Morgenstern, 2015; Kaandorp et al., 2018). In contrast, a different K_s field activates flow pathways in more permeable layers, deactivates flow pathways in less permeable layers, changes the spatial distribution of flow pathways, and therefore changes the shape of the SAS function (Harman et al.; Kim et al.).

We also underline the value of observational data in reducing predictive uncertainty in simulated TTDs. In this study, the majority of model parameters can be adequately conditioned by spatially distributed groundwater head observations (Figure 4). Provided that most of the hydraulic conductivities are constrained to the model-to-measurement misfit and reality, the TTD predictions ~~are also~~ can also be effectively bounded. This ~~can be~~ is evidenced by Figure 7, from which moderate values of ~~CVs~~ c_v ranging from 7.81% to 15.56% in different recharge realizations ~~with a mean value of 11.47%~~ can be observed. The ensemble averaged MTTs for ~~each recharge scenario effectively eliminate the variability caused by parameter uncertainty, and provides a benchmark for evaluating the influence of recharge to the MTT prediction. The ensemble averaged MTTs of~~

different recharge realizations also have a high ~~variability~~ c_v (15.70%), implying that the TTD prediction ~~seems~~ appears highly sensitive to recharge. Our findings are in line with Danesh-Yazdi et al. (2018), in which the interplay between recharge and subsurface heterogeneity was investigated and a strong dependency of TTDs on the recharge was ~~founded~~observed.

~~The assumption of spatially uniform input forcing has been widely applied in regional-scale hydrologic models (Zghibi et al., 2015; Yang . Our study indicates that the spatial variability of recharge significantly alters the shape and breadth of TTD predictions (Figure 10). This phenomenon is critically important for large-scale applications with typically a large spatial variability of topography, land cover, geology, and rainfall. As a result, the groundwater recharge show significant spatial variability (Figure 2). Our study indicates that both the shape and breadth of TTDs are sensitive to the spatial pattern of recharge. Therefore, the reasonable characterization of spatial pattern of recharge is crucial for the reliable TTD prediction. Unfortunately, it is quite challenging to confidently determine the groundwater recharge at the regional scale under today's technique due to the lack of reliable measurements (Healy and Scanlon, 2010; Cheng et al., 2017; Zink et al., 2017). Appropriate techniques should be chosen to estimate groundwater recharge according to the study goals and the spatial and temporal scale. Besides, a combination of multiple techniques is suggested to reduce the uncertainty of recharge estimation (Healy and Scanlon, 2010).~~

5.2 Analytical model and SAS functions

5.3 ~~Analytical model and SAS function~~

The analytical solution of TTD, assuming a random sampling of water, cannot properly replicate the TTD of numerical simulation in the study domain. In the stratigraphic aquifer with complex topography and ~~spatially distributed~~ diffuse recharge, the analytical solution using Eq. 6 ~~may underestimate~~ underestimates the MTT. Note that this conclusion holds when the simulated TTD has a relatively larger long-tail behavior than the exponential distribution. Such observations have also been reported for other real-world aquifers (Basu et al., 2012; Eberts et al., 2012; Kaandorp et al., 2018). This finding can be seen as an extension of Basu et al. (2012), ~~whereby they found a certainty degree of~~ in which the authors found a moderate discrepancy between the analytical solution using Eq. 6 and the numerical solution in a small catchment. ~~Unlike the simple model configuration used in Basu et al. (2012), we use a more realistic spatially distributed recharge and stratigraphic aquifer for the numerical model.~~

It is obvious that ~~lumped parameter model~~ the analytical solution under the RS assumption cannot explicitly include the impact of the distributed hydraulic properties of stratigraphic ~~aquifer~~ aquifers and the spatially ~~non-uniform~~ nonuniform recharge. The above limitations of analytical models may introduce a significant ~~aggregation~~ predictive error for the TTD predictions, as shown in Figure 7. Moreover, ~~a new method for estimating the effective storage (S_{eff}) is proposed to characterize the effective~~ the effective volume of storage related to the transport process ~~.~~ S_{eff} is calculated utilizing the numerical solutions of distributed models~~, therefore successfully avoiding the aggregation error through the explicit characterization of the hydraulic and topographic variability. Although being~~ . As a first-order approximation of effective ~~volume of storage , it provides a simple metric for quantifying effective storage in complex real-world applications~~storage volume, it suggests that only a small fraction of total groundwater volume is involved in the active hydrologic cycle.

The SAS function provides a good interpretability ~~of numerical solutions~~ for simulation results using the fully distributed model in terms of characterizing the preference for releasing water of different ages. We find that the SAS ~~function is visually insensitive to hydraulic conductivity fields and recharges in~~ functions are weakly dependent on the K_s fields in the stratigraphic aquifer system~~.~~, but the overall preference for young water does not change. This weak dependency can be explained by the fact that different realizations of K_s fields modify the spatial distribution of particle flow paths. The overall tendency for young groundwater ~~of~~ in the saturated aquifer has been ~~observed by Danesh-Yazdi et al. (2018), although the spatial organization of aquifer system is distinct from the stratigraphic aquifer used in our study~~reported by many past studies (Danesh-Yazdi et al., 2018; Kaandorp et al., 2018; Yang et al., 2018). Our study links the explicit simulations of travel times and the analytical SAS functions ~~,~~ and offers original insights into the uncertainty propagated from recharge and ~~hydraulic conductivity~~ the K_s fields to the SAS functions.

5.3 Dependency of TTDs and SAS functions on the spatial pattern of forcings

5.4 ~~Implications for applied groundwater modeling~~

The sensitivity of the TTDs and SAS functions to the spatial pattern of recharge forcings can be explained mainly by the different flow paths of particle tracers, resulting primarily from the spatially heterogeneous fields of recharge across the study catchment. For the regional groundwater system, the spatial variation of recharge determines the distribution of starting points of the flow pathlines of tracer particles. For example, more particles will be injected from recharge zones that are typically located in high-elevation regions, resulting in a higher weight of flowlines starting from high-elevation regions. The pronounced spatial variability of recharge also controls the systematic (water age) preference for particles existing from the system (to river discharge) that originated from different regions and therefore exerts a strong control on the shape of the SAS function.

~~Uncertainty limit~~ In the study catchment, an oversimplified spatially uniform recharge results in a smaller MTT and a stronger preference for discharging young water compared to those taking the spatial variability of recharge. Such observations are conditioned to site-specific features of the study catchment, which is noticed only when the following apply: (a) a site is located in a headwater catchment under a humid climate condition; (b) the recharge rate in areas close to the drainage network is generally lower than that in areas far away from the drainage network; and (c) the system is under (near) natural conditions, meaning that artificial drainage and pumping do not dominate the groundwater budget.

The assumption of spatially uniform input forcing has been widely applied in regional-scale subsurface hydrologic models (Zghibi et al., 2015; Yang et al., 2018). Our study indicates that a reasonable characterization of spatial pattern of recharge is crucial for reliable TTD prediction. Unfortunately, it is quite challenging to confidently quantify the groundwater recharge at the regional scale under today's technique due to the lack of reliable measurements (Healy and Scanlon, 2010; Cheng et al., 2017; Zink et al., 2 . Appropriate techniques should be chosen to estimate groundwater recharge according to the study goals and the spatial and temporal scales.

5.4 Implications for the applied groundwater modeling

Uncertainty limits the applicability of groundwater models. Most of the applied groundwater models are deterministic models ~~. The deterministic models~~ that use direct values of inputs and parameters instead of probabilistic distributions of them. Specifically, both the model inputs and the inversion process are deterministic, leading to a deterministic best-fit parameter ~~sets~~ set achieved during model calibration. Our study reveals limitations of the above modeling procedure ~~, and suggest~~ and suggests that the probabilistic distribution of inputs and parameters should be considered ~~in~~ for the applied modeling. The main limitation is that the single exclusive assignment of recharge is inadequate for the simulation of transport processes ~~. If the exclusive recharge estimation is biased from reality due to the insufficiency of data, the generated parameters (i.e.~~because the error can be propagated from inputs to the conditioned parameters (e.g., hydraulic conductivities) ~~will also become biased from the reality as they are strongly dependent on the accuracy of input data~~ through the calibration process (Figure 4). This accumulated error ~~of both input and parameter~~ will further lead to a seriously biased prediction of travel times. Additionally, the modeling workflow used in this study is computationally more efficient than the Bayesian approach and is suitable for complex real-world applications.

The degree of predictive uncertainty is highly dependent on the parameterization scheme. Some ~~complex aquifer systems can be described by highly-parameterized models . These models~~ highly parameterized models are potentially ill-posed due to the paucity of data ~~,~~ and therefore cannot be constrained by the available calibration dataset. In this case, the predictive uncertainty of TTDs is potentially ~~to be very~~ high (Weissmann et al., 2002; Danesh-Yazdi et al., 2018). ~~In applied groundwater modeling, stratigraphic~~ Stratigraphic aquifer models with zoned parameters are still widely used ~~,~~ for applied groundwater modeling because the field representation of local-scale heterogeneity is difficult. Given that the aquifer model is stratigraphic and the number of parameters is less than the number of observations, most of the adjustable parameters can be effectively bounded. ~~For this kind of over-determined problems~~In this case, the uncertainty of input data (~~recharge) seems~~ e.g., recharge) appears to have a primary influence on the TTD predictions. Note that here, we do not account for the error caused by model structural deficiency. ~~The trade-off of the measurement error and model structural error can be described by the Minimum Message Length (MML) curve (Wallace and Boulton, 1968; Moore and Doherty, 2006).~~

~~Therefore, we emphasize the primary impact of recharge estimation on the applicability of applied groundwater models. The uncertainty of hydraulic characteristics and its propagation to model predictions has been intensively studied (Weissmann et al., 2002; Moor . In contrast, the potential risk of deterministic recharge estimation and the oversimplified spatial representation of recharge seems to be constantly overlooked in the groundwater modeling efforts. We suggest the uncertainty of recharge, including the uncertainty of mean value and spatial variability, should be primarily considered in the applied modeling of groundwater transport process. Additionally, the modeling workflow used in this study is computationally more efficient than the Bayesian approach, and is suitable for the applied groundwater modeling.~~

6 Conclusions

In this study, we explore the relationship between the uncertainty of recharge, calibration-constrained hydraulic conductivity realizations, and predictions of groundwater TTDs. Using both a ~~physically-based~~ physically based numerical model and a

lumped analytical model, a comprehensive case study is performed in an agricultural catchment (the Nägelstedt catchment). The RWPT method is used to track the water samples through the modeling domain and compute their travel times. Moreover, the analytical model is fitted against the numerical solutions to provide a reference for the effective storage and ~~the~~ sampling behavior of the system. Based on this study, the following conclusions are made:

1. In the ~~stratigraphic aquifer system where most of parameters are effectively bounded by calibration, the uncertainty of TTD predictions are primarily controlled by the recharge . Meanwhile, the simulated TTD can be moderately magnified or attenuated by the different realizations of the post-calibrated hydraulic conductivity fields, given that most parameters are estimable based on the calibration dataset. These insights~~ Nägelstedt catchment model, the simulated MTTs are strongly dependent on the recharge rate and weakly dependent on the postcalibrated K_s field. We highlight the importance of recharge quantification and the worth of reliable data in reducing the predictive uncertainty of TTDs.

2. ~~The analytical solution under the random sampling assumption holds only in idealized aquifer system. It may deviate from simulated TTDs and underestimate the catchment groundwater MTT, particularly in a real-world catchment with typically complex topography, geometry, hydro-stratigraphic structure, and non-uniform recharge.~~

3. The framework of the SAS function provides a good interpretability of simulated TTDs in terms of characterizing the systematic preference for sampling young/old water as outflow. On the basis of this framework, we find that the ensemble simulations have a consistent preference for young water, despite the different recharge and hydraulic conductivity realizations. Our study provides a ~~new possibility to combine the strengths of numerical simulation and analytical SAS function~~ novel modeling framework to explore the effect of input uncertainty and parameter equifinality on TTDs and SAS functions through a combination of calibration-constrained Monte Carlo parameter generation, a numerical model, and a SAS function framework.

4. Both the shape and the breadth of catchment groundwater ~~TTD~~ TTDs and SAS functions are sensitive to the spatial distribution of recharge. Therefore, ~~the~~ a reasonable characterization of the spatial pattern of recharge is crucial for the reliable TTD prediction in the catchment-scale groundwater models.

Appendix A: Random walk particle tracking

Random walk particle tracking solves a diffusion equation at local Lagrangian coordinates rather than the classical advection-diffusion equation, which can be expressed as:

$$\mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + \mathbf{v}(\mathbf{x}(t_{i-1}))\Delta t + Z\sqrt{2\mathbf{D}(\mathbf{x}(t_{i-1})\Delta t} \tag{A1}$$

where $\mathbf{x}$ denotes the coordinates of the particle location, Δt denotes the time step size, and Z denotes a Gaussian random number, with the mean being zero and the variance being unity.

The velocity $\mathbf{v}$ in Eq. (A1) is replaced by $\mathbf{v}_i^*$ to maintain consistency with the classical advection-dispersion equation (Kinzelbach, 1986). The expressions of $\mathbf{v}_i^*$ and the hydrodynamic dispersion tensor $\mathbf{D}_{ij}$ are:

$$\mathbf{v}_i^* = \mathbf{v}_i + \sum_{i=1}^{3} \frac{\partial \mathbf{D}_{ij}}{\partial x_{ij}} \tag{A2}$$

$$\mathbf{D}_{ij} = \alpha_T |\mathbf{v}| \delta_{ij} + (\alpha_L - \alpha_T) \frac{\mathbf{v}_i \mathbf{v}_j}{|\mathbf{v}|} + \mathbf{D}_{ij}^d \tag{A3}$$

where δ_{ij} denotes the Kronecker symbol, α_L denotes the longitudinal dispersion length, α_T denotes the transverse dispersion length, $\mathbf{D}_{ij}^d$ denotes the tensor of the molecular diffusion coefficient, and $\mathbf{v}_i$ denotes the mean pore velocity component in the ith direction.

The stochastic governing equation of 3-D RWPT can therefore be expressed as:

$$x_{t+\Delta t} = x_t + \left(V_x(x_t, y_t, z_t, t) + \frac{\partial D_{xx}}{\partial x} + \frac{\partial D_{xy}}{\partial y} + \frac{\partial D_{xz}}{\partial z}\right)\Delta t + \sqrt{2D_{xx}\Delta t} Z_1 + \sqrt{2D_{xy}\Delta t} Z_2 + \sqrt{2D_{xz}\Delta t} Z_3 \tag{A4}$$

$$y_{t+\Delta t} = y_t + \left(V_y(x_t, y_t, z_t, t) + \frac{\partial D_{yx}}{\partial x} + \frac{\partial D_{yy}}{\partial y} + \frac{\partial D_{yz}}{\partial z}\right)\Delta t + \sqrt{2D_{yx}\Delta t} Z_1 + \sqrt{2D_{yy}\Delta t} Z_2 + \sqrt{2D_{yz}\Delta t} Z_3 \tag{A5}$$

$$z_{t+\Delta t} = z_t + \left(V_z(x_t, y_t, z_t, t) + \frac{\partial D_{zx}}{\partial x} + \frac{\partial D_{zy}}{\partial y} + \frac{\partial D_{zz}}{\partial z}\right)\Delta t + \sqrt{2D_{zx}\Delta t} Z_1 + \sqrt{2D_{zy}\Delta t} Z_2 + \sqrt{2D_{zz}\Delta t} Z_3 \tag{A6}$$

where x, y, z are the spatial coordinates of a particle, Δt is the time step, and Z_i is a random number with a mean of zero and a unit variance.

Appendix B: Composite parameter sensitivity

The PEST algorithm calculates the sensitivity with respect to each parameter of all observations (with the latter weighted as per user-assigned weights), namely the "composite sensitivity" (Doherty, 2015). The composite sensitivity of parameter i is defined as $csp_i = \frac{[\mathbf{J}^t\mathbf{Q}\mathbf{J}]_{ii}^{1/2}}{n}$, where $\mathbf{J}$ denotes the Jacobian matrix that includes the sensitivities of all predictions to all model parameters, $\mathbf{Q}$ is the weight matrix, and n is the number of observations with nonzero weights. In this study, all weights assigned to observations are equally set to 1. Table B1 displays the composite parameter sensitivities in each recharge realization. The mean composite parameter sensitivities of calibrations in all recharge realizations are also included in this table. The hydraulic conductivity of the Middle Muschelkalk (mm) is highly sensitive to groundwater head observations, whereas that of the Lower Muschelkalk (mu) is insensitive to groundwater head observations. The sensitivity of mu, however, varies widely between different recharge realizations, from the highest one in R3 (0.82) to the lowest one in R5 (0.015).

Competing interests. The authors declare that they have no conflict of interest.

Table B1. Composite parameter sensitivities to the groundwater head observations.

Recharge realizations	Parameter sensitivities [-]						
	soil	alluvium	km	ku	mo	mm	mu
R1	6.01	1.89	0.58	1.50	9.47	20.45	0.23
R2	6.15	1.93	0.49	1.52	9.62	20.61	0.35
R3	4.05	1.78	1.38	1.91	7.20	25.84	0.82
R4	5.97	1.91	0.39	1.56	9.41	20.99	0.19
R5	7.31	1.86	0.31	1.34	10.09	19.40	0.015
R6	5.87	1.90	0.39	1.67	9.50	21.03	0.23
R7	7.77	1.93	0.57	1.93	10.53	19.07	0.41
R8	5.03	1.87	0.48	1.81	9.07	22.64	0.11
mean	6.77	1.88	0.57	1.66	9.36	21.25	0.29

Acknowledgements. This study receives support from the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich CRC 1076 AquaDiva. We ~~thank the~~ kindly thank Sabine Sattler from Thuringian State office for the Environment and Geology (TLUG) for providing basic geological data. We acknowledge the EVE Linux Cluster team at UFZ for their support for this study. We ~~thank~~ also acknowledge the Chinese Scholarship Council (CSC) for supporting Miao Jing's stay in Germany.

References

Aigner, T.: Calcareous Tempestites: Storm-dominated Stratification in Upper Muschelkalk Limestones (Middle Trias, SW-Germany), in: Cyclic and Event Stratification, edited by Einsele, G. and Seilacher, A., pp. 180–198, Springer Berlin Heidelberg, Berlin, Heidelberg, 1982.

Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resources Research, 43, https://doi.org/10.1029/2005WR004745, http://dx.doi.org/10.1029/2005WR004745, 2007.

Ameli, A. A., Amvrosiadi, N., Grabs, T., Laudon, H., Creed, I. F., McDonnell, J. J., and Bishop, K.: Hillslope permeability architecture controls on subsurface transit time distribution and flow paths, Journal of Hydrology, 543, 17–30, https://doi.org/10.1016/j.jhydrol.2016.04.071, http://dx.doi.org/10.1016/j.jhydrol.2016.04.071, 2016.

Basu, N. B., Jindal, P., Schilling, K. E., Wolter, C. F., and Takle, E. S.: Evaluation of analytical and numerical approaches for the estimation of groundwater travel time distribution, Journal of Hydrology, 475, 65 – 73, https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.08.052, http://www.sciencedirect.com/science/article/pii/S0022169412007317, 2012.

Bear, J.: Dynamics of fluids in porous media, Courier Corporation, 2013.

Benettin, P., Kirchner, J. W., Rinaldo, A., and Botter, G.: Modeling chloride transport using travel time distributions at Plynlimon, Wales, Water Resources Research, pp. 3259–3276, https://doi.org/10.1002/2014WR016600, 2015.

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., and Rinaldo, A.: Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments, Water Resources Research, 53, 1864–1878, https://doi.org/10.1002/2016WR020117, http://dx.doi.org/10.1002/2016WR020117, 2017.

Benson, D. A., Aquino, T., Bolster, D., Engdahl, N., Henri, C. V., and Fernandez-Garcia, D.: A comparison of Eulerian and Lagrangian transport and non-linear reaction algorithms, Advances in Water Resources, 99, 15–37, https://doi.org/10.1016/j.advwatres.2016.11.003, http://dx.doi.org/10.1016/j.advwatres.2016.11.003, 2017.

Berne, A., Uijlenhoet, R., and Troch, P. A.: Similarity analysis of subsurface flow response of hillslopes with complex geometry, Water Resources Research, 41, 1–10, https://doi.org/10.1029/2004WR003629, 2005.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in water resources, 16, 41–51, 1993.

Böhlke, J. K.: Groundwater recharge and agricultural contamination, Hydrogeology Journal, 10, 153–179, https://doi.org/10.1007/s10040-001-0183-3, 2002.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox, Water Resources Research, 46, 1–18, https://doi.org/10.1029/2009WR008371, 2010.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Catchment residence and travel time distributions: The master equation, Geophysical Research Letters, 38, 1–6, https://doi.org/10.1029/2011GL047666, 2011.

Böhlke, J. K. and Denver, J. M.: Combined Use of Groundwater Dating, Chemical, and Isotopic Analyses to Resolve the History and Fate of Nitrate Contamination in Two Agricultural Watersheds, Atlantic Coastal Plain, Maryland, Water Resources Research, 31, 2319–2339, https://doi.org/10.1029/95WR01584, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95WR01584, 1995.

Cartwright, I. and Morgenstern, U.: Transit times from rainfall to baseflow in headwater catchments estimated using tritium: The Ovens River, Australia, Hydrology and Earth System Sciences Discussions, 12, 5427–5463, https://doi.org/10.5194/hessd-12-5427-2015, 2015.

Cheng, Y., Zhan, H., Yang, W., Dang, H., and Li, W.: Is annual recharge coefficient a valid concept in arid and semi-arid regions?, Hydrology and Earth System Sciences, 21, 5031–5042, https://doi.org/10.5194/hess-21-5031-2017, https://www.hydrol-earth-syst-sci.net/21/5031/2017/, 2017.

Danesh-Yazdi, M., Klaus, J., Condon, L. E., and Maxwell, R. M.: Bridging the gap between numerical solutions of travel time distributions and analytical storage selection functions, Hydrological Processes, https://doi.org/10.1002/hyp.11481, http://doi.wiley.com/10.1002/hyp.11481, 2018.

de Rooij, R., Graham, W., and Maxwell, R. M.: A particle-tracking scheme for simulating pathlines in coupled surface-subsurface flows, Advances in Water Resources, 52, 7–18, https://doi.org/10.1016/j.advwatres.2012.07.022, http://dx.doi.org/10.1016/j.advwatres.2012.07.022, 2013.

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

Doherty, J. and Hunt, R.: Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration, U. S. Geological Survey Scientific Investigations Report 2010-5169, p. 70, https://doi.org/2010-5211, http://pubs.usgs.gov/sir/2010/5169/, 2010.

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

Engdahl, N. B., McCallum, J. L., and Massoudieh, A.: Transient age distributions in subsurface hydrologic systems, Journal of Hydrology, 543, 88–100, https://doi.org/10.1016/j.jhydrol.2016.04.066, http://dx.doi.org/10.1016/j.jhydrol.2016.04.066, 2016.

Fiori, A. and Russo, D.: Travel time distribution in a hillslope: Insight from numerical simulations, Water Resources Research, 44, 1–14, https://doi.org/10.1029/2008WR007135, 2008.

Ginn, T. R.: On the distribution of multicomponent mixtures over generalized exposure time in subsurface flow and reactive transport: Theory and formulations for residence-time-dependent sorption/desorption with memory, Water Resources Research, 36, 2885–2893, https://doi.org/10.1029/2000WR900170, 2000.

Goode, D. J.: Direct simulation of groundwater age, Water Resources Research, 32, 289–296, https://doi.org/10.1029/95WR03401, 1996.

Haitjema, H.: On the residence time distribution in idealized groundwater sheds, Journal of Hydrology, 172, 127–146, 1995.

Hale, V. C. and McDonnell, J. J.: Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison, Water Resources Research, 52, 1358–1374, 2016.

Harman, C. J.: Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, Water Resources Research, 51, 1–30, 2015.

Harman, C. J., Ward, A. S., and Ball, A.: How does reach-scale stream-hyporheic transport vary with discharge? Insights from rSAS analysis of sequential tracer injections in a headwater mountain stream, Water Resources Research, 52, 7130–7150, https://doi.org/10.1002/2016WR018832, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018832.

Healy, R. W. and Scanlon, B. R.: Estimating Groundwater Recharge, Cambridge University Press, https://doi.org/10.1017/CBO9780511780745, 2010.

Heße, F., Zink, M., Kumar, R., Samaniego, L., and Attinger, S.: Spatially distributed characterization of soil-moisture dynamics using travel-time distributions, Hydrology and Earth System Sciences, 21, 549–570, https://doi.org/10.5194/hess-21-549-2017, http://www.hydrol-earth-syst-sci.net/21/549/2017/, 2017.

Howden, N. J. K., Burt, T. P., Worrall, F., Whelan, M. J., and Bieroza, M.: Nitrate concentrations and fluxes in the River Thames over 140 years (1868–2008): are increases irreversible?, Hydrological Processes, 24, 2657–2662, https://doi.org/10.1002/hyp.7835, http://dx.doi.org/10.1002/hyp.7835, 2010.

Jing, M., Heße, F., Kumar, R., Wang, W., Fischer, T., Walther, M., Zink, M., Zech, A., Samaniego, L., Kolditz, O., and Attinger, S.: Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwater model OpenGeoSys (OGS), Geoscientific Model Development, 11, 1989–2007, https://doi.org/10.5194/gmd-11-1989-2018, https://www.geosci-model-dev.net/11/1989/2018/, 2018.

Jochen, L., , R. D., and Röhling, H.-G.: Lithostratigraphie des Buntsandstein in Deutschland, Schriftenreihe der Deutschen Gesellschaft für Geowissenschaften, 69, 69–149, https://doi.org/10.1127/sdgg/69/2014/69, http://dx.doi.org/10.1127/sdgg/69/2014/69, 2014.

Kaandorp, V. P., de Louw, P. G. B., van der Velde, Y., and Broers, H. P.: Transient Groundwater Travel Time Distributions and Age-Ranked Storage-Discharge Relationships of Three Lowland Catchments, Water Resources Research, pp. 1–18, https://doi.org/10.1029/2017WR022461, http://doi.wiley.com/10.1029/2017WR022461, 2018.

Kim, M., Pangle, L. A., Cardoso, C., Lora, M., Volkmann, T. H. M., Wang, Y., Harman, C. J., and Troch, P. A.: Transit time distributions and StorAge Selection functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport variability, Water Resources Research, 52, 7105–7129, https://doi.org/10.1002/2016WR018620, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018620.

Kinzelbach, W.: Groundwater modelling: an introduction with sample programs in BASIC, vol. 25, Elsevier, 1986.

Kirchner, J. W.: Aggregation in environmental systems-Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrology and Earth System Sciences, 20, 279–297, https://doi.org/10.5194/hess-20-279-2016, 2016.

Kohlhepp, B., Lehmann, R., Seeber, P., Küsel, K., Trumbore, S. E., and Totsche, K. U.: Aquifer configuration and geostructural links control the groundwater quality in thin-bedded carbonate–siliciclastic alternations of the Hainich CZE, central Germany, Hydrology and Earth System Sciences, 21, 6091–6116, https://doi.org/10.5194/hess-21-6091-2017, https://www.hydrol-earth-syst-sci.net/21/6091/2017/, 2017.

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J. O., Fischer, T., Görke, U. J., Kalbacher, T., Kosakowski, G., McDermott, C. I., Park, C. H., Radu, F., Rink, K., Shao, H., Shao, H. B., Sun, F., Sun, Y. Y., Singh, A. K., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, Y., Xie, M., Xu, W., and Zehner, B.: OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media, Environmental Earth Sciences, 67, 589–599, https://doi.org/10.1007/s12665-012-1546-x, http://link.springer.com/10.1007/s12665-012-1546-x, 2012.

Kollet, S. J. and Maxwell, R. M.: Demonstrating fractal scaling of baseflow residence time distributions using a fully-coupled groundwater and land surface model, Geophysical Research Letters, 35, 1–6, https://doi.org/10.1029/2008GL033215, 2008.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, https://doi.org/10.1029/2012WR012195, http://dx.doi.org/10.1029/2012WR012195, 2013.

La Venue, M., Andrews, R. W., and RamaRao, B. S.: Groundwater travel time uncertainty analysis using sensitivity derivatives, Water Resources Research, 25, 1551–1566, 1989.

Leray, S., Engdahl, N. B., Massoudieh, A., Bresciani, E., and McCallum, J.: Residence time distributions for hydrologic systems: Mechanistic foundations and steady-state analytical solutions, Journal of Hydrology, 543, 67–87, https://doi.org/10.1016/j.jhydrol.2016.01.068, http://dx.doi.org/10.1016/j.jhydrol.2016.01.068, 2016.

McCallum, J. L., Engdahl, N. B., Ginn, T. R., and Cook, P. G.: Nonparametric estimation of groundwater residence time distributions: What can environmental tracer data tell us about groundwater residence time?, Water Resources Research, 50, 2022–2038, https://doi.org/10.1002/2013WR014974, 2014.

McCann, T.: The Geology of Central Europe Volume 2: Mesozoic and Cenozoic, https://doi.org/10.1144/CEV2P, https://doi.org/10.1144/CEV2P, 2008.

Menning, M.: Deutsche Stratigraphische Kommission (2002) Eine geologische Zeitskala 2002, Deutsche Stratigraphische Kommission (ed) Stratigraphische Tabelle von Deutschland, 2002.

Merz, G.: Zur Petrographie, Stratigraphie, Paläogeographie und Hydrogeologie des Muschelkalks (Trias) im Thüringer Becken, Zeitschrift der geologischen Wissenschaften, 15, 457–473, 1987.

Molnat, J. and Gascuel-Odoux, C.: Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality, Hydrological Processes, 16, 479–492, https://doi.org/10.1002/hyp.328, 2002.

Moore, C. and Doherty, J.: The cost of uniqueness in groundwater model calibration, Advances in Water Resources, 29, 605–623, https://doi.org/10.1016/j.advwatres.2005.07.003, 2006.

Neuman, S. P.: Theory of flow in unconfined aquifers considering delayed response of the water table, Water Resources Research, 8, 1031–1045, https://doi.org/10.1029/WR008i004p01031, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR008i004p01031.

Park, C. H., Beyer, C., Bauer, S., and Kolditz, O.: Using global node-based velocity in random walk particle tracking in variably saturated porous media: Application to contaminant leaching from road constructions, Environmental Geology, 55, 1755–1766, https://doi.org/10.1007/s00254-007-1126-7, 2008.

Remondi, F., Kirchner, J. W., Burlando, P., and Fatichi, S.: Water Flux Tracking With a Distributed Hydrological Model to Quantify Controls on the Spatiotemporal Variability of Transit Time Distributions, Water Resources Research, 54, 3081–3099, https://doi.org/10.1002/2017WR021689, 2018.

Rinaldo, A., Beven, K. J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., Russo, D., and Botter, G.: Catchment travel time distributions and water flow in soils, Water Resources Research, 47, 1–13, https://doi.org/10.1029/2011WR010478, 2011.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, n/a–n/a, https://doi.org/10.1029/2008WR007327, http://doi.wiley.com/10.1029/2008WR007327, 2010.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resources Research, 44, n/a–n/a, https://doi.org/10.1029/2008WR006836, http://dx.doi.org/10.1029/2008WR006836, w00B03, 2008.

Selle, B., Rink, K., and Kolditz, O.: Recharge and discharge controls on groundwater travel times and flow paths to production wells for the Ammer catchment in southwestern Germany, Environmental Earth Sciences, 69, 443–452, https://doi.org/10.1007/s12665-013-2333-z, 2013.

Stewart, M. K., Morgenstern, U., McDonnell, J. J., and Pfister, L.: The 'hidden streamflow'challenge in catchment hydrology: a call to action for stream water transit time analysis, Hydrological Processes, 26, 2061–2066, 2012.

Stewart, M. K., Morgenstern, U., Gusyev, M. A., and Maloszewski, P.: Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future appli-

cations of tritium, Hydrology and Earth System Sciences Discussions, pp. 1–26, https://doi.org/10.5194/hess-2016-532, http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-532/, 2016.

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., and Soulsby, C.: Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions, Water Resources Research, 50, 969–985, https://doi.org/10.1002/2013WR014147, http://dx.doi.org/10.1002/2013WR014147, 2014.

Tompson, A. F. and Gelhar, L. W.: Numerical simulation of solute transport in three-dimensional, randomly heterogeneous porous media, Water Resources Research, 26, 2541–2562, 1990.

Tonkin, M. and Doherty, J.: Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, Water Resources Research, 45, 1–17, https://doi.org/10.1029/2007WR006678, 2009.

Toth, J.: A Theoretical Analysis of Groundwater Flow in Small Drainage Basins, Journal of Geophysical Research, 68, 4795–4812, https://doi.org/10.1029/JZ068i016p04795, 1963.

Van Der Velde, Y., Torfs, P. J. J. F., Van Der Zee, S. E. A. T. M., and Uijlenhoet, R.: Quantifying catchment-scale mixing and its effect on time-varying travel time distributions, Water Resources Research, 48, 1–13, https://doi.org/10.1029/2011WR011310, 2012.

van der Velde, Y., Heidbüchel, I., Lyon, S. W., Nyberg, L., Rodhe, A., Bishop, K., and Troch, P. A.: Consequences of mixing assumptions for time-variable travel time distributions, Hydrological Processes, 29, 3460–3474, https://doi.org/10.1002/hyp.10372, http://doi.wiley.com/10.1002/hyp.10372, 2015.

Van Meter, K., Basu, N., Veenstra, J., and Burras, C. L.: The nitrogen legacy: emerging evidence of nitrogen accumulation in anthropogenic landscapes, Environmental Research Letters, 11, 035 014, 2016.

Van Meter, K. J., Basu, N. B., and Van Cappellen, P.: Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River Basins, Global Biogeochemical Cycles, 31, 2–23, https://doi.org/10.1002/2016GB005498, 2017.

Wallace, C. S. and Boulton, D. M.: An Information Measure for Classification, The Computer Journal, 11, 185–194, https://doi.org/10.1093/comjnl/11.2.185, http://dx.doi.org/10.1093/comjnl/11.2.185, 1968.

Wang, H., Richardson, C. J., Ho, M., and Flanagan, N.: Drained coastal peatlands: A potential nitrogen source to marine ecosystems under prolonged drought and heavy storm events—A microcosm experiment, Science of The Total Environment, 566-567, 621 – 626, https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.04.211, http://www.sciencedirect.com/science/article/pii/S0048969716309573, 2016.

Wechsung, F., Kaden, S., Behrendt, H., and Klöcking, B., eds.: Integrated Analysis of the Impacts of Global Change on Environment and Society in the Elbe Basin, Schweizerbart Science Publishers, Stuttgart, Germany, http://www.schweizerbart.de//publications/detail/isbn/9783510653041/Wechsung_Integrated_Analysis_of_the_Imp, 2008.

Weissmann, G. S., Zhang, Y., LaBolle, E. M., and Fogg, G. E.: Dispersion of groundwater age in an alluvial aquifer system, Water Resources Research, 38, 16–1–16–13, https://doi.org/10.1029/2001WR000907, http://doi.wiley.com/10.1029/2001WR000907, 2002.

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., and Fleckenstein, J. H.: Exploring the Dynamics of Transit Times and Subsurface Mixing in a Small Agricultural Catchment, Water Resources Research, pp. 2317–2335, https://doi.org/10.1002/2017WR021896, http://doi.wiley.com/10.1002/2017WR021896, 2018.

Zghibi, A., Zouhri, L., Chenini, I., Merzougui, A., and Tarhouni, J.: Modelling of the groundwater flow and of tracer movement in the porous and fissured media: chalk aquifer (Northern part of Paris Basin, France), Hydrological Processes, pp. n/a–n/a, https://doi.org/10.1002/hyp.10746, http://doi.wiley.com/10.1002/hyp.10746, 2015.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrology and Earth System Sciences, 21, 1769–1790, https://doi.org/10.5194/hess-21-1769-2017, http://www.hydrol-earth-syst-sci.net/21/1769/2017/, 2017.