October, 7 2017

Re: Response to the review of Anonymous Referee #2 of the manuscript "Groundwater withdrawal in randomly heterogeneous coastal aquifers" by Martina Siena and Monica Riva.

We appreciate the efforts Anonymous Referee #2 has invested in our manuscript and we are grateful for his/her insightful comments. Following is an itemized list of his/her comments (in italic) and our responses.

> *GENERAL COMMENTS*
> *Seawater intrusion is a major problem in coastal aquifers, and several studies are attempting to improve its numerical simulation. The authors want to underline how 1) the heterogeneity of the porous media impacts the numerical simulations of coastal aquifers and 2) different configurations of the pumping scheme effect the position of the saltwater wedge and the width of the mixing zone. To answer these questions, the numerical solutions of the coupled flow and transport equations are compared considering homogenous and randomly heterogeneous permeability of the porous media. I find the topic of the manuscript of interest for HESS readers. The methodology presented is clear and the manuscript is well written.*

We thank the Reviewer for his/her appreciation of our work.

1. *However, in my opinion further investigation is needed to better support the conclusions proposed. In particular I am concerned with the following points: 1) The Monte Carlo analysis is performed using only 60 random realizations. I can understand that MC simulations of this 3D, coupled system are computational intensive, however a brief analysis on the convergence of the MC scheme is required to understand the sensitivity of the first and second moments of the computed metrics to the ensemble size (e.g. in the case without pumping).*

We agree with the Reviewer in that the number of Monte Carlo (MC) realizations, n, is critical. The choice of n typically results from a trade-off between CPU time and accuracy in reproducing the statistical moments of the quantities of interest. For example, Pool et al. (2015) rely on $n = 50$ to analyze the effect of tidal fluctuations on SWI in a three-dimensional heterogeneous system. Additionally, the choice of the number of Monte Carlo realizations depends on the type of quantity that one is interested in (for example, either point-/local- or integral- quantities, as also detailed below) and on the type of behavior which is intended to be highlighted. Note that, for the test cases analyzed in the manuscript, a single MC simulation on a i7-3930K Intel core with 32GB memory requires about 3 hours.
In the following we assess the stability of the MC-based results on the basis of the methodology proposed by Ballio and Guadagnini (2004). Figures R1 and R2 respectively depict the sample mean and standard deviation of all metrics, ξ'', considered in the manuscript versus n. Results are evaluated (for the test case without pumping) at the end of the 8-year period considered. The estimated 95%-confidence intervals, computed according to eqs (3) and (8) of Ballio and Guadagnini (2004), are also shown (see also our reply to item 8). Figures R1 and R2 show that the oscillations displayed by the quantities of interest are in general limited and do not hamper the strength of the main massage of our work. As expected, integral quantities (i.e., A_T'', A_S'', V_T'', V_S'', W_{MZ}'') tend to stabilize faster than local ones (L_T'' and L_S''). We plan to (*i*) increase n up to 100 (for each scenario investigated) and (*ii*) include this convergence analysis as supplementary material in the revised manuscript.

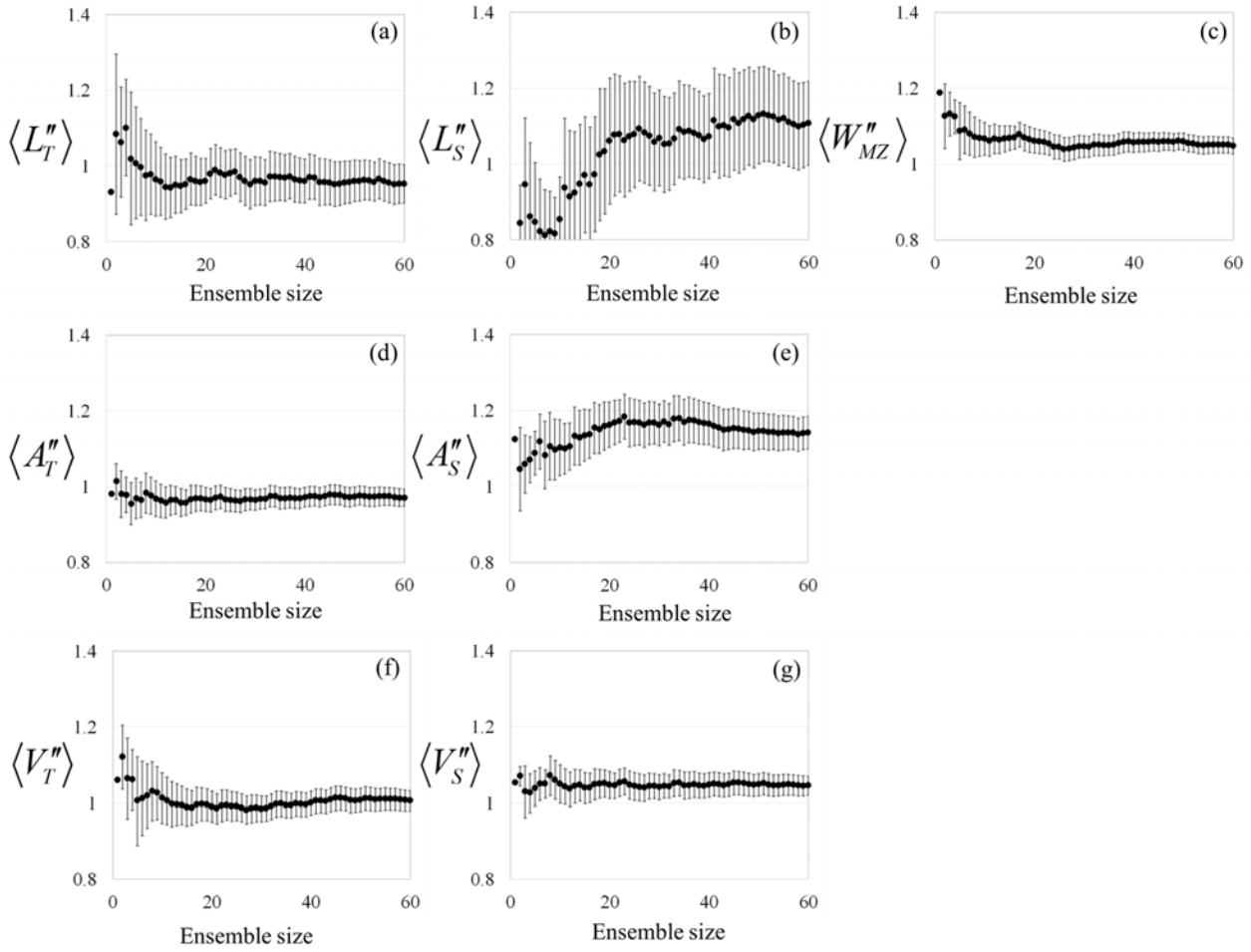

Figure R1. Sample mean of metrics ξ'' versus the ensemble size. The associated 95% confidence intervals are also shown.

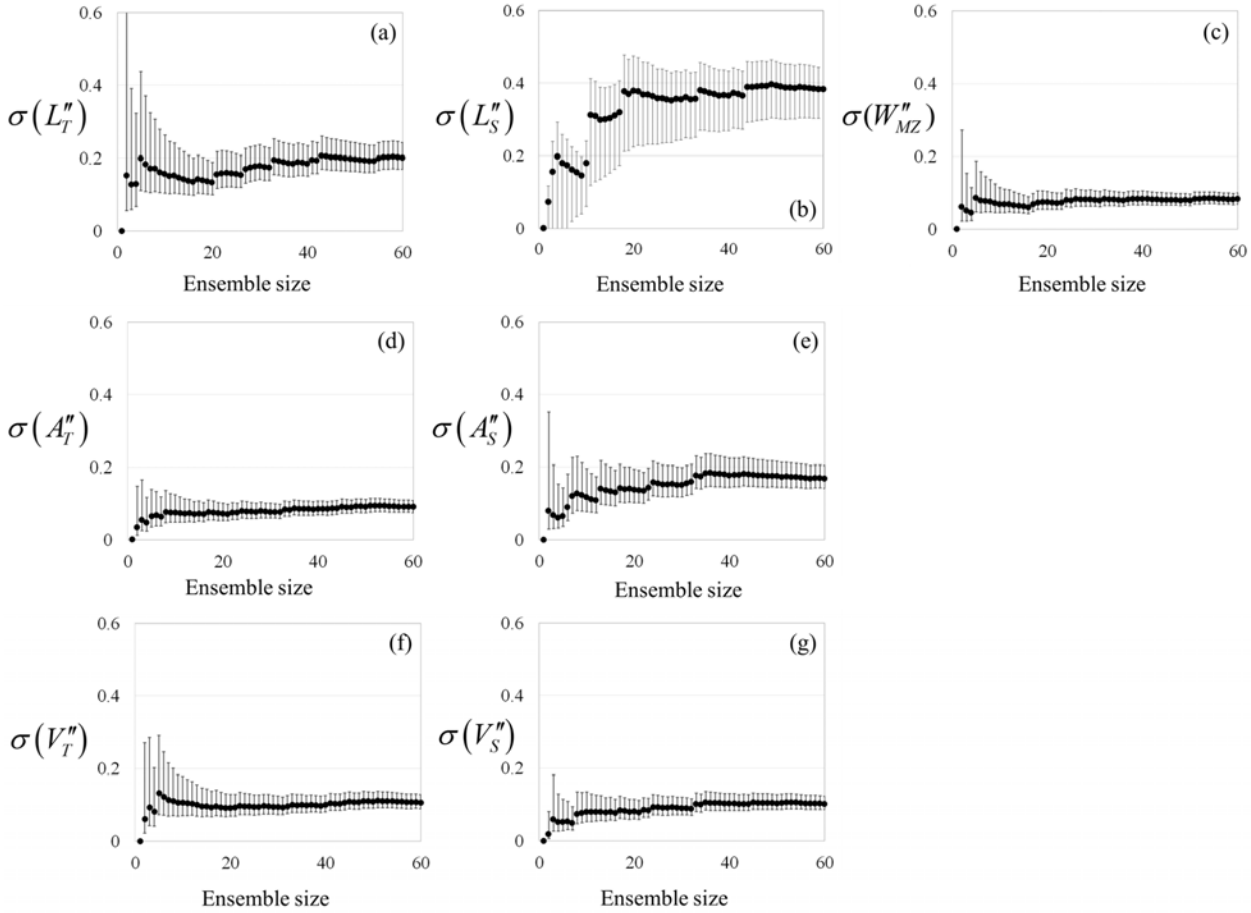

Figure R2. Sample standard deviation of metrics ξ'' versus the ensemble size. The associated 95% confidence intervals are also shown.

> *2. Most of the conclusions are not fully supported by the results, as only one aquifer and one heterogeneous configuration have been considered (e.g., the first point: 'heterogeneous aquifer systems are characterized by toe penetration and extent of the mixing zone that are respectively smaller and larger than their counterparts…'. An analysis of the variability of the considered metrics with respect different configurations of the permeability random field (e.g. large/small variance and large/small correlation length) would better support the proposed general conclusions. Otherwise, the conclusions should be revised referring only to the case studied.*

The present work aims at investigating the effect of random heterogeneity on a three-dimensional domain patterned after a real aquifer. As discussed in the manuscript, only a few contributions studying SWI within a stochastic framework have been published to date. Amongst these, the studies most relevant to our work have been listed in the Introduction of the manuscript. It has to be noted that the vast majority of these works consider idealized synthetic showcases and/or simplified systems (typically in a two-dimensional context) and/or simple flow conditions (usually steady state mean uniform flow). To the best of our knowledge, there are only two contributions where a probabilistic approach is employed to analyze the transient behavior of a real (three-dimensional) costal aquifer: (*a*) Lecca and Cau (2009), and (*b*) Kerrou et al. (2013), respectively targeting the Oristano (Italy) and the Korba aquifer (Tunisia). Key elements of novelty of our manuscript with respect to these works include the introduction and the detailed analysis of an original set of metrics, aimed at characterizing quantitatively the effects of heterogeneity on the extent of seawater wedge penetration and of the seawater/freshwater mixing zone.

These metrics yield a quantitative depiction of SWI in a global sense across a three-dimensional system (not only at the bottom of the aquifer and/or along the vertical direction, as is usually done in the literature).
With reference to the type of random heterogeneity analyzed, we note that the variogram sill we consider represents a domain with moderate variability. As we state in the manuscript, the value of the correlation scale has been selected consistent with documented analyses according to which the integral scale of log conductivity and transmissivity values inferred worldwide using traditional (such as exponential and spherical) variograms tends to increase with the length scale of the sampling window at a rate of about 1/10 (Gelhar, 1993; Neuman et al., 2008). We concur that a systematic analysis of the influence of variogram shape and variogram parameter values would be of interest and will be the subject of a future study.
We stress that ours is one of the first attempts at including the effect of random heterogeneity within a three-dimensional, transient density-variable system. In this context, our results can be considered as exemplary for the type of representative field conditions we analyze. We will revise the conclusions highlighting the novel elements of our study.

> *3. By considering only three pumping schemes, I find hard to conclude that the position proposed in S3 is the best. How did the authors select the position of the well in S3? Is it possible to select the position in such a way to minimise the considered metrics (e.g. for one configuration of the random permeability)?*

The three pumping schemes have been selected to investigate the effects of the distance of the wellbore from the coastline and from the freshwater-saltwater mixing zone on SWI. In this context, in S2 and S3 we also analyzed the impact of an additional pumping rate of seawater (at the bottom of the aquifer). It has to be noted that in this work we evaluate for the first time the effectiveness of the simultaneous extraction of fresh- and seawater in limiting SWI intrusion within a three-dimensional random heterogeneous aquifer. We are aware that our study does not cover the totality of feasible combinations of pumping scenarios. The analysis proposed by Reviewer to identify the optimal well location, albeit of interest, is beyond the scope of the present work and could constitute by itself the topic of a future study. At the same time, we are convinced that such a study should be performed within a stochastic framework (not in a single realization context), thus requiring a remarkable (and possibly prohibitive, in case one would also consider multiple variogram parameters and functional shapes) CPU time.

> *4. Page 6, line 19: I was not able to find the reference Almagro Landò et al. (2010). Please, report in the manuscript the details about the recharge and the head in the inland. It should be stated that these boundary conditions as well as the assumption of a fully saturated domain play a fundamental role in the determination of the SWI.*

The area of interest is characterized by 5 recharge zones (see Fig. 1 of the manuscript) specified on the basis of the land use (inferred from the SIGPAC2005 dataset). Rodriguez Fernandez (2015) provides calibrated values of the recharge associated with each zone. The total recharge slightly varies in time (on a monthly basis), with mean value equal to about 7.6 l/s. Head values at the inland boundary are inferred by interpolating the time-dependent hydraulic-head distribution taken from the two-dimensional model of Rodriguez Fernandez (2015). It has to be noted that iso-potential curves are approximately parallel to the coast in the region of interest. Therefore, the head values set along the inland boundary are approximately constant in time. The average value of hydraulic head at the inland boundary over the 8-year period is $h = 2.4$ m.

5. Section 2.3: which are the initial conditions for the flow and concentration equations? Section 3.1: during the 8 years of the simulation, has the recharge any impact on the SWI? Is the solution after 8 years independent from the choice of the initial conditions?

We set $h = 0$ as initial condition. Adopting $h = 2.4$ m (equal to the mean value of h set along the inland boundary) did not lead to significantly diverse results at the end of the 8-year time period in the homogeneous system. As it is commonly done in the literature, (e.g., Bear et al., 2001; Koussis et al., 2002; Jakovovic et al., 2016) we set initially $C = C_F = 0$. The impact of recharge on SWI has not been investigated. However, due to the limited recharge in the investigated area, its effect appears to be negligible. This can be inferred, for example, from Fig. 3 of the manuscript. The isolines $C/C_s = 0.5$ for the homogeneous system (red curves) do not change appreciably amongst different cross sections along the coast (characterized by diverse recharges).

6. Page 8, lines 18-29: these metrics should be presented in the 'Materials and Methods' section. A table summarizing the meaning of the seven metrics could be of great help to better follow the results. Section 3.2: the description of the four pumping schemes (S0-S3) should be presented in the 'Materials and Methods' section. Page 13, line 11: replace 'associated a' with 'associated with a'. Figure 2: please indicate the depth of the left and right boundaries in panels (b) and (c). Figure 1: Could you provide a small map of Spain indicating where is the Argentona aquifer? It would also help to delineate the boundary of the model grid in panel (a).

We agree with the Reviewer's suggestions and we will implement them in the revised manuscript.

7. Figure 4: the variability of the considered metrics with respect to the single random realisations is not of interest, as it is already expressed in the confidence interval associated to the ensemble mean. It would be more interesting to see their sensitivity to different parameters describing the spatial correlation of the permeability (e.g., short vs long correlation length, high vs low variance).

We prefer plotting not only the ensemble results but also the single realization outcomes. Please see also our answer to item 2.

8. Figure 10: the vertical bars representing the 95 % confidence interval should be much wider. Why the authors divided the standard deviation by the square root of n (page 12, line 12)? This operation should already be done in the computation of the standard deviation. Please, check the result and correct the figure.

Vertical bars in Figure 10 represent an estimate of the error in the evaluation of the sample mean $\langle C_T \rangle / C_S$, due to the finite value n of Monte Carlo (MC) simulations. The error in the evaluation of the sample mean scales with $1/\sqrt{n}$. Assuming that C_T has a Normal distribution, we can write (see e.g., Ballio and Guadagnini 2004)

$$Pr\left[\overline{\Re}_n - t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_n}{\sqrt{n}} \le \mu \le \overline{\Re}_n + t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_n}{\sqrt{n}}\right] = 1-\alpha \quad \text{(R1)}$$

where μ is the ensemble mean of C_T/C_S, $\overline{\Re}_n$ is the sample mean (computed on the basis of n realizations, denoted $\langle C_T \rangle / C_S$ in the manuscript), S_n is the sample standard deviation, t_{n-1} is the Student distribution with (n-1) degree of freedom and $1-\alpha$ is the probability that μ lies within the confidence intervals around the sample mean $\overline{\Re}_n$. When n is large (about $n > 30$), t_{n-1} can be approximated by a

standard normal distribution. Therefore, setting $\alpha = 0.05$, the 95% confidence intervals are given by $\overline{\Re}_n - 1.96\frac{S_n}{\sqrt{n}} \le \mu \le \overline{\Re}_n + 1.96\frac{S_n}{\sqrt{n}}$.

9. References: Almagro Landò et al. (2010): is this document public? This document is cited several times along the manuscript, but it seems to be not available online. Could the author upload this report?

The references concerning the preliminary model of the Argentona basin will be made available if needed.

References: (note that only the references not already included in the manuscript are listed)

Ballio, F., and Guadagnini, A., Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology, Water Resour. Res., 40, 4, W04603, doi: 10.1029/2003WR002876, 2004.

Bear, J., Zhou, Q., Bensabat, J.: Three dimensional simulation of seawater intrusion in heterogeneous aquifers, with application to the coastal aquifer of Israel. First International Conference on Saltwater Intrusion and Coastal Aquifers – Monitoring, Modelling and Management, April 23–25, Essaouira, Morocco, 2001.

Kerrou, J., Renard, P., Cornaton, F. and Perrochet, P., Stochastic forecasts of seawater intrusion towards sustainable groundwater management: application to the Korba aquifer (Tunisia), Hydrogeology J., 21, 425-440, doi: 10.1007/s10040-012-0911-x, 2013.

Jakovovic, D., Werner, A. D., de Louw, P. G. B., Post, V. E. A., Morgan, L. K., Saltwater upconing zone of influence, Adv. Water Resour., 94, 75-86, doi: 10.1016/j.advwatres.2016.05.003, 2016.

Lecca, G., and Cau, P., Using a Monte Carlo approach to evaluate seawater intrusion in the Oristano coastal aquifer: A case study from the AQUAGRID collaborative computing platform, Phys. Chem. Earth, 34, 654-661, 2009.