Dear referee #3:

General comments:

Thanks very much for taking your time to review this manuscript. I really appreciate your comments and suggestions. In the following we will address them point by point.

Response to general comments:

1. *The terminology of high-to-low fidelity is confusing or perhaps even misleading.*

   We have made effort to refine the definition of high-to-low fidelity model, and differentiate high-to-low fidelity model. Actually, the multi-fidelity model consists of two components, namely high-fidelity model and low-fidelity model. The high-fidelity model is the ADE model, and the low-fidelity model is the particle tracking model without random walk. The high-fidelity model and low-fidelity model are related by non-linear Gaussian process regression model to build up the multi-fidelity model. The multi-fidelity model combines the efficiency of low-fidelity model and accuracy of high-fidelity model. The advantages and characteristics of the multi-fidelity model related to the high-fidelity and low-fidelity models have been fully expanded to clarify their difference and applications. They are also detailed in the response to the following comments.

2. *It is also unclear the proposed mixed approach is "better" in a computational sense, the authors jump to this conclusion seemingly without comprehensive evaluation.*

   Thanks for the comments. 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 reduced number of particles 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 at most personal computers, rather than to be run parallelly in a specific cluster system.

   To clarify and highlight the advantage of the multi-fidelity, 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."

Specific Comments

1. S2.2 / lines 141-154. *The approach presented has similarities and other approaches (e.g. Maxwell 2019, Maxwell 2016, Remondi 2018, Wilusz, 2020), the statements about inefficiency*

of the particle tracking (low-fidelity) approach are not well justified.

Answer: Thank you for suggesting these constructive papers. In section 2.2/ lines 141-154, two particle tracking models are summarized, i.e., particle tracking models with or without random walk. The number of particles required between the two models is compared in line 146-154: as the molecular dispersion and mechanical diffusion are not considered in the low-fidelity model, the number of particles required in low -fidelity model will be much smaller than the number in particle tracking model with random walk, especially when the spatial scale of the model increases. Particle tracking without random walk (low-fidelity model) is taken as the component in the multi-fidelity model to improve the computational efficiency. So we assume the comments of 'inefficiency of the particle tracking (low-fidelity), would actually to be "efficiency of the particle tracking". Moreover, based on the suggested literatures, extra comparative statement will be added to justify the difference among these particle tracking models and limitations of the particle tracking (low-fidelity) models.in the revised manuscript. . To address this comment, parts of section 2.2 will be fully revised and extended to address as follows:

"Particle tracing model can be classified into two categories generally: particle tracking model with random walk, and particle tracking model without random walk. The particle tracking model without random walk will be taken as the low-fidelity model, and is set to be one of the components of the multi-fidelity model. Until now, there are several methods to improve the efficiency of particle tracking model (Yang et al, 2021a, b), but these are algorithm improvement, and the improved efficiency is very limited. The computation efficiency of particle tracking model without random walk can be several orders of magnitude lower than the efficiency of the particle tracking model with random walk by Yang et al, 2021a ,b. The improvement of the efficiency will be discussed after section 2.4 in detail. "

The improvement of the efficiency is also discussed in the response to general comment 2.

[References]

YANG, C., Maxwell, R. M., & Valent, R. (2021). Accurate load balancing accelerates Lagrangian simulation of water ages on distributed, multi-GPU platforms.

Yang, C., Zhang, Y. K., Liang, X., Olschanowsky, C., Yang, X., & Maxwell, R. (2021). Accelerating the Lagrangian particle tracking of residence time distributions and source water mixing towards large scales. Computers & Geosciences, 151, 104760.

2. *Subsurface heterogeneity can impart significant differences on macro scale behavior (e.g. Benson 2019; Danesh-Yazdi 2018, Engdahl 2014)*

Answer: Thank you for this remark. Firstly, we admit the influence of subsurface heterogeneity and other physical properties on macro scale behaviors. However, as is stated in the abstract and last paragraph of the introduction section, the main purpose of this study is to propose a multi-fidelity model to quantify the time-variant travel time distribution by solving the 5-dimensional age distribution. The illustrative example established in section 3 is a 2-dimensional homogeneous hillslope aquifer. The purpose to construct the 2-dimensional model is to perform the error analysis and convergence analysis and to illustrate the performance of the multi-fidelity (see line 282-285). However, comment 2 indeed let us to consider the question: can the subsurface heterogeneity influence the performance of the multi-fidelity model? In the future, we are happy to perform extensive and separate study to explore this question with our model. To make our statement more inclusive and comprehensive, we will add a new section 3.4 in the revised manuscript as follows:

"The performance of multi-fidelity model has been proved in the 2-dimensioal homogeneous aquifer. It can be further inferred that the multi-fidelity model can be applied to arbitrary 2-dimensional/3-dimensional heterogeneous aquifers based on the construction of multi-fidelity model in section 2.4. The cumulative BTCs calculated by multi-fidelity model is based on the mapping between the cumulative BTCs calculated by high-fidelity model and low-fidelity model. High-fidelity and low-fidelity components of the multi-fidelity model are based on the same physical model, i.e., the 2-dimensional/3-dimensional homogenous/heterogeneous hillslopes or catchments. Therefore, the multi-fidelity model is also applicable to study the influences of heterogeneity on macro behaviors (i.e., TTD), which will be also a direction of future studies on the multi-fidelity model."

3. *Effects beyond the hillslope are important (e.g. Kollet 2008; Maxwell 2016; Remondi 2018; Wilusz 2020) and should be discussed.*

   Answer: multi-fidelity model can also be optimistically applied to the spatial scales beyond the hillslope (as we mentioned in the response to general comment 2). Those suggested literatures are quite helpful and constructive, and we have included them in the revised manuscript for better justification.

4. Some references included below that the authors may want to read / consider. This is not a comprehensive list.

   Answer: Thank you for suggesting these constructive literatures, we will add these in the revised manuscript.