Reply to Reviewer #3

We would like to thank the reviewer for his thorough review. We appreciate his time and effort and all his suggested comments, which improved the quality of our work. We followed the reviewer's suggestions and revised the manuscript accordingly. Below please find our detailed response to the reviewer's comments.

In the paper by Gharamti et al., the authors compare three data assimilation strategies for a subsurface state-parameter estimation problem: the standard ensemble Kalman filter (EnKF), a hybrid EnKF including optimal interpolation (EnKF-OI) and a second order sampling formulation of EnKF (EnKF-ESOS). Synthetic data assimilation experiments are performed with a reactive transport problem for migration, sorption and degradation of chlorinated hydrocarbons. This set-up should mimic a contaminated aquifer in the port of Rotterdam. Concentration data and first-order degradation rates are updated within the three assimilation schemes.

The paper is well written and points out important limitations of the ensemble Kalman filter in subsurface characterization (under-sampling of forecast covariances and observation errors) and how they could be ameliorated with EnKF-IO and EnKF-ESOS. However, I have two major concerns regarding the content of the paper:

1- It seems to me that there is a considerable overlap with earlier work from Gharamti et al. (2014). Large parts of the paper related to the EnKF-OI contain very similar information as in the earlier work and also the overall model set-up is quite similar in both studies (see below) leading to almost the same conclusions regarding EnKF-OI. Therefore, the authors should give a clear motivation why the comparison EnKF/EnKF-OI is repeated in this paper and they should point out what is the innovative aspect of this study compared to their previous work (i.e., what did we learn from this study regarding EnKF/EnKF-OI that was not already covered in Gharamti et al., 2014).

We thank the reviewer for pointing this out. The previous work only introduced the hybrid EnKF-OI formulation to state-parameters estimation problems. We agree that part of the methodology has some overlaps but the overall goal of the two studies is quite different. The major differences between both studies are listed here:

a. The update step of the hybrid EnKF-OI algorithm is extended. We allow the observations to be processed serially, and therefore the optimization presented in *Gharamti et al. (2014)* to be performed for each single observation separately. We believe that this is a more convenient approach given that different observations carry varying degrees of information to the system. As such the weighting between the ensemble and the background covariances changes when assimilating the observations serially. The serial assimilation makes the update scheme consistent with that of the EnKF-ESOS, which requires the observations to be processed serially.

b. The application presented in this study is based on a large-scale and more realistic problem. Although the reviewer seems to think that the model setup is similar, there are many differences to the one used in *Gharamti et al. (2014)*. In particular:
    i. The model is three-dimensional (unlike the 2D problem in the previous article) and the hydraulic parameters such as porosity and permeability

are based on real geologic facies. In the revised manuscript, we now outline the procedure we follow to construct the parameters (using the GeoTOP software package). The masking of the domain location (in the port area) and the confidentiality of the contaminant data are two conditions imposed by the municipality of Rotterdam (it is not in our control). Further details on the flow model parameters are now included in Section **3.2.2**.

ii. The vertical resolution of the model is fine and quite unique compared to many other model setups found in the literature. Many studies assume a single layer (maybe 2 or 3 at most) for each aquifer. We however discretize the vertical model domain into 120 layers, covering 4 aquifer systems, each of length 0.5 m. This helps to understand the interaction between the components and eventually provide more insights on the correlations between the parameters and the associated component concentrations.

c. Different optimization strategies for determining the weighting factors in the hybrid algorithm are now examined. *Gharamti et al. (2014)* only considered maximizing the information gain to weight between the flow-dependent and the static covariances. In the current study, we test and analyze different optimization scenarios (Section **4.1.2**). For instance, maximizing the information gain when hybridizing the state statistics (i.e., α) and minimizing information gain when hybridizing the parameters statistics (i.e., β). We also assess the performance when the information gain is minimized for both state and parameters statistics (refer to lines 71-73 and 543-548).

d. The EnKF-ESOS algorithm is not yet tested for state-parameters estimation problems. It was only presented for state estimation only (refer to lines 85-87).

e. One important message we emphasize in this manuscript is the efficiency of each approach (OI and ESOS) within the EnKF. In other words, we quantify the improvements that could be achieved when we tackle the under-sampling issues that are related to the limited ensemble size or the observational sampling errors. This is discussed in Section **4.2** (and lines 621-626).

2- The authors claim to use a 'reality-inspired' test case for the comparison of the different data assimilation schemes. In fact, only a limited amount of information about the site characterization is given in section 3.1, which makes it difficult for the reader to judge how realistic the model set-up is. For example, how many measurements were available to derive the parameter fields for hydraulic conductivity, porosity and distribution coefficients and how uncertain are the derived parameter fields? Is the model discretization fine enough to account for the spatial variability of subsurface parameters? Another question is whether the assumption of steady state groundwater flow is valid for the chosen site. Usually, one would expect transient groundwater flow due to temporally variable recharge, pumping activities or density-driven flow in such environments. Transient groundwater flow could have important implications for the data assimilation, e.g., for the determination of the background covariance matrix in the EnKF-OI scheme (see below). Overall, the current set-up is very similar to what has been used in Gharamti et al. (2014) except that groundwater flow is 3D in this example (which should not be a major issue when a steady state flow field is used) and the chemical reactions are different (but follow a very similar mathematical description).
So in fact, I think that there is not much more complexity in this 'reality inspired' setup than in the 'purely' synthetic set-up used in previous studies. Therefore, I suggest the authors to

add more complexity in their model set-up in order to test the different assimilation schemes under more realistic conditions. This could be accomplished e.g., by considering more sources of uncertainty (e.g. hydraulic parameters, forcing terms) and by using transient flow conditions.

We thank the reviewer for bringing this discussion about the model. We now discuss the initialization process for the parameters and further study the impact of model uncertainty on the performance of the schemes. Our response for each point is detailed below:

a. The hydraulic conductivity is provided in the database GeoTOP. The GeoTOP for the province of South-Holland is constructed using 46.000 borehole data. Using the borehole data, the most probable lithostratigraphy and lithofacies have been estimated in each voxel of 100x100x0.5 m. In the next step the GeoTOP uses relations between the lithostratigraphical units and the lithofacies with parameters such as hydraulic conductivity, porosity and organic carbon content in order to provide these parameters on the voxel scale. Further details and related references are now included in Section **3.2.2** of the revised manuscript.

b. The horizontal model discretization (50 m) is finer than the resolution of the parameters such as the hydraulic conductivity (100 m) and the vertical dimensions are equal. The vertical discretization is 0.5 m (for each layer) and this is a considerably fine resolution.

c. In our opinion, steady state groundwater flow is a valid assumption. We agree that there are temporal variations on a small scale, such as tidal influences and yearly fluctuations of precipitation and evapotranspiration. Effects of tidal influences are expected to be minor (yearly averaged additional advection is zero but it may increase spreading that is accounted for by a relatively high effective dispersivity values). Effects of yearly fluctuations of precipitation and evapotranspiration are also expected to be minor as the near surface groundwater levels are controlled by the drainage levels of the drainage systems in the port area (3 – 4 m above sea level) and the deeper groundwater levels are predominantly influenced by the surface water levels in both the polders area (managed levels around or below sea level) and the large surface waters (approximately sea level). Temporal variations due to density driven flow can be also neglected as we would expect only minor changes in the most lower part of the model domain on the time scale of 50 years. Similar discussion has been added to Section **3.2.2**.

d. The reviewer is raising an interesting point regarding the background covariance of the hybrid EnKF-OI and the connection to perfect flow conditions. Following the reviewer’s suggestion, we have added a new section (**4.3**) to the results in which we run a new set of experiments while perturbing the hydraulic conductivity and the porosity. We test the performance when strong and moderate uncertainties are imposed. We found that imposing large uncertainties on the hydraulic parameters strongly degrades the performance of all filtering schemes. Given that the performance of the hybrid EnKF-OI depends on the quality of the background statistics, satisfactory results were obtained only when the uncertainty imposed on the background information is moderate and not very high. Further details can be found in the Abstract, Section **4.3** and Figures 18 and 19 of the revised manuscript. Thank you.

Specific comments

Line 191-192: The same applies for the alpha and beta values in EnKF-OI.

Yes, if they are set manually. Our proposed adaptive scheme does not require any tuning effort.

Line 195-196: What do you mean with '...dynamically constants quantities....'?

We meant to say that they are static in time unlike the state (e.g., concentration) of model, which evolves based on the dynamics of subsurface. We clarify this in the revised manuscript.

Line 216-217: Incomplete sentence.

We modified the sentence, which reads now: "This decomposition is useful in practice in order to reduce computational burden and memory storage." Thank you.

Line 368-370 and Figure 5: Why does PCE appear in layer 40, when the contaminant source is located in layer 60 and the pre-dominant flow direction is downward? Is the groundwater flow rate so low compared to molecular diffusion?

Overall, the groundwater flow rate (in the downward direction) is stronger than molecular diffusion. However, since we consider a constant source term for PCE in layer 60 and given the long simulation time (i.e., 50 years) a small amount of this component appear in layer 40 (under the effect of molecular diffusion). We checked the other 3-contaminant components, and none of them reach layer 40 by the end of the simulation period. Thank you.

Line 396-417: In this example, the background covariance matrix for EnKF-OI is derived on the basis of a steady-state flow field with perfectly known hydraulic parameters.
Additionally, the background covariances are derived from the same time period, where the assimilation experiments are performed. This means that the derived background covariance matrix contains a very precise description of the relation between concentrations and degradation rates in your system. However, under real-world conditions the uncertainties in hydraulic parameters may have a considerable impact on the quality of the background covariance matrix. Additionally, under transient flow conditions it might be much more difficult to derive a good estimate of the background covariance matrix. Therefore, I suggest the authors to discuss such practical issues in more detail and also to perform additional simulation experiments where these influences on the derivation of the background covariance matrix are assessed in more detail, e.g. by introducing uncertainty in the hydraulic parameters and by using transient flow conditions. This would provide a more realistic assessment of the EnKF-OI assimilation scheme.

The reviewer has a good point. As mentioned in our response to the reviewer's second major comment, we now include a set of experiments that are based on imperfect hydraulic parameters. In Figure 17, we analyze the effect of perturbing the flow model on the background cross-correlations. Essentially, we found that the dominant correlation patterns are similar to those obtained using perfect flow conditions, especially in the shallow aquifer layers. The magnitude of the new background correlations, however, is considerably smaller. Generally, porosity and conductivity affect the speed and the movement of groundwater in the aquifer and thus the degradation process would be expected to either slow down or accelerate. A similar discussion has been included in the revised manuscript (lines 602-613). Thank you.

Figure 11: It would be helpful in this plot to also show the evolution of concentration values without data assimilation as a comparison. Additionally, why does the optimized EnKF-OI simulation (grey lines) for PCE update in the wrong direction between year 5 and 10?

We have added the free run concentration evolution to Figure 11. The update of PCE between year 5 and 10 does not entirely go in the wrong direction. Over all, the concentration is close to the reference solution by year 10. The difference in the behavior to that of the EnKF could be related to the rapid adjustment of the biodegradation rates right after incorporating information about the background state-parameters cross-correlations.