Dear referee #1:

General comments:

Thanks very much for taking your time to review this manuscript, and thanks for your comments and valuable suggestions. Below are our point-by-point responses to the referee's comments (in Italics). We hope you will find them to be comprehensive and satisfactory.

1. *Not convincing motivation and unclear computational efficiency.*

   Answer: Thanks for your comment. We will reply to this comment in the two aspects: (1) the motivation of this study; (2) the advantage of the multi-fidelity model (in the aspects of applications); (3) The interpretation of the multi-fidelity model.

   (1) The motivation of this study

   The initial motivation of this study is to calculate time-variant TTD by the direct numerical modeling approach, that is, to quantify time-variant TTD given physical properties (e.g., permeability architecture, stratification etc.) and corresponding boundary conditions. In the introduction section, several approaches to estimate time-variant TTD are reviewed. The age master equation model based on SAS function is an inverse modeling approach. i.e., calculating the parameters of SAS function, TTD and RTD using the long-term observed data, i.e., gauge data (flow rates, rainfall etc.) and tracer data in both rainfall and discharge which are the modeling results. As the long-term tracer data are not available in all catchments, the inverse modeling can be limitedly applied to the long-term monitored catchments. Then several direct modeling approaches are reviewed. The direct modeling approach calculates SAS function, time-variant TTD and RTD based on the physical properties of the system and boundary conditions which may be easily obtained in some catchments. The direct modeling approach mainly includes particle tracking model and groundwater age distribution model. However, the computation cost (including the computation time, computation memory, etc.) will increase as the spatial scale increases or as the Péclet number decreases. The multi-fidelity model is then introduced to overcome this problem, i.e., to reduce the compotation cost especially for the model with large spatial scale and low Péclet number.

   (2) The advantage of multi-fidelity model

   Firstly, the multi-fidelity model is a direct modeling approach relying on readily obtainable data. The input data is the physical properties i.e., permeability distribution, and boundary conditions. This input data can be readily obtained for most of the catchment compared to the input data (long-term observed data) of inverse modeling approach.

   Secondly, the multi-fidelity model consists of three component: low-fidelity model, high-fidelity model and non-linear Gaussian process regression model, and in multi-fidelity model, the ADE model (high-fidelity model) only need to be run for several rainfall events, and the particle tracking model does not need to consider random walk. The results of multi-fidelity model are further calculated from the non-linear Gaussian process regression model. The computation cost of the multi-fidelity model mainly depends on the low-fidelity model, and the accuracy mainly depends on the non-linear Gaussian process regression model. The computation cost includes computation time and usage of computation memory. As for the computation time, the number of rainfall events considered in ADE model of the multi-fidelity model is reduced compared to the groundwater age distribution model; and the smaller number of particles in low-fidelity model will also reduce the computation time. As for the usage of computation memory, if the transport is strongly influenced by molecular diffusion and mechanical dispersion (low Péclet number scenario), or if the model has a large spatial scale, the multi-fidelity model can reduce the number of particles significantly compared to the particle tracking model with random walk (How significantly the number of particles can be reduced will be discussed in the following paragraph in detail). This reduction of the usage of computation memory can be far more than the reduction by simply using a coarser grid as stated in this comment. The reduction of the usage of the computation memory allows the multi-fidelity model to be run on most personal computers, rather than to be run parallelly in a specific cluster system.

To clarify and highlight the advantage of the multi-fidelity and show the case "*where we really need the multi-fidelity model*", line 244-250 will be fully reorganized and extended as a separate section as follows:

"Section 2.5 The advantage of the multi-fidelity model

The advantage of multi-fidelity model is manifested in two aspects: (1) advantage compared to the inverse modeling approach; (2) advantage compared to the approaches in direct modeling approach.

Firstly, the multi-fidelity model is a direct modeling approach which calculates time-variant TTD, RTD and SAS function by simulating water and solute transport in the system with the given physical properties and boundary conditions. The input data of direct modeling approach can be readily obtained for most of the catchment compared to the input data of inverse modeling approach which requires long-term observational data i.e., gauge data (flow rates, rainfall etc.) and tracer data in both rainfall and discharge.

Secondly, the multi-fidelity model consists of three component: low-fidelity model, high-fidelity model and non-linear Gaussian process regression model, and it takes advantage of both the accuracy of high-fidelity model and the low computation cost of low-fidelity model; and it also makes the evaluation of time-variant TTD to be easily implemented in real coordinate space compared with the complex approach that evaluates time-variant TTD numerically in Laplace space (Cornaton, 2012). The computation cost includes computation time and usage of computation memory. As for the computation time, the number of rainfall events considered in ADE model of the multi-fidelity model is reduced compared to the groundwater age distribution model; and the smaller number of particles in low-fidelity model also reduces the computation time. As for the usage of computation memory, when the spatial scale of model increases, the number of particles is reduced significantly compared with the random walk particle tracking model. Specifically, one particle is enough to simulate the transport of a certain water parcel injected at a certain time and a certain point in the model without molecular diffusion and mechanical dispersion. However, in the model considering molecular diffusion and mechanical dispersion, it is likely that more than 10 particles are required to simulate the transport of that water parcel to reduce random noise at laboratory scale; it is also likely that more than 100 particles are required to simulate the transport of that water parcel at hillslope scale; it is even likely more than 1000 particles are required to simulate the transport of that water parcel at large river basin scale. Given a certain Péclet number, the increase of particle number (10 - 100 - 1000) from laboratory scale to large basin scale depends on the increase of the spatial scales of the models. Besides, if the spatial scale of a model is given, the particle number also depends on the Péclet number, and the particle number will increase as the Péclet number decreases (i.e., the increase of the influence of diffusion). The significant reduction of the number of particles leads to the significant reduction of the usage of the computation memory and allows the multi-fidelity model to be run on most personal computers, rather than to be run parallelly in a specific cluster system."

(3) The interpretation of multi-fidelity model

As for the worry of the accuracy reduced in the multi-fidelity model and the difference between the results of multi-fidelity model and high-fidelity model, it has been discussed in section 3 in the manuscript comprehensively. We will also answer these questions in the next comment.

2. *Performance of the multi-fidelity model and design of the test.*

Answer: Thanks for your comment. In summary, there are two questions in this comment: (1) the model error does not seem to converge to zero as shown in Figure 4; (2) Can the multi-fidelity model be applied to a hydrologic system with larger time variability of hydrologic transport dynamics? We will answer the two questions and the question in the last comment (i.e., the difference between the results of multi-fidelity model and high-fidelity model) point by point as follows:

(1) The model error does not seem to converge to zero as shown in Figure 4

Figure 4 presents the results of convergence analysis. The error of multi-fidelity model must converge to zero based a specific theoretical analysis in section 2.3. If the rainfall interval is equal to zero, this means that the ADE model need to be run for all rainfall events, and the multi-fidelity model is identical to the high-fidelity model, so the error of the model with zero rainfall interval must equal zero.

With the specific concern of "*When the rainfall interval of 5 is used, the mean error is a bit larger than the mean error for the case with the rainfall interval of 10*". Figure 4 illustrates the decrease of model error as the rainfall interval decreases. As for the error difference between the rainfall interval of 5 and rainfall interval of 10, the absolute errors of the two cases are almost close to zero. A bit bigger error of rainfall interval of 5 result from the evaluation of the 2-norm error. Each point in Figure 4 represents the mean error of a certain breakthrough curve. Moreover, as the error decreases slowly and approach zero when the rainfall interval is less than 10, rainfall of 10 is recommended in this example for the multi-fidelity model. Lines 381-382 also present the reason for the recommendation of rainfall interval of 10 for other models.

(2) The difference between the results of multi-fidelity model and high-fidelity model

This is to answer the question in the first comment: "*what should a user expect to be the difference between the high-fidelity model result and the multi-fidelity model result*?" The difference between the two models arises from the molecular diffusion and mechanical dispersion process. Specifically, the BTCs of multi-fidelity model are calculated from the non-linear Gaussian regression model whose input is the BTCs of low-fidelity model. This non-linear regression model is also determined by BTCs of ADE model and the number of ADE models which is a sub-component of the multi-fidelity model. As shown in Figure 4 in the manuscript, the errors decrease as the rainfall interval decreases. Figure a1 shown below also illustrates the performance of the Gaussian process regression. The BTCs calculated from ADE models (high-fidelity model) are similar to constrain points in Figure a1, and the BTCs of low-fidelity model are similar to the x-axis in Figure a1. The solid blue line represents the predicted value, and the shadow area represents the uncertainty of the predicted value. The uncertainty can decrease as the number of constrain points increase. Therefore, the accuracy of BTCs calculated from multi-fidelity model is determined by the BTCs of selected ADE models and the number of ADE models (i.e., the numbers of the constraint points in Figure a1). The difference between the results of multi-fidelity model and high-fidelity model represent the errors of BTCs calculated from multi-fidelity model. The error analysis presented in section 3 has already shown the mean error of BTCs of multi-fidelity approximate to be zero (Figure 5 in the manuscript). The error analysis of TTD, RTD and SAS function (Figures 7a3, 7b3, 8a3, 8b3, and 9a3) also show the results calculated from the multi-fidelity model follows the trend of the results of high-fidelity model

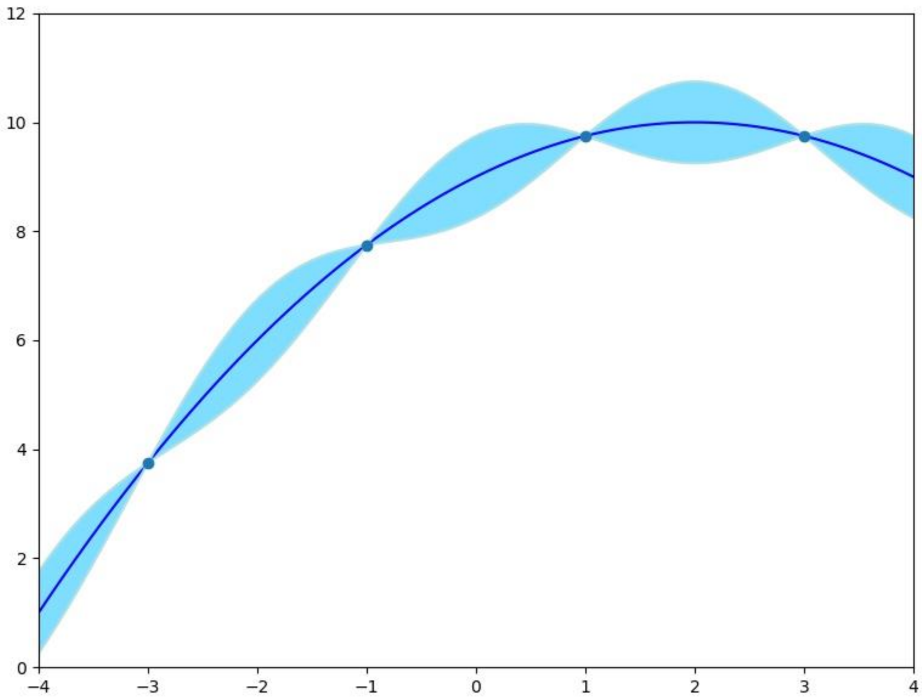

Figure a1 Illustration of the performance of Gaussian process regression

(3) Can the multi-fidelity model be applied to a hydrologic system with larger time variability of hydrologic transport dynamics?

We are happy that the reviewer pointed out some of the future research gaps that is the study hydrologic system with large time variability. In this kind of hydrologic system, if the BTCs vary largely in time, the rainfall interval should be set to be smaller to constrain the results of multi-fidelity model. Moreover, the rainfall interval can also set to vary in time, for example, it is set to be smaller in wet season and bigger in dry season.

3. *Model results.*

Answer: Thanks for your comment. We have double checked the calculation of BTC, TTD, RTD and SAS function. We are sure that the calculations are correct. The calculation follows equation (24) – (28) strictly. Firstly, the cumulative BTCs are of particle tracking model and are calculated by counting the particles. 200 particles are released in particle tracking model during each rainfall event. (If the particle tracking model considers random walk, 200 particles are far from adequate to simulate the diffusion and dispersion process. This also illustrates how efficiency the multi-fidelity model is.). Then the BTCs of particle tracking model are determined by calculating the difference between two successive days. For example, if the cumulative BTC at day 10 and day 11 are cBTC(10) and cBTC(11), the BTC at day 11 is determined as follows:

BTC(11)=cBTC(11)-cBTC(10) for the case cBTC(11)>cBTC(10)

If cBTC(10)=cBTC(11) and cBTC(9)<cBTC(10)=cBTC(11)<cBTC(12), the BTCs at day 10,11,12 are calculated as follows:

$$BTC(10)=BTC(11)=cBTC(11)=(cBTC(12)-cBTC(9))/(12-9)$$

The BTCs of low-fidelity model calculated in this method not only avoid zero value caused by small number of particles released during each rainfall event, but also keep the accuracy of the calculated BTCs.

Moreover, the BTCs of high-fidelity model are calculated by tracking the mass of water from each rainfall event. Then the non-linear mapping is constructed by non-linear Gaussian regression model based on the two sets of BTCs, and the BTCs of high-fidelity model are calculated from the non-linear mapping.

Finally, the TTD, RTD and SAS function are calculated following equation (24) – (28) in the manuscript strictly. The python codes to conduct the calculation have been archived and are available at Zenodo (Mao, 2021).

Therefore, we are sure that the spiky-shaped fSAS functions are correct. Actually, these spiky-shaped fSAS functions can be also found in the previous study (Yang et al., 2018). We hypothesis that this spiky-shaped fSAS function may result from the temporal variation of rainfall injection rate. Meanwhile, as is shown in Figure 9 in the manuscript, the molecular diffusion and mechanical dispersion may smooth the fSAS function. The reason why the fSAS function is spiky may deserve more studies.

Mao, Rong. (2021). Codes and dataset used in the manuscript entitled "Quantifying time-variant travel time distribution by multi-fidelity model in hillslope under nonstationary hydrologic conditions" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5195801

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, https://doi.org/10.1002/2017wr021896, 2018.

4. *Section 2.3.*

Answer: Thanks for your constructive comment. We will reply to this comment in the following three aspects: (1) the objective of this section; (2) the meaning of the “correlation”; (3) explanation of equations (10) to (15).

(1) the objective of this section

As the theoretical foundation of the multi-fidelity model is the mapping between the BTCs of particle tracking model without random walk and ADE model, this section aims to illustrate that there is only one mapping between the results of the two models. The control factor of mapping is also elaborated in this section.

(2) The meaning of the “correlation”

The ‘correlation’ here means ‘mapping”. It denotes the mapping between the results of the two models. The mapping will be used to solve the problem: find the results of ADE model with the given diffusion coefficient and the result of particle tracking model without random walk.

(3) explanation of equations (10) to (15)

As is stated in this referee’s comment, equations (10) to (15) are exactly physical manifestations of “*as the diffusion coefficient decreases, the transport is dominated by advection*”. Equation (15) also shows the connection between the results of the particle tracking model without random walk and ADE model in a simple one-dimensional case. According to equation (15), it is difficult to find the mathematical form of mapping between the analytical solutions of the two models even in such a simple one-dimensional scenario. Undoubtedly, the mathematical form of the non-linear mapping for a much more complicated real-world case is infeasibly to be derived. The motivation of this section aims to delineate why the non-linear Gaussian process regression model is proposed in this study to build up the non-linear mapping.

Therefore, to make the objective of this section clearer, the leading sentence of this section (line 156) will be revised as follows:

“Prior to establishing the multi-fidelity model, the connection between the results of particle tracking model and ADE model is investigated in a one-dimensional case. This connection is related to the key problem of the multi-fidelity model: find the results of ADE model with the given diffusion coefficient and the result of particle tracking model without random walk.”

The last paragraph of this section will also be revised as follows:

“Based on the analytical solutions, if the diffusion parameter D is given, the mapping between the results of high-fidelity model and low-fidelity model is unique. This is the theoretical foundation of the multi-fidelity model. However, according to equation (15), it is difficult to find the mathematical form of the mapping between the analytical solutions of the two models in such a simple one-dimensional scenario. Undoubtedly, the mathematical form of the non-linear mapping for a much more complicated real-world case is difficult to be derived.”

Moreover, the title is also changed as “Connection between the results of low-fidelity model and high-fidelity model in one-dimensional scenario”.

5. *Minor comments.*

5.1 ***Needs less focus on the hillslope scale****: In many places throughout the manuscript, the authors talked about hillslope (e.g., in the title, L3, L18, and so on). However, I am not sure if the multi-fidelity model could be useful at the hillslope scale, which is relatively small. For a small hydrological domain like a hillslope, running the full ADE simulations or the* *particle tracking model with the random component will be feasible.*

Answer: Thanks for pointing out this problem. The multi-fidelity model will be more useful for a model of large spatial scale with low Péclet number, and the performance of the multi-fidelity model (error analysis and convergence analysis) is validated in the hillslope scale. We will change "hillslope" to "catchment" in the revised manuscript except the example section.

5.2 ***Diffusion process****: It is unclear what the diffusion process simulated in the 2D hillslope model is. The authors may have assigned a specific concentration for the rainfall event of interest while the concentration of all other water is set to zero. Since the diffusion is controlled by the concentration gradient and the diffusion coefficient, it needs to be more precise what the gradient in the model (which depends on the assigned value for the rainfall event of interest) means and what the diffusion coefficient means.*

Answer: What the diffusion process simulated in the 2D hillslope is already discussed in section 2.1, i.e., the transport of age density function $\rho'(\vec{x},t,\eta)$, where $\vec{x}$ is the position vector, t is the clock-time, η is the injection time (Gomez and Wilson, 2013). Age density function is an analogue to solute concentrations in the simulations. The specific "concentration" of a certain rainfall event is set to be 1000 kg/m^3 which means that all the rainfall water has the same age. When the rainfall water is infiltrated into the system, the proportion of water from a certain rainfall event (i.e., the age density function) is less than one. The transport of age density function is similar to the transport of solute. Therefore, the gradient in the model means the spatial distribution of the age density function, and the diffusion coefficient describes the molecular diffusion and mechanical dispersion of water parcels.

To clearly clarify what the diffusion process simulated, line 310 will be revised as follows:

"The transport of dilute species in the porous media module in COMSOL is used to solve for the age density function of water determined by groundwater age distribution model (Gomez and Wilson, 2013)."

[Reference]

Gomez, J. D., & Wilson, J. L. (2013). Age distributions and dynamically changing hydrologic systems: Exploring topography - driven flow. Water Resources Research, 49(3), 1503-1522.

5.3 *L5: "transient ground water flow" is not the only process that could result in time-variant TTDs.*

Answer: Thanks for your comment. We agree with this comment. Line 5 introduces the general origin of time-variant TTD. The time-variant TTD emerges in transient groundwater flow system, and the transient groundwater flow system may result from multiple processes except for the non-stationary rainfall input. We will change this sentence as follows in the revised manuscript:

"Under nonstationary hydrologic conditions such as nonstationary rainfall input, the TTD varies with transient groundwater flow system, leading to a time-variant TTD."

5.4 *L11-13: In is not clear how significantly the number of particles can be reduced.*

Answer: Thanks for your comment. Due to limited space of abstract, this issue is discussed later in the manuscript. This question has already been answered in the response to the major comment 1.

5.5 *"yearly averaged" does not seem to describe the time-invariant TTD correctly.*

Answer: Thanks for your comment, and we agree it. We will revise this sentence as follows:

"The time-invariant TTD quantifies transport process under steady-state condition theoretically and may capture the general trend of the time-variant TTD under unsteady state condition."

5.6 *L67-69 and L146-148: While I agree, this sentence is not enough to motivate this study. I think the authors need to cite studies that show that considering diffusion or dispersion requires significantly more particles (if such a study exists) or need to show it clearly somewhere in the manuscript.*

Answer: Thanks for your comment. L67-69 and L146-L148 show one of the motivations of this study. The detailed explanation is already presented in the response to major comment 1.

5.7 *L76: Not clear what message the authors are trying to convey.*

Ambiguous sentence in L76 will be revised as follows:

"but most current software or codes do not have the module to discretize the 5D space, so they cannot be further used to numerically solve the 5D equation directly."

5.8 *L83: Not clear if the authors solved the full 5D equation or just combine the results of the high-fidelity model and the low fidelity model.*

Answer: Thanks for your comment. The ambiguous sentence in Line 83 will be revised as follows:

"Then, time-variant TTD can be calculated by the multi-fidelity model."

5.9 L152: No evidence is provided to support the "drastic" decrease of the number of particles

Answer: Thanks for your comment. The detailed explanation is already presented in the response to major comment 1.

5.10 L175: While I agree with the "no matter what boundary conditions" argument, what the authors show in this section is for a specific boundary condition used to get (14).

Answer: Thanks for your comment. The ambiguous statement in Line 175 will be extended as follows to avoid the misunderstanding:

"As D approaches zero, the governing equation (13) (considering molecular diffusion and mechanical dispersion) converges to the governing equation (10), so no matter under what boundary conditions, the solution of equation (13) will always converge to the solution of equation (12) when D approaches zero."

5.11 L265: There are many studies where RTDs are estimated using BTCs. Most of the experimental studies of the SAS function and many the SAS function studies that utilize process-based model estimated RTDs using BTCs.

Answer: Thanks for your comment. We will delete this inappropriate sentence.

5.12 L280-281: This sentence is unnecessary for the purpose of this paper.

Answer: Thanks for your comment. We will delete this unnecessary sentence.

5.13 L291: Lower concentration of which solute?

Answer: Thanks for your comment. We will change "low solute concentration" to "lower reactive solute concentration".

5.14 L305: "quasi-steady state" means something different.

Answer: Thanks for your comment. This ambiguous statement in Line 305 will be revised as follows:

"but in order to get rid of the influence of the uncertainty of initial condition rapidly"

5.15 L313: Each single rainfall event at daily time step?

Answer: The modeling time-step for a certain rainfall event is adjusted by the convergence rate of the current time-step dynamically. If the convergence rate is too low, the length of modeling time step will decrease, and vice versa.

5.16 L320-321: Richards' equation-based model has many assumptions. Also, perhaps better to avoid the term "real physical process".

Answer: Thanks for your comment. We will delete this sentence.

5.17 Figure 3: Please clarify what the particles at x = 100 m are.

Answer: The particles at x=100m are the particles arrived at x=100m before the time=1050 day. This will be explained in the caption of Figure 3 in the revised manuscript.

5.18 L373-376: A model result of solute concentration cannot be lower during the whole time than the concentration estimated using another model.

Answer: Thanks for the comment. There is a misunderstanding of the two models. In Figure 5, the low-fidelity model does not consider the molecular diffusion and mechanical dispersion, but the high-fidelity model and multi-fidelity model consider these processes. This leads to the concentration difference.