General concern about our use of uncorrelated conductivity fields:

Since both referees commented on this, we are responding to them together.

The reviewers correctly point out that hydraulic conductivity is a random field; furthermore, it is correlated at some scales. Indeed, the effect of small-scale correlations on regional velocities and mass transport is completely understood at least through second-order (Gelhar and Axness, 1983; Winter et al., 1984; Neuman and Orr, 1992), and there is no doubt that scaled up regional velocities can be affected by the degree of correlation between conductivities at small scales, even when the conductivity field is heterogeneous at local and regional scales (Winter and Tartakovsky, 2002). But the degree of correlation, and hence its effect on regional velocities and quantities like stream-aquifer exchanges that are influenced by them, depends on the meaning of "small". The small scale of our study is set by the 1 km x 1 km size of our grid cells. Many alluvial systems have correlation lengths that are much smaller than that, e.g., Rehfeldt et al. (1992) and Riva et al. (2006), so at the "small" scale of our study, which is typical of simulations used to support regional watershed management, conductivities are effectively uncorrelated in some real settings. Hence, it is informative, and in some cases realistic, to investigate the effect that a field of independent, identically distributed conductivities has on the state variables of regional models. Our experiments show that stream-aquifer exchanges estimated by a typical regional simulation (one that ignores local heterogeneity) can be biased even when local conductivities are uncorrelated. We agree with the reviewers that more remains to be done with regard to the study of locally correlated fields, and we (and we hope others) are at work on such studies. But the work must start somewhere, and it seems sensible to start with the maximum entropy model: one like ours that makes the fewest assumptions about the structure of conductivity, yet is relevant to watershed management. We hope the reviewers and editor will agree that such a model provides a baseline for additional studies and is useful in its own right. If the reviewers and editor believe it will help, we will be glad to put a discussion to this effect in our paper.

Responses to Referee #2:

Thank you very much for your comments and suggestions. Based on them, we revised the paper carefully, and we respond to each of your comments in the following.

General comments

(1) The Authors simulate the regional exchanges of water between stream and aquifer systems using a two-dimensional (single-layer) flow model. And they assume the streambed conductance the same for all simulations. The results of the mathematical model could be a little more realistic respect to the Middle Heihe River Basin (MHRB) and they cannot contribute to the management of water resources in this area.

- We used a two-dimensional (single-layer) flow model to simulate the regional exchanges of water between stream and aquifer systems because that is standard practice in watershed management and it is the one that we wish to test for bias, etc. in this study; in that sense, we are using the MHRB as a convenient example. With regard to the Heihe River Basin specifically, Hu et al. (2007) state that vertical flow is significant in the study area, but water exchanges are more influenced by the connected stream and aquifer system, which means we can treat all layers as a whole. It is also questionable how many layers truly exist. Since our study focuses on comparing the effects of heterogeneous conductivity (or transmissivity, if that is preferred) to a typical two-dimensional model used in many applications of groundwater modeling to water resources management, we have used a single layer.

Additionally, several studies indicate that aquifer heterogeneity influences water exchanges more strongly than do variations in the streambed (Woessner, 2000; Fleckenstein et al., 2006; Kalbus et al., 2009). Our focus on sensitivity to the effects of locally heterogeneous hydraulic conductivity, led us to hold streambed conductance constant across realizations. Since the locally heterogeneity of streambed conductance is affected by many factors (landform, river morphology, sediment, et al.) in different time and space, its effects are much more complicated than those of aquifer heterogeneous conductivity and will need further study. For now, it seems reasonable to stick to experiments with one control variable, the level of heterogeneity in the conductivity field.

The results of the mathematical model can reveal these non-linear effects and contribute to the understanding of aquifer heterogeneity effects on stream-aquifer water exchanges. The calibration results showed the numerical model for the MHRB could simulate the groundwater movement and water cycle relatively satisfactorily, and it reflects realistic conditions in the MHRB. Since stochastic computational tools are more and more capable of approximating the range of dynamics of basin-scale hydrologic systems, we believe the results will also contribute to the management of water resources in this area.

(2) The assumption of random hydraulic conductivities uncorrelated in space is questionable, and a great CV could be not so consistent with the heterogeneity patterns of the aquifer.

- Our discussion of the relevance of uncorrelated random conductivity fields is given above.
- The Coefficient of Variance (CV) quantifies the level of aquifer conductivity heterogeneity (refer to Kalbus et al. (2009), for example). Levels of CV = 2, and indeed much higher, are frequently found in groundwater studies, and indeed can be a serious limitation for approximations based on perturbation expansions when the perturbation parameter depends on the variance of the conductivity field (cf., Zhang, 1998).

(**3) They assume, among the others, that i) small-scale heterogeneities of hydraulic conductivity significantly affect simulated stream–aquifer water exchanges in river basins and ii) systematic biases arise in estimates of exchanges if small-scale heterogeneities are smoothed by aggregation into a few sub-regions. With these assumptions they try to prove, by means of computational experiments, that the biases result from slow-paths in groundwater flow that emerge due to small-scale heterogeneities.**

We are sorry to have been unclear: First, the points identified above as (i) and (ii) are hypotheses (called H1 and H2 in our paper) that we test through our computational experiments. They are important in themselves, since they add weight to the practical conclusion that watershed management decisions based on typical deterministic models, e.g., the base case of our paper, may over-estimate exchanges of water between aquifers and streams, especially aquifer discharge. They also motivate hypothesis H3, that local heterogeneity of conductivity produces slow paths in the local flow field. We do not mean to imply that our experiments "prove" anything, since it is not possible to do that with a few experiments; only the weight of evidence produced by many experiments can do anything like that. We actually make a comment similar to this in our concluding remarks, but we will be happy to amplify that, and other such discussions in our paper, if the editor and reviewer believe that is needed.

Specific Comments

(1) “The zones were defined in previous hydrogeological studies of the MHRB

(Hu et al., 2007): but in their work, Hu et al. divide the aquifer in 8 layers. Could be, please, more precise?

- Line 6 on page 4: There were 8 layers in their work of Hu et al., and we adopt the thickest layer (Layer Six), which can most represent the aquifer properties in space. This layer in their work also had relatively more zones than other layers.

(2) In section 2.2 the author state “All stream–aquifer interactions are simulated using the numerical modeling tool MODFLOW with the stream package (STR) for one-dimensional stream flow and two dimensional groundwater flow”: I suppose that is true for this study and not for all. And after “…… This is an acceptable assumption for typical alluvial sediments of the kind found in the MHRB (Spanoudaki et al., 2009; Huang, 2012)”: this assumption depends on the objective study too!

- Line 12 on page 8: Thanks for your idea. It is true for this study and not for all. This assumption also depends on our objective study, so we corrected it in the manuscript.

(3) I suppose that seepage was calculated from the product of the head difference times the streambed conductance.

- Line 23 on page 8: Thanks for your suggestion. That is right, and we corrected this in the revised manuscript.

(4) “Stream inflow at the YLX Gauge and groundwater lateral recharges from mountain areas are used as an upper boundary, and outflow at the ZYX Gauge is taken as a lower boundary (Zhou et al., 2011)”: Explain as these flowrates were estimated.

- The stream inflow at the YLX Gauge (15.8×m^3/a) and out flow at the ZYX Gauge (9.5×m^3/a) are collected from the observed runoff data. In their study, there were lateral recharges from mountain areas, which can be obtained from the total water resources quantity and surface river inflow amounts.

(5) Figure 2c), is the legend correct?

- Yes, the legend is correct. The horizontal axis represents the distance from the upper boundary (i.e. YLX) along the river. The left vertical axis represents the river stage changes in December, compared with that in June. The right vertical axis represents the groundwater table changes for each corresponding cell along the river. We added a bit more explanation to make this more explicit.

References

Fleckenstein, J. H., Niswonger, R. G., & Fogg, G. E. (2006). River-aquifer interactions, geologic heterogeneity, and low-flow management. Ground Water, 44(6), 837-852.

Gelhar, L. W. and Axness, C. L. (1983) 3-Dimensional stochastic analysis of macrodispersion in aquifers. Water Resources Research, 19 (1), 161-180.

Kalbus, E., Schmidt, C., Molson, J. W., Reinstorf, F., & Schirmer, M. (2009). Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge. Hydrology and Earth System Sciences, 13(1), 69-77.

Neuman, S.P.and Orr, S. (1993) Nonuniform geologic media by conditional moments' exact nonlocal formalism, effective conductivities, and weak approximation. Water Resources Research, 29 (2), 341-364.

Rehfeldt, K. J., Boggs, J. M., Gelhar, L. W. (1992), Field Study of Dispersion in a Heterogeneous Aquifer 3. Geostatistical Analysis of Hydraulic Conductivity. Water Resources Research, 28 (12), 3309-3324.

Riva, M., Guadagnini, L., Guadagnini, A., Ptak, T., Martac, E. (2006) Probabilistic study of well capture zones distribution at the Lauswiesen field site. Journal of Contaminant Hydrology 88 (2006) 92–118.

Winter, C. L., Newman, C.M., Neuman, S. P. (1984) A perturbation expansion for diffusion in a random velocity field. SIAM Journal of Applied Mathematics, 44(2), 411-424.

Winter, C.L. and Tartakovsky, D.M. (2002) Groundwater flow in heterogeneous composite aquifers. Water Resources Research, 38 (8), 23.1-23.11.

Woessner, W. W. (2000). Stream and fluvial plain ground water interactions: rescaling hydrogeologic thought. Ground Water, 38(3), 423-429.

Zhang, D. (1998). Numerical solutions to statistical moment equations of groundwater flow in nonstationary, bounded, heterogeneous media. Water Resources Research, 34 (3), 529-538.