

## Reviewer 2

We thank reviewer 2 for his/her thoughtful review that greatly helped to improve our paper. In this document, we have entered your comments in italics, added our response as regular text, and subsequently added suggested changes to the paper in red. After our responses you can find the suggested changes made to the text in red. Line numbers are written as “[P...L...]”, indicating page number, line number in the no markup document.

### General comments

*General comments: This paper describes numerical modelling used to reconstruct the salinity distribution in the Nile Delta aquifer over the past 32 ka. The model simulations include the Holocene marine transgression and various scenarios relating to possible aquifer lithology, and sources of hypersaline seawater. Through consideration of palaeo-environments, the authors aim to better represent the Nile Delta aquifer and an observed freshening. The authors state that simulation of the freshening of the aquifer has not been possible previously using realistic model parameters and steady-state modelling. The objectives of this modelling study are stated as being (p4, L8) to: 1) investigate the physical plausibility of the Holocene transgression hypothesis for the Nile Delta; 2) investigate the influence of the uncertain geology; 3) provide estimates of the present-day fresh groundwater [FGW] volumes; 4) assess the importance of using palaeo-reconstructions compared to less expensive steady-state modelling. These objectives have been addressed within the study and the description of how this was achieved is generally clearly detailed. There are not very many palaeo-reconstruction type studies looking at saltwater-freshwater dynamics in coastal aquifers and therefore there is some novelty in this contribution. However the significance is not clear. The study relates specifically to the Nile Delta aquifer and the more generalised extension of the science is not apparent. As such the study is primarily a case study, albeit with some interesting insights for other similar deltaic systems. For this manuscript to be suitable for publication in HESS the authors need to indicate how the study provides a substantial contribution, and is not just a case study.*

We think with the help of also the general comments of Reviewer 1, we managed to change the focus of this paper from the regional problem to the scientific problem. We start the introduction [P2 L1-7] and abstract now with the general scientific problem. Furthermore, we discuss the scientific problem now at the start of our discussion [P15 L1-30] and the local management issues are kept as a later part of the discussion [P16 L19-30]. Furthermore, the start of the abstract is changed [P1 L11-12]. This adaptation should more clearly aid in stressing the scientific significance of this research.

Furthermore, given the wide range of lithological scenarios, several of the findings are also applicable to other deltas: The severe decrease in FGW volume after the marine transgression (at the start of the Holocene), the large variation between the different palaeohydrogeological reconstructions compared to steady-state scenarios, and the lithologies that favor offshore fresh groundwater volumes in deltas.

To show this stress this more, we added this to the conclusion [P18 L6-L9]:

**Given the large range variation in delta-architectures analyses, we can conclude that steady-state models are not likely to result in realistic FGW distributions in deltaic areas. Our results also show that the occurrence of past marine transgressions constitute a valid hypothesis explaining the occurrence of the extensive saline zone land inward.**

## Specific comments

*Specific comments: The use of iMOD-SEAWAT has presumably assisted in reducing run times and making modelling of the various scenarios used in the study tractable. The paper would benefit from more information describing the run times and benefits of using iMOD-SEAWAT. I expect that other researchers will find this interesting for similar palaeo-reconstruction type work.*

The added benefits of this code were already described (namely faster runtimes on computers with multiple CPUs), we however could add some information on run times [P6 L32 – P7 L1].

Simulations were conducted on the Dutch national computational cluster Cartesius (Surfsara, 2014), using Intel Xeon E5-2690 v3 processors. With this new code the model scenarios had a wall clock time ranging of 44 hours to 108 hours on 48 cores, depending on model complexity.

*The study involved numerous scenarios and I found many of these confusing as I read through the paper. The section describing the coding of the scenarios needs improvement. For example, I was not able to discern the meaning of P in the scenario H-F-B-P (p10, L1), despite that this was used as part of the explanation of how the coding system works.*

We can understand that this might be confusing, as it was very difficult to come up with an appropriate coding system. To reduce the confusion we expanded the example and added each feature-member combination behind each piece of the sentence where it was specified [P10 L19-23]:

For each feature, the corresponding letters in table 1 are converted to a code as follows: {sea}-{clayer}-{prov}-{temp}. For example, the palaeohydrogeological reconstruction (temp: P) of an aquifer with a half-open sea connection (sea: H), fluvial horizontal clay layers (clayer: F), and HGw seeping in from the bottom (prov: B) gets the following code: H-F-B-P. Its equivalent steady-state model is attributed the code H-F-B-S.

*Additionally, the term ‘behavioural’ model scenarios (p. 12, L1) seems odd and I suggest changing.*

Copying from response to comment to P11L19 by Reviewer 1:

With the term “behavioural” we mean the ability of models to reproduce certain patterns observed, following Beven and Binley (1992) *The future of distributed models: Model calibration and uncertainty prediction*. Whatever these patterns specifically are, is up to the researcher to decide. So, on second thought, perhaps the word “acceptable” captures the inherent arbitrariness of this decision better. Hence, we have changed the word “behavioural” to “acceptable” in the paper.

In the text [P12 L29 – P13 L3]:

More striking differences are observed in the hypersaline zone, where we observe a division around  $\Lambda = 0.07$  into two groups. There are the scenarios with  $\text{Md}[\Lambda] < 0.07$ , that predict the location of the HGw with similar skill as to with which they predict the location of saline groundwater. We call these scenarios “acceptable”. Specifically, these are the following five model scenarios: C-M-B-P, C-N-T-P, H-M-T-P, H-F-T-P, H-N-T-P. The other scenarios perform considerably worse in predicting the location of the HGw.