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 #1:

Thank you very much for your comments. We paid careful attention to your concerns and respond to each of them as follows.

(1) Main concern: Justify the assumption of uncorrelated conductivity fields.

- This is clearly a critical issue. Please see above.

We added a paragraph to response the comments about why it makes sense to study the influence of uncorrelated fields.

(2) Line 25 on Page 5: Does “second” follow “First” in Line 19 of Page 5?

- Yes, “second” in Line 26 of Page 5 follow “First” in Line 18 of Page 5. We put them into the same paragraph in the reversed manuscript.

(3) Line 20 on Page 12: Is this conclusion based on Figure 2c?

- This conclusion is based on Fig. 2a and 2b, so we added reference here in the revised version.

(4) It is better to add a dashed zero-line on Figure 2c so that the differences of groundwater table below (or above) 0 can be explicitly shown.

- There are two vertical axes in Fig. 2c, and their scales are different. The groundwater table changes are much larger than river stage changes, so it may not be necessary to add a dashed zero-line. We will follow the advice of the editor and reviewer regarding modifying this figure, but we have left it the same for now

(5) Figure 3b is a little hard to read. It is suggested to use gray color for water table depth and improve the contrast with monthly leakage data.

- Using gray color for water table depth is better to improve the contrast with monthly leakage data, but it is not so easy to see the distinct groundwater table changes. As we tried to demonstrate that the groundwater table distribution affects the groundwater flow rate distribution and stream-aquifer exchanges in the later text, it seems better to highlight the colorful water table depth. We will follow the editor's and reviewer's further advice on this point.

(6) Line 2 on page 15: Is this confidence interval computed based on 10 realizations generated using equation (1) and (2)? Is the realization number sufficient to make a reasonable statistics?

- Yes, this confidence interval, computed based on 10 realizations, is generated using equations (1) and (2). We have compared the different realization results, and found they are almost the same with more realizations. So 10 realizations seem sufficient to make a reasonable statistics. We have added a statement 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. We added a paragraph to response the comments about why it makes sense to study the influence of uncorrelated fields.
- 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) Line 6 on page 4: "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?

- 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) Line 21 on page 8: I suppose that seepage was calculated from the product of the head difference times the streambed conductance.

- Thanks for your suggestion. That is right, and we corrected this in the revised manuscript.

(4) Line 12 on page 9: "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$\times$m^3/a) and out flow at the ZYX Gauge (9.5$\times$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.

author's changes in manuscript：

Nonlinear effects of locally heterogeneous hydraulic conductivity fields on regional stream-aquifer exchanges

Jinfeng Zhu[1], C Larrabee Winter[2,*], Zhongjing Wang[1,3,*]

[1] Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China

[2] Department of Hydrology and Water Resources and Program in Applied Mathematics, University of Arizona, Tucson, AZ, 85721, USA

[3] State Key Laboratory of Hydro-Science and Engineering, Tsinghua University, Beijing, 100084, China

*Corresponding author information:

Email: winter@email.arizona.edu

Tele: +1 520 621 7120

Email: zj.wang@mail.tsinghua.edu.cn

Tele: +86 10 62782021

Manuscript submitted to *Hydrology and Earth System Sciences*

2015.07.09

Abstract

Computational experiments are performed to evaluate the effects of locally heterogeneous conductivity fields on regional exchanges of water between stream and aquifer systems in the Middle Heihe River Basin (MHRB) of northwestern China. The effects are found to be nonlinear in the sense that simulated discharges from aquifers to streams are systematically lower than discharges produced by a base model parameterized with relatively coarse effective conductivity. A similar, but weaker, effect is observed for stream leakage. The study is organized around three hypotheses: (H1) small-scale spatial variations of conductivity significantly affect regional exchanges of water between streams and aquifers in river basins, (H2) aggregating small-scale heterogeneities into regional effective parameters systematically biases estimates of stream-aquifer exchanges, and (H3) the biases result from slow-paths in groundwater flow that emerge due to small-scale heterogeneities. The hypotheses are evaluated by comparing stream-aquifer fluxes produced by the base model to fluxes simulated using realizations of the MHRB characterized by local (grid-scale) heterogeneity. Levels of local heterogeneity are manipulated as control variables by adjusting coefficients of variation. All models are implemented using the MODFLOW simulation environment, and the PEST tool is used to calibrate effective conductivities defined over 16 zones within the MHRB. The effective parameters are also used as expected values to develop log-normally distributed conductivity (K) fields on local grid scales. Stream-aquifer exchanges are simulated with K fields at both scales and then compared. Results show that the effects of small-scale heterogeneities significantly influence exchanges with simulations based on local-scale heterogeneities always producing discharges that are less than those produced by the base model. Although aquifer heterogeneities are uncorrelated at local scales, they appear to induce coherent slow-paths in groundwater fluxes that in turn reduce aquifer-stream exchanges. Since surface water-groundwater exchanges are critical hydrologic processes in basin-scale water budgets, these results also have implications for water resources management.

1. Introduction

Exchanges of water between streams and aquifers are critical elements in the coupled dynamics of watersheds. Groundwater discharge to streams maintains natural hydrologic systems through base flow during periods of low stream flow, while leakage from streams is an important source of groundwater recharge (Sophocleous et al., 1995; Hantush, 2005; Newman et al., 2006). The magnitudes of such fluxes vary on scales ranging from sub-regional (or zonal) to local scales. Local variations in the spatial distribution of aquifer characteristics are known to affect groundwater flow and stream-aquifer exchanges on sub-regional scales (Schmidt et al. 2006; Kalbus et al, 2009; Mendoza et al., 2015), but it is not usually possible to measure system parameters and states with sufficient accuracy and level of detail to specify local variations throughout a watershed (Wroblicky et al., 1998; Kalbus et al., 2006). Furthermore, hydrologic system parameters can vary on multiple spatial and temporal scales and are subject to experimental error (Molz, 2000; Genereux et al., 2008).

Computational experiments, on the other hand, allow the effects of heterogeneity to be investigated by consistently varying the values of parameters, such as hydraulic conductivity, as control variables and observing the resulting effects on simulated system states (Winter et al., 2004; Bruen and Osman, 2004; Hantush, 2005). In this study the sensitivity of regional hydrologic systems to locally heterogeneous aquifer hydraulic conductivity is explored by simulating stream-aquifer exchanges in the Middle Heihe River Basin (MHRB) of northwestern China, a typical semi-arid basin. Fluxes in alluvial aquifers of the MHRB are usually represented as two-dimensional processes (Huang, 2012).

In addition to their importance for water resources management, stream-aquifer exchanges are a convenient measure of overall watershed performance because they summarize the states of system fluxes at well-defined locations where they are relatively easy to quantify. Two scales of heterogeneity are represented in the simulations, zonal and local (or grid-scale). Local-scale conductivity is manipulated as the only control variable in the computational experiments. All other system parameters, including streambed conductance, are the same in every simulation.

Multi-physics models of the coupled hydrologic fluxes of the MHRB are implemented using the MODFLOW simulation environment (McDonald and Harbaugh, 2003). The PEST tool (Doherty et al., 1994) is used to calibrate an effective aquifer hydraulic conductivity for each of sixteen zones within the MHRB. The zones were defined in previous hydrogeological studies of the MHRB (Hu et al., 2007). The effective zonal parameters are also used as expected values to develop realizations of log-normally distributed conductivity fields on local grid scales. Standard deviations are defined by means of coefficients of variation (CV) ranging from 0.5 to 2.0. Ten realizations of a random conductivity field are produced for each CV. Random conductivity values are uncorrelated in space. Monthly stream-aquifer exchanges are calculated for each realization and for the zonally parameterized model.

The study is organized in terms of three explicit hypotheses about the effects of local-scale heterogeneity:

H1. Local-scale heterogeneities significantly affect regional stream-aquifer water exchanges in the MHRB.

H2. Systematic biases arise in estimates of stream-aquifer exchanges if local-scale heterogeneities are smoothed by aggregating conductivity into effective values for the set of 16 sub-regional zones.

H3. The biases are the result of slow-paths in groundwater flow that emerge due to small-scale heterogeneities in conductivity fields.

The hypotheses H1-3 are tested by comparing simulated stream-aquifer exchanges produced by the local realizations to the zonal simulation.

Hypotheses related to H1 and H2 have been investigated in a few recent studies. Kalbus et al. (2009) evaluated the relative effects of aquifer conductivity and streambed conductance by simulating stream-aquifer interactions in a 220m reach of a small stream in Germany. They found that heterogeneity in the aquifer influenced discharge to the stream more strongly than did variations in the streambed itself.

Schmidt et al. (2006) also studied the spatial distribution and magnitude of groundwater discharge to a stream with a simple analytical model that was mainly focused on groundwater discharge at the reach scale. Kurtz et al. (2013) generated multiple realizations of stream-aquifer interactions in the Limmat aquifer system in Zurich Switzerland. They allowed riverbed hydraulic conductivities to take one of four different levels of heterogeneity ranging from local variability at each grid-point to effective conductances of only 5, 3, and 2 values. They found that effective conductance did not always reproduce fluxes obtained from base simulations where system parameters were perfectly known, and furthermore, simulations based on effective parameters gave biased estimates of net exchanges between aquifer and stream. Mendoza et al. (2015) used a land surface model with strong a priori constraints on parameters to argue that such models can perform poorly when spatial variability and hydrologic connectivity of a region are represented coarsely. Lackey et al. (2015) used a synthetic stream-aquifer system to show that modeling streambed conductance as a homogeneous property can lead to errors in estimated stream depletion.

The current study contributes to the understanding of these topics in two ways. First, results indicate that locally heterogeneous hydraulic conductivity fields lead to systematic reductions of regional stream-aquifer exchanges, especially in the direction of aquifer-stream discharges. The causal mechanism appears to be slow paths that emerge in groundwater flows in response to local heterogeneities. These local-scale fluxes are spatially coherent and are not averaged out by aggregating conductivity into a few effective zonal parameters. The nonlinear effects of these coherent structures are propagated to regional scales. Second, hypotheses H1-3 have practical implications for allocations of water resources since many resource management decisions depend on the results of computational models like the base model used here (Perkins and Sophocleous, 1999; Fleckenstein et al., 2006). A regional basin is partitioned into a few zones, on the basis of physical, geographical and geological criteria (Eaton, 2006), and effective zonal parameters are set by calibration

批注 [Jinfeng1]: Following Referee#1's concern, we merged this two paragraphs.

(Christensen et al., 1998). Since effectively parameterized zonal models cannot produce the nonlinearities observed here, the base model is found to systematically overestimate regional stream-aquifer exchanges. This holds true for all levels of heterogeneity investigated here as controlled by CV.

The paper is organized as follows. In Section 2, the experimental framework used to model the MHRB and the stochastic methods used to produce locally heterogeneous hydraulic conductivity fields are described. Section 3 reports methods and results of investigations of H1 and H2. It presents results comparing simulations of stream-aquifer exchanges derived from (1) a base model whose effective conductivity parameters are specified by zone and (2) stochastic realizations of heterogeneous local-scale conductivity. Section 4 focuses on H3. The emergence of coherent slow paths in the flow fields is investigated as the source of the bias in base model estimates. Results are summarized and discussed in Section 5, and a few directions for future experiments are sketched.

2. Experiment setting and model development

The experimental approach is to compare alternative numerical models of hydrologic system dynamics that produce stream-aquifer exchanges in the Middle Heihe River Basin (MHRB), a closed basin in northwestern China of area of 8778 km^2. Stream-aquifer exchanges are estimated by (1) a so-called base model that specifies hydraulic conductivity on zonal scales and (2) a set of related models that incorporate local-scale heterogeneity in their parameterizations. Zonal conductivity in the base model is calibrated using the PEST parameter estimation tool (Doherty et al., 1994; Sophocleous et al., 1999), and the base model assigns the same calibrated value of hydraulic conductivity to every grid point in a given zone.

The level of heterogeneity of hydraulic conductivity is the experimental control variable in local-scale simulations. It is set by adding varying levels of randomness to the zonal values of the base model. The result is a random field of conductivities with expected values that are the same in a given zone as the base model's effective conductivity. Otherwise, system parameters including specific yield and streambed conductance and boundary conditions are the same for all simulations. All computations are conducted using the MODFLOW simulation environment (Leake and Prudic, 1991; McDonald and Harbaugh, 2003; Rodríguez et al., 2006).

2.1. Study area

The MHRB is contained in the Heihe River Basin, the second largest closed river basin in China, which drains a total area of 142.9 thousand km^2. The MHRB is in a semi-arid region with average annual rainfall of approximately 50-200mm/y, most of which occurs from May to October. The river originates in the Qilian Mountains of Qinghai Province and then flows through Gansu Province to western Inner Mongolia. The average annual inflow from mountain areas upstream is 1.58 billion m^3 as measured at the Yingluoxia Gauge (YLX), and the discharge through the Zhengyixia Gauge (ZYX) is 0.95 billion m^3. The 184 km long reach from YLX to ZYX defines the MHRB study area (Fig. 1), which is known to be influenced by temporal patterns and spatial distributions of stream-aquifer exchanges (Zhou et al., 2011).

Intensive stream seepage and groundwater discharge occur in the study area along the river in fluvial and alluvial fans. The exchanges proceed in two directions: groundwater flows through the streambed into the stream (discharging gaining stream), and stream water infiltrates through the streambed into the groundwater system (leaking losing stream) which is the same as general states presented in (Winter, 1995; Kalbus et al., 2006). Groundwater discharge to streams maintains natural hydrologic systems during periods of low stream flow through base flow (Sophocleous et al., 1995; Hantush, 2005). Streams are an important source of groundwater recharge at higher elevations. Conjunctive use of surface and subsurface waters is a potentially important resource management tool in semi-arid regions like the Heihe River Basin whose potential also depends on system state estimates.

The Middle Heihe River leaks into a sandy fluvial area immediately after YLX and then enters an alluvial plain composed of fine soil. At that point substantial stream seepage, diversion, and groundwater discharge (base flow) occur along the stream. The Middle Heihe River is a wide and shallow stream, and the area of the stream channel is much greater than the lateral areas of both banks. Thus the exchange of stream water and groundwater occurs primarily in the vertical direction. The plain of fine-grained soil along the river is the main area where springs and groundwater emerge. Thirty irrigation areas are distributed throughout the study area. Groundwater tables range from 0m to 300m beneath the land surface, and aquifer thickness varies from tens to a couple of hundred meters.

2.2. Numerical model of stream-aquifer exchanges

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 in this study. The stream package (Leake and Prudic, 1991) simulates stream flow with Manning's equation and interactions with groundwater flow using Richard's equation, which assumes vertical, gravity-driven flow and neglects capillarity. This is an acceptable assumption for the typical alluvial sediments of the kind found in the MHRB (Spanoudaki et al., 2009; Huang, 2012).

批注 [Jinfeng2]: We added this to make this sentence more accurate.

The stream and aquifer systems are coupled in the model through an iterative process in which convergence of state variables linking the two domains is used as the criterion for accepting a solution. The stream stage and groundwater table are iterated at each time step until the differences between two iterations are within a small tolerance. Seepage is calculated from the product of the head difference times a streambed conductance. In regional-scale groundwater models it is reasonable to assume a homogeneous low-conductivity streambed within a heterogeneous aquifer (Kalbus et al., 2009). Conductance is assigned based on an existing calibration and measured data values between 1-3 m/d (He and Zhao, 2007; Zhou et al., 2011).

批注 [Jinfeng3]: We changed "gradient" into "difference" here.

Since aquifer thickness is small compared to its horizontal dimensions, one layer is sufficient for the vertical discretization. Workman et al. (1997) pointed out that the discrete cell size can be 1-2km in regional groundwater flow models, as a river can affect water table elevations as far as 1525 m from the middle of the stream. Thus the study domain is discretized into 155 rows and 172 columns, with each cell size 1km x 1km. The aquifer is simulated as a free-surface boundary able to fluctuate in response to recharge from irrigation fields, evapotranspiration, flow to drains, and interaction with streams. The recharge package (RCH) is used to input rainfall and irrigation seepages as vertical boundaries, and the recharge distribution is set according to the data from Hu et al. (2007). The ETS package is employed to simulate phreatic evaporation considering the multiple-linear relationship with groundwater depth. Based on previous studies (Zhou et al., 2011; Tian et al., 2015), spring flow is also calculated in irrigation areas by representing the drainage area as drain cells.

The groundwater table observed on January 1 of the year 2000 is input and interpolated as initial head for all simulations. The time step is 5 days, and the simulation period is one year. 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). In the mountain front region, the aquifer has a high proportion of pebbles, sand and gravel, while the plain of the lower area is composed of thick unconsolidated alluvial fills with alluvial sand, silt, and clay (Hu et al., 2007).

2.3. Base model

Previous research has identified 16 major hydrologic zones within the MHRB based on variations of hydraulic conductivity and specific yield (Hu et al., 2007; Jia et al., 2009). In this study, the PEST parameter estimation system is used to calibrate aquifer parameters (zonal horizontal conductivity K_z and specific yield S_z) for the different zones, z = 1, ..., 16. The parameterization of conductivity by constant effective values, one each for the 16 coarse zones, is the foundation of the base model. The 16 zonal conductivities are also used as expected values for the stochastic generation of locally refined conductivity fields as described in the next sub-section.

The objective function for PEST estimates is the sum of squared differences between observations and simulations of groundwater tables at 34 wells distributed through the basin. Effective parameters K_z and S_z are optimized and estimated separately for the 16 zones of the domain due to computational limitations. The lower and upper levels of parameter settings for conductivity are 0.5~50 m/d and are 0.01~0.25 for specific yield without considering anisotropy. The procedure of PEST includes an initial MODFLOW simulation, calculation of the Jacobian matrix for each parameter, and different Lambda upgrade. After the calibration of K_z values the specific yields of the system, S_z, are optimized with PEST for transient conditions by minimizing the groundwater table errors at observation wells.

2.4. Stochastic realization of conductivity fields

Stochastic simulations are the basis for local-scale realizations of hydraulic conductivity fields, $K(x)$, used in comparisons. The variable x corresponds to a point in the 155 x 172 computational grid used in all simulations of hydrologic system dynamics in the MHRB. Local parameters $K(x)$ are defined by log-normally distributed perturbations on the zonal conductivity field of the base model,

$$ln\mathrm{K}(x) = \mathrm{Y}(x) = \mathrm{Y}_z + \mathrm{Y}'(x), \quad \mathrm{Y}'(x) \sim N(0, V_z) \text{ for all } x \in \text{zone } z. \qquad (1)$$

The zonal mean, Y_z, and variance, V_z, are defined in terms of the base model parameters, K_z,

$$\mathrm{Y}_z = ln\frac{K_z}{\sqrt{1 + C}}, \ V_z = \mathrm{Var}[\mathrm{Y}(x)] = ln(1 + C), \ x \in \text{zone } z \qquad (2)$$

where C is the Coefficient of Variation (CV).

In order to compare the effects of different levels of total heterogeneity, realizations of conductivity fields are generated corresponding to six values C = 0.1, 0.2, 0.5, 0.8, 1.0, 2.0. Grid-scale analyses at different levels of heterogeneity are based on a sample of 10 realizations of log-normally distributed random conductivity fields for each value of C. Local scale simulations of groundwater flow and stream-aquifer exchanges are produced using realizations of K fields based on Eqns. (1) and (2) in the general model.

Grid point values of conductivity are sampled independently, so no correlation structure is imposed on local conductivities. It is generally accepted that hydraulic conductivity is a random field that is typically correlated over a continuous range of small scales, becoming uncorrelated once a characteristic correlation length is reached (Zhang, 2002; Rubin, 2003). The effect of small-scale correlations on regional velocities and mass transport is understood through second-order for spatially stationary conductivity fields (Gelhar and Axness, 1983; Winter et al., 1984; Neuman and Orr, 1993) and for a wider range of fields that are heterogeneous at local and regional scales (Winter and Tartakovsky, 2002). The small scale of our study is set by the 1 km x 1 km size of our grid cells. Since many alluvial systems have correlation lengths that are much less than that, e.g., Rehfeldt et al. (1992) and Riva et al. (2006), it is informative, and in some cases realistic, to investigate the effect that a field of locally independent, identically distributed conductivities has on the state variables of regional models. Indeed, our experiments show that stream-aquifer exchanges estimated by a typical regional simulation like the base case can be biased even when local conductivities are uncorrelated.

批注 **[Jinfeng4]:** We added a paragraph to response the comments about why it makes sense to study the influence of uncorrelated fields.

2.5. Tests of Hypotheses

批注 **[Jinfeng5]:** This section 2.5 is separated since the topic (hypothesis testing) is different from the previous discussion.

Comparisons between simulations produced by the base model and local-scale models depend on either normalized squared departures between fluxes produced by the base model and the locally heterogeneous models or ratios between the same fluxes. The results of comparisons are said to support H1 and H2 if simulated values produced by the base model exhibit systematic bias with respect to local-scale simulations of stream-aquifer exchanges. H3 is evaluated by comparing normalized squared departures of groundwater fluxes produced by the base model with samples produced by locally resolved simulations.

3. Simulated stream-aquifer exchanges

This section focuses on H1 and H2, the hypotheses that simulations of exchanges of water between the stream and aquifer systems produced using locally heterogeneous conductivity fields are systematically less than estimates produced by the base model. Results are given separately for the base model and for models with parameterizations affected by local scales of heterogeneity and then compared. Levels of CV are manipulated as control parameters in the experiments. Simulated groundwater fluxes are used later (Section 4) to evaluate H3, the hypothesis that locally heterogeneous conductivity affects the emergence of preferential paths in the aquifer.

3.1. Base model calibration and simulation results

The lower fluvial plain of the MHRB is mainly composed of relatively low permeability silts and clay, while the upland aquifer, containing mostly sand mixed with gravel, is more permeable (Hu et al., 2007). This distribution is reflected in estimates of zonal conductivity and specific storage, K_z and S_z, derived from a PEST calibration based on a record of 5-day groundwater table variations in the year 2000 (Fig. 1). The differences between well heights simulated using the PEST calibration and observations yield a residual mean of -0.19 m, standard error of 0.021 m, root mean square (RMS) of 1.4 m, and normalized RMS of 0.77%. The PEST calibration results specify higher conductivity for upstream areas near mountain fronts than for the lower fluvial plain in accord with the characteristic geology of the basin.

Simulated river stages and groundwater tables for June and December along the 183km river distance are displayed in Fig. 2a and Fig. 2b. The groundwater table is generally much lower than the stream stage for about 25 km below Yingluoxia Gauge (YLX), indicating a losing reach of stream, as can be seen from these two figures. The groundwater table becomes quite shallow in the plain downstream of YLX, and the stream may be losing or gaining at different seasons and locations until the outlet at the Zhengyixia Gauge (ZYX).

批注 [Jinfeng6]: We added the reference here.

River stage in June is relatively higher than that in December (Fig. 2c), as the stream flow is higher in flood season. Nevertheless, the groundwater table along the river in December is a bit higher because there is almost no pumping for irrigation in that month. Stream inflow from YLX exhibits clear seasonal variations within the year with a large pulse of water in Spring-Summer due to snowmelt (Fig. 2d). This effect is clearly reflected by monthly stream leakage to groundwater (m^2/km/day). On

批注 [Jinfeng7]: We added this to make it more clear.

the other hand, groundwater discharge to the river (or base flow) is relatively stable and less controlled by the seasonal inflow from station YLX.

The pattern of groundwater flow is not only controlled by the distribution of hydraulic conductivity, but also by the configuration of the water table. Magnitudes of average groundwater flow velocity are higher near the mountain fronts than in the lower plains (Fig. 3a). Average monthly stream leakage rates ($\overline{Q_b}$),where the average is taken over all monthly values, exhibit critical transition points between upwelling and downwelling zones (Fig 3b). The stream is losing in deep groundwater areas near mountain fronts and becomes a gaining stream in the lower plain where groundwater tables are shallow.

The intensity of stream-aquifer interactions also varies seasonally. Interactions in summer and autumn show similar patterns with water exchanges of losing streams and gaining streams stronger than the monthly average. Stream inflow is higher and recharge to groundwater (precipitation and irrigation seepage) is relatively larger in these seasons. Thus, leakage from losing streams and groundwater discharge along gaining streams both become larger than the corresponding seasonal average. The magnitudes of water exchange are more spatially variable in the high-flow seasons of winter and spring: the magnitude augmentation (compared with seasonal average) of water exchanges for losing streams and gaining streams can be as large as 1041 and 317 m^2/km/day respectively. However, the water exchanges of the losing streams and gaining streams become much weaker in spring and winter, as the stream inflow from YLX is relatively low and the groundwater recharge (precipitation and irrigation seepage) is quite small.

3.2. Influence of local heterogeneity on simulated results

Groundwater simulations based on local (grid-scale) simulations of hydraulic conductivity $K(x)$ display more spatially heterogeneous system states than the base model. Realizations of highly heterogeneous $K(x)$ fields are produced by applying the method of Eqns. (1) and (2) in Section 2. Fig. 4 displays grid-scale perturbations of $K(x)$ from the zonal mean K_z for coefficients of variation (CV) of magnitude C =

0.1 and C = 2, as indicated by K(x)/K_z. Higher perturbations from zonal mean conductivity occur for higher CV, which follows from the definition of zonal standard deviation (Eqn. 2).

Grid-scale simulations, $W_G(x)$, of groundwater tables at the end of the simulation period (Dec) are more variable than the base result, $W_B(x)$. Point-wise differences between the two, $\Delta W(x) = W_B(x) - W_G(x)$, are about equally distributed between grid cells where the grid-scale water table is higher than the base case ($\Delta W(x) < 0$) and lower ($\Delta W(x) > 0$), although there is a slightly greater chance of a rise. The number of cells where the absolute change in the water table is large ($|\Delta W(x)| > 1$ m) increases systematically with CV, and the number of cells where $|\Delta W(x)| < 1$ m decreases correspondingly. The point-wise differences exhibit a pattern of compact spatially coherent areas where water tables consistently increase or decrease together. This is despite the lack of any correlation in the values of grid-scale conductivity.

It is convenient to use an indicator,

$$I_m = Q_m / Q_m^{(B)} \tag{3}$$

to compare monthly stream-aquifer exchanges produced by local grid-scale models, Q_m, with monthly exchanges produced by the base model. Here m = 1, ..., 12 indicates the month.

The base model simulation, $Q_m^{(B)}$, of total water exchanged between the stream and aquifer during month m summed over all x = 1, ..., n grid cells along the stream is,

$$Q_m^{(B)} = \sum_{x=1}^{n} Q_m^{(B)}(x) \tag{4}$$

Similarly, Q_m, the mean total exchange for month m averaged across all r = 1, ..., N_r (=10) locally refined models is

$$Q_m^{(r)} = \sum_{x=1}^{n} Q_m^{(r)}(x) \text{ and } Q_m = \frac{1}{N_r} \sum_{r=1}^{N_r} Q_m^{(r)} \quad . \tag{5}$$

The mean monthly value of the aquifer-stream discharge indicator is always less than 1 for all values of CV: the base model produces discharge values that are systematically larger than the locally heterogeneous models (Fig. 5). When two standard deviation confidence intervals are calculated for I_m, a few intervals corresponding to the smallest CVs include 1 for some months. In those cases small-scale heterogeneities increase discharge in some local areas and the overall effect of heterogeneity is reduced. When CV equals 0.1, the monthly aquifer-stream discharge is quite close to the base model value (i.e., the expected value) since perturbations around the expected values are small in that case. On the other hand, mean indicator values decrease to nearly 0.8 as CV increases from 0.1 to 2, and standard deviations continue to increase. As heterogeneity in local models increases, the monthly aquifer-stream discharge indicator decreases and uncertainty as measured by standard deviation increases. The discharge ratio also shows some seasonal patterns, with smaller effects of heterogeneity in high-flow months, such as August and September. Compared with other months, the aquifer-stream discharge quantity is larger and the uncertainty is less in these months.

The mean monthly value of the corresponding stream-aquifer leakage indicator is also less than 1 in most months, and the standard deviations show increasing trends with increasing CVs, but the relative magnitudes of leakage are less than those of discharge and are closer to the base model (Fig. 6). Taking CV equaling to 2 for example, the ranges of mean values for stream-aquifer leakage and discharge are respectively 0.97-0.99 and 0.81-0.93, and the respective ranges of standard deviations for stream-aquifer leakage and discharge are 0.004-0.013 and 0.020-0.044.

The entire Middle Heihe River includes upstream segments where the dominant exchange mechanism is leakage and downstream segments where discharge is dominant. To separate these two confounding classes of fluxes, spatial effects of heterogeneous conductivity are further investigated by calculating indicator values ($Q_m/Q_m^{(B)}$) for four major sub-streams in the MHRB. Results for discharge and leakage are shown for all 10 realizations of conductivity for each value of CV (Figs. 7 and 8). Stream 1 is an upstream segment while Streams 2-4 comprise the main stem of the river in the alluvial plain. The indicator values are almost all less than 1 for average monthly discharge from these four sub-streams (Fig. 7), corresponding to the relatively larger impact of conductivity heterogeneity on groundwater processes leading to discharge. Stream-aquifer leakage is larger than the corresponding base model in every sub-stream during at least some months (Fig. 8).

The stream-aquifer water exchanges of these four sub-streams respond differently to local heterogeneity, reflecting the different groundwater flow regimes of the different sub-regions. For example, Stream 1 is mainly a losing stream located in a mountain front area with a deep groundwater table, thus aquifer heterogeneity has a relatively small effect on stream-aquifer exchanges. Streams 2-4 go through the plain where water exchanges are affected by pumping and recharge schemes, and the differences between Q_m and $Q_m^{(B)}$ are more obvious. Since groundwater depths are shallow and extensive stream-aquifer interactions occur in these areas, the effects of aquifer heterogeneity on water exchanges of these streams are relatively strong. The values of $Q_m/Q_m^{(B)}$ are smaller during summer and autumn, when the stream inflow from YLX and precipitation are relatively large.

4. Slow paths

In this section evidence is presented for the hypothesis (H3) that the systematic bias observed in Section 3 is due to the response of groundwater flow to local heterogeneities in conductivity fields. Spatially coherent slow paths are seen to emerge in groundwater flows simulated in realizations of heterogeneous conductivity fields, and their density increases as the CV of experiments increases. This occurs despite the lack of any correlation in the spatial distribution of local-scale conductivity itself.

4.1. Local scale conductivity field heterogeneity

It is well-known that groundwater system parameters exhibit heterogeneity on multiple scales ranging down to the smallest (Neuman et al., 1987; Neuman, 1990, 1994). Here the statistics of locally heterogeneous conductivity fields are analyzed and compared to the mean fields corresponding to the calibrated conductivity parameters of the base model. A single realization is chosen at random from the 10 available per level of CV, and its statistics are analyzed. Results for other realizations are essentially the same.

The point x is a grid cell in the z^{th} zone. The ratio

$$\rho_z(x) = K(x)/K_z \tag{6}$$

compares $K(x)$, the local conductivity at x to K_z, the expected value of $K(x)$ defined by the calibrated zonal conductivity parameterizing the base model. The summary statistics of $\rho_z(x)$ describe the overall behavior of normalized local conductivity when adjusted for zone (Table 1). The arithmetic means of the $\rho_z(x)$ equal 1 for all values of CV, implying that the conductivity field is indeed a realization drawn from an ensemble of fields whose zonal expected values are the base model values derived from the initial calibration. The standard deviations of the $\rho_z(x)$ increase with CV. Both these behaviors are expected. Since hydraulic conductivity is bounded below by 0, the ranges of minimum values are also limited; nonetheless, minimum values decrease by two orders of magnitude, which is consistent with the increase in CV. Maximum values increase with increasing CV. Calculations of Moran's I (not shown) indicate that values of K are not spatially correlated for any CV.

The geometric mean of $\rho_z(x)$ declines from 0.96 to 0.76 as the heterogeneity of the conductivity field (CV) increases. Since the effective conductivity of a heterogeneous 2D field is equal to its geometric mean (Matheron, 1967; Gutjahr et al., 1978), groundwater flow is expected to be slower in more heterogeneous fields where CV is higher. A related effect is observed in the simulations of aquifer-stream discharge reported in Section 3.2 where both overall discharge and discharge in individual stream segments (Fig. 5 and Fig. 7) decline as CV increases. This is further suggested by the trend in the percentage of grid cells with $K(x)$ values less than K_z (i.e. $\rho_z(x) < 1$) which increase from 54.8% to 64.2% with CV increasing from 0.1 to 2. It should be noted that the median value also shows a decreasing trend from 0.99 to 0.84.

These trends are evident in grid-scale $K(x)$ field comparison maps based on the realization used to construct Table 1 (Fig. 9). Maps for other realizations are similar. Black grid cells are locations where $K(x) < K_z$, while white grid cells have values $K(x) > K_z$. The number of black grids clearly increases with heterogeneity (CV), illustrating the increasing percentage of area occupied by locations with smaller conductivity than expected.

4.2. Emergence of slow paths

To evaluate hypothesis H3, we compare groundwater fluxes $q(x)$ and $q_b(x)$ obtained respectively from (i) the flow simulation based on the same realization of a locally heterogeneous conductivity field, $K(x)$, that was selected in Section 4.1 and (ii) the simulation based on the calibrated base model field, K_z. We adopt the normalized difference

$$\Delta q(x) = \left(q(x) - q_b(x)\right) / \sqrt{q(x) \bullet q_b(x)} \tag{7}$$

to indicate differences between flow simulations. Here $\Delta q(x) < 0$ indicates a point where flow due to local variations of conductivity, $q(x)$, is less than flow due to the base model, $q_b(x)$. The normalization removes large-scale effects due to zonal averages from the analysis.

Summary statistics of flow (Table 2) support the hypothesis that locally heterogeneous conductivity fields induce reduced overall flows even though the heterogeneous conductivities are spatially independent. The arithmetic mean difference drops from -0.05 to -0.51 as CV increases from 0.1 to 2, and the median and minimum values exhibit a declining trend as well, implying that groundwater flow estimated by the local-scale model is slower at most points than the groundwater flow estimates from the base model. Meanwhile, standard deviations steadily increase from 0.175 to 0.833, indicating that groundwater flow becomes more variable as CV increases. Additionally the percent of area that is occupied by relatively lower groundwater flow (q(x) < q_b(x)) increases with increasing heterogeneity.

This fact is amplified by examining the spatial distribution of normalized differences Δq(x) (Fig. 10). Other realizations exhibit similar patterns. Black grid cells correspond to locations where Δq(x) < 0, while white grid cells are points where Δq(x) > 0. As aquifer heterogeneity increases, the area covered by black grid cells expands, consistent with the last line of Table 2. This is somewhat more pronounced in low-lying regions where zonal mean hydraulic conductivities are low (cf. Fig. 1).

The distribution of locations where Δq(x) > 0 (Δq(x) < 0) is fairly uniform in the simulation with relatively low heterogeneity (CV = 0.1). Areas in zone 1 where there seems to be a slightly higher concentration of Δq(x) < 0 may be exceptional. It should also be noted that relatively strong local anisotropies in the direction of the stream appear throughout the basin, except in zone 1. The relatively uniform spatial distribution of the Δq(x) index begins to disappear as *CV* increases. When heterogeneity is high (CV = 1.0, 2.0), areas where the $\Delta q(x)$ index is less than zero generally coincide with zones of low conductivity (like zones 1, 2, 3, etc. in Fig. 1). Zones of high conductivity are somewhat harder to pick out, but there are obvious areas where Δq(x) > 0 that correspond to areas of high zonal conductivity (like zones 10, 8, 9, etc. in Fig. 1).

Coherent areas of relatively high and low flow build up in the groundwater simulations (Fig. 10), especially at high levels of heterogeneity (high CV). This seems related to the increase of the area occupied by points of low conductivity with increasing heterogeneity (Table 1). Yet points of relatively low conductivity do not exhibit spatial coherence (cf. Fig. 9), nor should they since the random perturbation to conductivity that is added at any given point is chosen independently of all other points. The spatial coherence of low flux areas, as measured by $\Delta q(x)$ < 0, results from the cumulative effect on continuous flows of blockages arising from spatially independent low conductivities. The flow paths of other simulations are statistically the same as this one, although their actual locations and geometry are different.

5. Summary and Concluding Remarks

This study focuses on three related hypotheses about the effects of locally variable hydraulic conductivity fields on regional exchanges of water between streams and aquifers: (H1) small-scale heterogeneities of hydraulic conductivity significantly affect simulated stream-aquifer water exchanges in river basins, and hence, computational projections of them; (H2) systematic biases arise in estimates of exchanges if small-scale heterogeneities are smoothed by aggregation into a few sub-regions; and (H3) the biases result from slow-paths in groundwater flow that emerge due to small-scale heterogeneities.

The study addresses these hypotheses through computational experiments by simulating system states of the hydrologic systems of the Middle Heihe River Basin (MHRB) of northern China. The study compares (i) estimates of stream-aquifer exchanges and groundwater fluxes produced by a base model that reflects current practice of parameterizing hydraulic conductivity on sub-regional scales to (ii) estimates produced by a related set of models whose conductivity parameters are spatially heterogeneous on local (grid-point) scales. The computational experiments provide evidence that regional system states, specifically exchanges of water between the MHRB's streams and aquifers, respond significantly to local variations of conductivity.

Local-scale heterogeneities in conductivity fields cannot be resolved by effective zonal parameterizations used in the base model. The cumulative effects of local heterogeneities combine to produce spatially coherent "slow paths" in the resulting simulations of groundwater flows. Flow paths in locally heterogeneous systems appear to intersect enough relatively low conductivity areas compared to the base model to reduce groundwater flow regardless of zone. The systematic behavior observed in the local-scale experiments is also consistent with theoretical results of Matheron (1967) and Gutjahr et al. (1978).

The aquifer-stream discharge response is stronger than the response of stream-aquifer leakage. That may result from discharge's greater sensitivity to variations of hydraulic conductivity in the MHRB. Stream segments that are further downstream are also more affected by heterogeneity than upstream segments. This seems consistent with the overall pattern of leakage in the Heihe River Basin where recharge occurs primarily in upland reaches that are not directly connected to the water table and average conductivities are relatively high, and thus are relatively unaffected by variations in conductivity.

The results presented here strongly suggest that local-scale variations in the conductivity system parameter are not averaged out as scales of representation increase, but instead significantly affect regional exchanges of water between streams and aquifers (H1): in these experiments local heterogeneity has non-linear effects on states of the regional system. That is so even when the level of heterogeneity in conductivity is fairly small, as measured by the coefficient of variation of the simulated conductivity fields (CV). These effects increase with CV. This is especially true of discharge, which the sub-regionally scaled base model systematically over-estimates (H2). The emergence of slow paths in the groundwater flow field (H3) appears to be the source of the non-linearities.

These non-linear effects have implications for water resources management in addition to their intrinsic scientific interest. Since the introduction of computational tools capable of approximating the dynamics of basin-scale hydrologic systems, it has become common to use effectively (zonally) parameterized coupled models to manage regional water resources (Chen and Shu, 2006; Werner et al., 2006; Dragoni et al., 2013). Although estimates of system states produced by such aggregated models have been assumed to approximate watershed dynamics up to reasonable amounts of unsystematic random error, this assumption is generally not tested. The base model used here, and the methods used to produce it, is typical of such zonally parameterized models. It has been an open question whether zonal parameterizations can produce estimates of system states that account for the effects of local-scale heterogeneity that are known to exist in parameters (Schmidt et al., 2006; Rodríguez et al., 2013). The results of this study reveal a systematic bias in approximations of stream-aquifer exchanges in the MHRB produced by the base model. This is especially true of discharge. Similar effects have been observed in other studies (Kalbus et al., 2009; Kurtz et al., 2013; Mendoza et al., 2015; Lackey et al., 2015).

Most of the evidence for (or against) hypotheses like H1-3 will continue to be computational. Several additional lines of investigation suggest themselves immediately. First is the extension of this kind of study to other basins, both natural and synthetic. Extensions to three dimensions may be required in some cases. The respective roles of zonal and local heterogeneity can be investigated explicitly by applying Analysis of Variance (Winter et al., 2006), or other multivariate statistical techniques.

Local-scale heterogeneities are independent in this study. This is a weak assumption in the sense that it places a minimal number of restrictions on the spatial structure of conductivity fields, but its influence on conclusions about H1-3 is unclear. Hydraulic conductivity fields are generally correlated, so experiments that evaluate the effects of different levels and kinds of correlation should be pursued. Synthetic conductivity fields that reproduce specified correlation structures can provide large numbers of realizations of parameter fields with given values of control variables like correlation length and spatial anisotropy.

Finally, the existence of slow paths in groundwater flows, is inferred indirectly here from summary statistics (cf. the last line in Table 2) or by qualitative interpretations of results (Fig. 10). A Langrangian analysis based on particle tracking would allow the interaction between variations in conductivity and groundwater fluxes to be quantified more directly.

Acknowledgements

This work is supported financially by the National Natural Science Foundation of China (Grant 91125018), National Science and Technology Support Program of China (Grant 2013BAB05B03) and Public Non-profile Project of the Ministry of Water Resources of China (Grant 201401031). Part of the work was finished during a 8-month academic visit at the University of Arizona financially supported by China Scholarship Council. The authors wish to thank Prof. Thomas Maddock III and Prof.

Thomas Meixner of the University of Arizona for their comments on an early version of this paper.

References

Bruen, M. P., and Osman, Y. Z.: Sensitivity of stream–aquifer seepage to spatial variability of the saturated hydraulic conductivity of the aquifer, Journal of Hydrology, 293(1), 289-302, 2004.

Chen, X., and Shu, L.: Groundwater evapotranspiration captured by seasonally pumped wells in river valleys, Journal of Hydrology, 318(1), 334-347, 2006.

Christensen, S., Rasmussen, K. R., and Moller, K.: Prediction of regional ground water flow to streams, Ground Water, 36(2), 351-360, 1998.

Doherty, J., Brebber, L., and Whyte, P.: PEST: Model-independent Parameter Estimation, Watermark Numerical Computing, Brisbane, 1994.

Dragoni, W., Mottola, A., and Cambi, C.: Modeling the effects of pumping wells in spring management: The case of Scirca spring (central Apennines, Italy), Journal of Hydrology, 493, 115-123, 2013.

Eaton, T. T.: On the importance of geological heterogeneity for flow simulation, Sedimentary Geology, 184(3), 187-201, 2006.

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

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

Genereux, D. P., Leahy, S., Mitasova, H., Kennedy, C. D., and Corbett, D. R.: Spatial and temporal variability of streambed hydraulic conductivity in West Bear Creek, North Carolina, USA, Journal of Hydrology, 358(3), 332-353, 2008.

Gutjahr, A. L., Gelhar, L. W., Bakr, A. A., and MacMillan, J. R.: Stochastic Analysis of Spatial Variability in Subsurface Flows 2, Evaluation and Application, Water Resources Research, 14 (5), 953-959, 1978.

Hantush, M. M.: Modeling stream–aquifer interactions with linear response functions, Journal of Hydrology, 311(1), 59-79, 2005.

He, Z., and Zhao, W.: Measurement of streambed hydraulic conductivity and anisotropy analysis, Advances in Water Science, 18(3), 351-355, 2007. (in Chinese with English abstract).

Huang, P. F.: Characteristics of groundwater flow and its response to changing environment in the Middle Shule River Basin, Dissertation, Tsinghua University, 2012.

Hu, L. T., Chen, C. X., Jiao, J. J., and Wang, Z. J.: Simulated groundwater interaction with rivers and springs in the Heihe river basin, Hydrological Processes, 21(20), 2794–2806, 2007.

Jia, Y., Ding, X., Qin, C., and Wang, H.: Distributed modeling of landsurface water and energy budgets in the inland Heihe river basin of China, Hydrology and Earth System Sciences, 13(10), 1849-1866, 2009.

Kalbus, E., Reinstorf, F., and Schirmer, M.: Measuring methods for groundwater-surface water interactions: a review, Hydrology and Earth System Sciences, 10(6), 873-887, 2006.

Kalbus, E., Schmidt, C., Molson, J. W., Reinstorf, F., and Schirmer, M.: Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge, Hydrology and Earth

批注 **[Jinfeng8]:** New reference added.

System Sciences, 13(1), 69-77, 2009.

Kurtz, W., Hendricks Franssen, H.-J., Brunner, P., and Vereecken, H.: Is high-resolution inverse characterization of heterogeneous river bed hydraulic conductivities needed and possible? Hydrology and Earth System Sciences, 17(10), 3795-3813, 2013.

Lackey, G., Neupauer, R., and Pitlick, J.: Effects of Streambed Conductance on Stream Depletion. Water, 7(1), 271–287, 2015. doi:10.3390/w7010271

Leake, S. A., and Prudic, D. E.: Documentation of a computer program to simulate aquifer-system compaction using the modular finite-difference ground-water flow model, US Department of the Interior, US Geological Survey, Washington, 1991.

Matheron, G.: Ele'ments pour une the'orie des milieux porueux, Masson, Paris, 1967.

McDonald, M. G., and Harbaugh, A. W.: The history of MODFLOW, Ground Water, 41(2), 280-283, 2003.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models? Water Resources Research, 51(1), 716–728, 2015. doi:10.1002/2014wr015820

Molz Fred J.: Nature and Measurement of Hydraulic Properties: Overview of Evolving Results, Methodology and Concepts, Clemson University, 2000.

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

批注 [Jinfeng9]: New reference added.

Neuman, S. P., Winter, C. L., and Newman, C. M.: Stochastic theory of field-scale fickian dispersion in anisotropic porous media, Water Resources Research, 23(3), 453-466, 1987.

Neuman, S. P.: Universal scaling of hydraulic conductivities and dispersivities in geologic media, Water Resources Research, 26(8), 1749-1758, 1990.

Neuman, S. P.: Generalized scaling of permeabilities: Validation and effect of support scale, Geophys. Res. Lett., 21(5), 349-352, 1994.

Newman, B. D., Vivoni, E. R., and Groffman, A. R.: Surface water-groundwater interactions in semiarid drainages of the American southwest, Hydrological Processes, 20(15), 3371-3394, 2006.

Perkins, S. P., and Sophocleous, M.: Development of a comprehensive watershed model applied to study stream yield under drought conditions, Ground Water, 37(3), 418-426, 1999.

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

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

批注 [Jinfeng10]: New references added.

Rodríguez, L. B., Cello, P. A., and Vionnet, C. A.: Modeling stream-aquifer interactions in a shallow aquifer, Choele Choel Island, Patagonia, Argentina, Hydrogeology Journal, 14(4), 591-602, 2006.

Rodríguez, L., Vives, L., and Gomez, A.: Conceptual and numerical modeling approach of the Guarani Aquifer System, Hydrol. Earth Syst. Sci., 17(1), 295-314, 2013.

Rubin, Y.: Applied Stochastic Hydrogeology, Oxford University Press, New York, NY, 2003.

批注 [Jinfeng11]: New reference added.

Schmidt, C., Bayer-Raich, M., and Schirmer, M.: Characterization of spatial heterogeneity of groundwater-stream water interactions using multiple depth streambed temperature measurements at the reach scale, Hydrology and Earth System Sciences Discussions, 3(4), 1419-1446, 2006.

Sophocleous, M. A., Koussis, A., Martin, J. L., and Perkins, S. P.: Evaluation of simplified stream-aquifer depletion models for water rights administration, Ground Water, 33(4), 579-588, 1995.

Sophocleous, M. A., Koelliker, J. K., Govindaraju, R. S., Birdie, T., Ramireddygari, S. R., and Perkins, S. P.: Integrated numerical modeling for basin-wide water management: The case of the Rattlesnake Creek basin in south-central Kansas, Journal of Hydrology, 214(1), 179-196, 1999.

Spanoudaki, K., Stamou, A. I., and Nanou-Giannarou, A.: Development and verification of a 3-D integrated surface water–groundwater model, Journal of Hydrology, 375(3), 410-427, 2009.

Tian, Y., Zheng, Y., Wu, B., Wu, X., Liu, J., and Zheng, C.: Modeling surface water-groundwater interaction in arid and semi-arid regions with intensive agriculture, Environmental Modeling & Software, 63, 170-184, 2015.

Werner, A. D., Gallagher, M. R., and Weeks, S. W.: Regional-scale, fully coupled modelling of stream–aquifer interaction in a tropical catchment, Journal of Hydrology, 328(3), 497-510, 2006.

Winter, C. L., Guadagnini, A., Nychka, D., and Tartakovsky, D. M.: Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers. Journal of Computational Physics, 217(1), 166–175, 2006. doi:10.1016/j.jcp.2006.01.047

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

批注 [Jinfeng12]: New reference added.

Winter, C. L., Springer, E. P., Costigan, K., Fasel, P., Mniewski, S., and Zyvoloski, G.: Virtual watersheds: simulating the water balance of the rio grande basin. Comput. Sci. Eng., 6(3), 18–26, 2004. doi:10.1109/mcise.2004.1289305

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

批注 [Jinfeng13]: New reference added.

Winter, T. C.: Recent advances in understanding the interaction of groundwater and surface water, Reviews of Geophysics, 33(S2), 985-994, 1995.

Workman, S. R., Serrano, S. E., and Liberty, K.: Development and application of an analytical model of stream/aquifer interaction. Journal of Hydrology, 200(1), 149-163, 1997.

Wroblicky, G. J., Campana, M. E., Valett, H. M., and Dahm, C. N.: Seasonal variation in surface-subsurface water exchange and lateral hyporheic area of two stream-aquifer systems, Water Resources Research, 34(3), 317-328, 1998.

Zhang, D.: Stochastic Methods for Flow in Porous Media: Coping with Uncertainties, Academic Press, San Diego, Calif., 2002.

批注 [Jinfeng14]: New reference added.

Zhou, J., Hu, B. X., Cheng, G., Wang, G., and Li, X.: Development of a three-dimensional watershed modeling system for water cycle in the middle part of the Heihe rivershed, in the west of China, Hydrological Processes, 25(12), 1964-1978, 2011.

Tables

Table 1 Grid-scale K field statistics

$\rho_z = K(x)/K_z$	CV=0.1	CV=0.5	CV=1	CV=2
Arithmetic mean	1	1	1	1
Standard deviation	0.3	0.65	0.98	1.12
Min	0.2	0.04	0.024	0.006
Max	4.1	14.62	37.47	21.64
Median	0.99	0.942	0.89	0.84
Geometric mean	0.96	0.85	0.78	0.76
Percent of $K(x)/K_z < 1$	54.8%	59.0%	62.0%	64.2%

Table 2 Groundwater flow field statistics with local heterogeneity

$\Delta q(x)$	CV=0.1	CV=0.5	CV=1	CV=2
Arithmetic mean	-0.05	-0.21	-0.34	-0.51
Standard deviation	0.175	0.398	0.569	0.833
Min	-1.833	-3.338	-4.901	-12.4
Max	1.54	1.429	3.458	2.721
Median	-0.017	-0.075	-0.14	-0.234
Percent of $q(x) < q_b(x)$	61.9%	71.0%	75.5%	79.5%

Figures

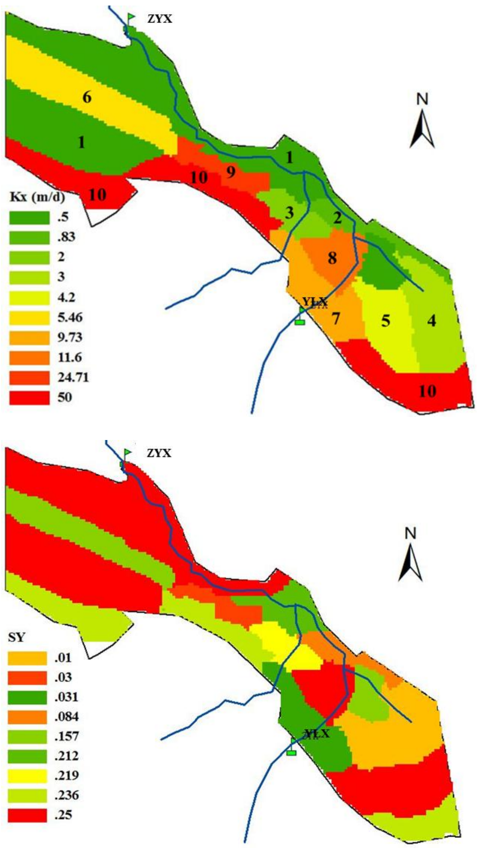
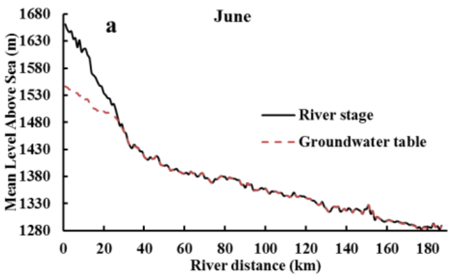


Fig. 1. PEST estimated aquifer conductivity K_z and Specific yield S_z



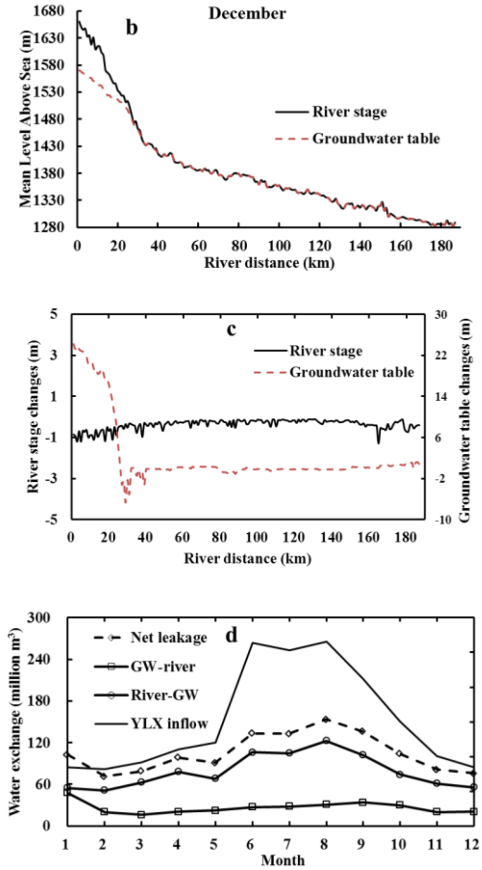
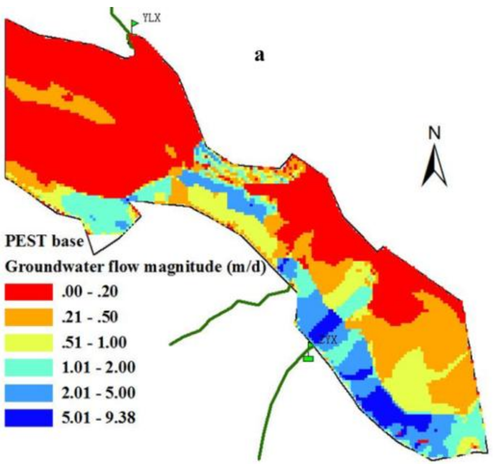


Fig. 2. Simulated groundwater table and river stage along river distance in (a) Jun and (b) Dec; (c) difference between Dec and Jun (Dec-Jun); (d) temporal stream-aquifer exchanges within year



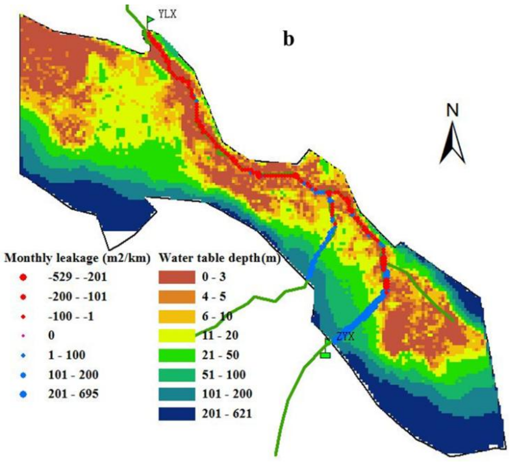


Fig. 3. Groundwater flow rate q_0 distribution (a); Average monthly leakage rate $\overline{Q_b}$ and groundwater table depth (b)

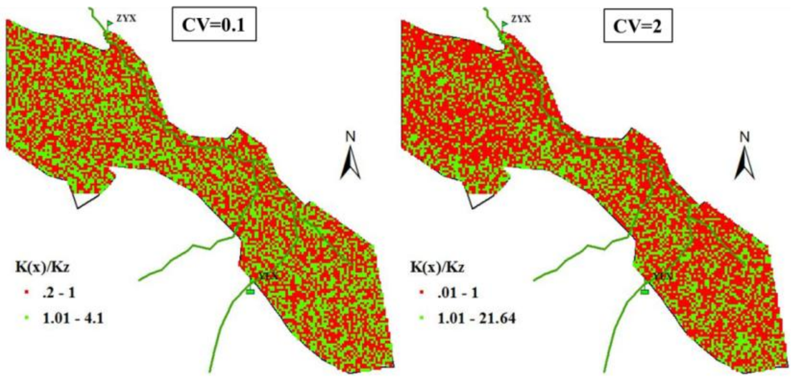


Fig. 4. Grid-scale realization of *K* field perturbation when CV equals 0.1 and 2

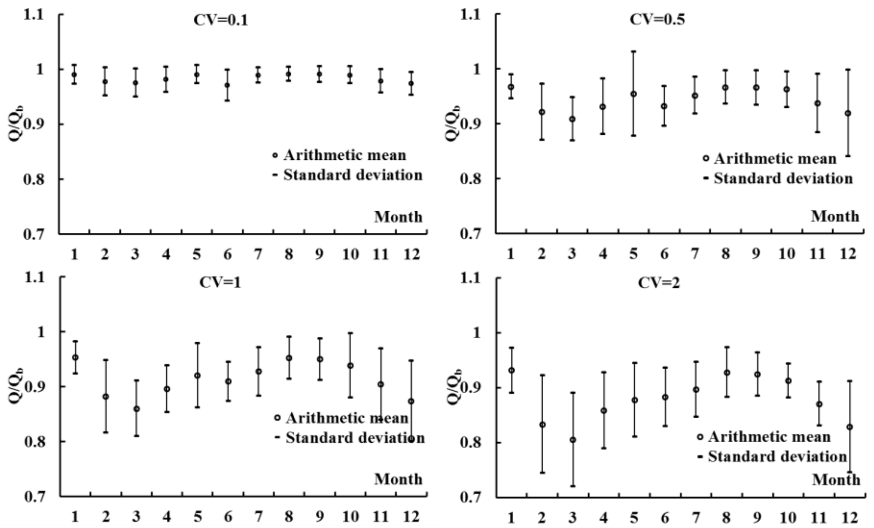


Fig. 5. Means and confidence intervals of aquifer-stream discharge from local model

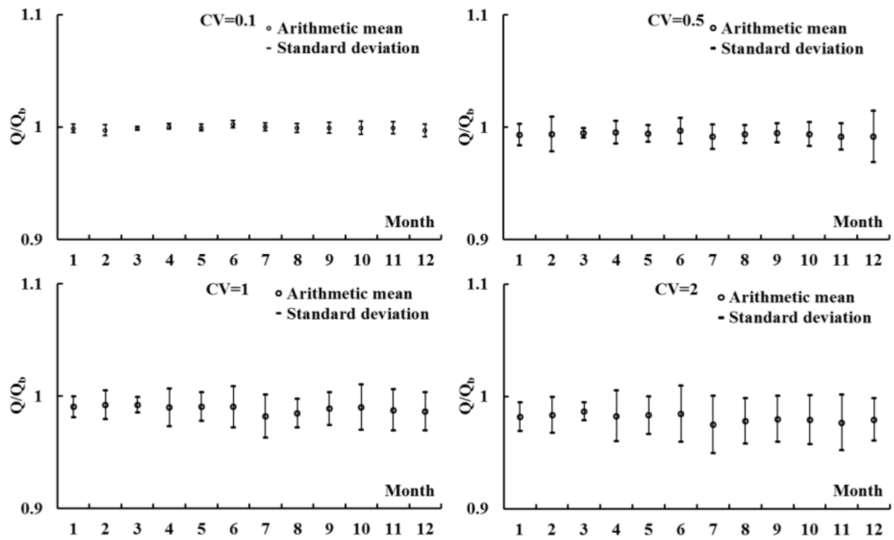


Fig. 6. Means and confidence intervals of stream-aquifer leakage from local model

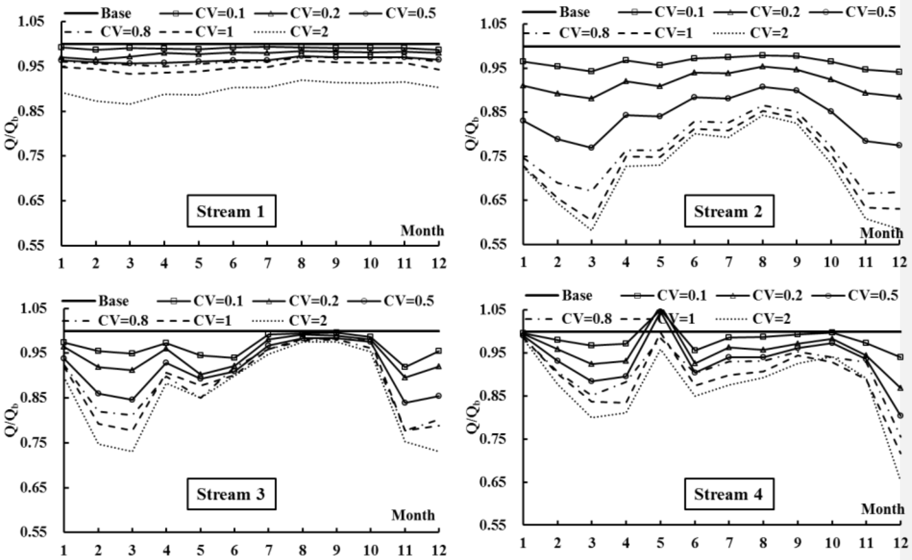


Fig. 7. Aquifer-stream discharges of different streams from refined model

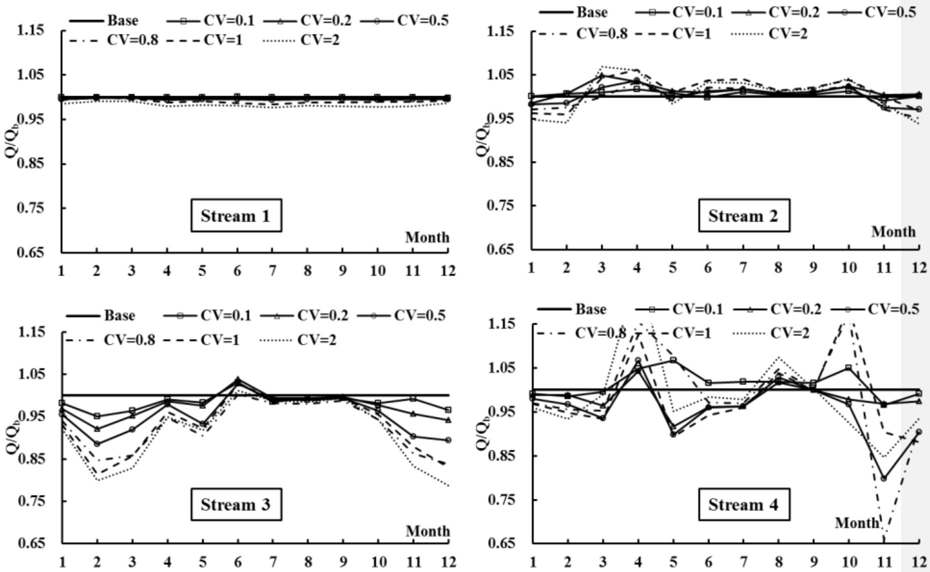


Fig. 8. Stream-aquifer leakages of different streams from refined model

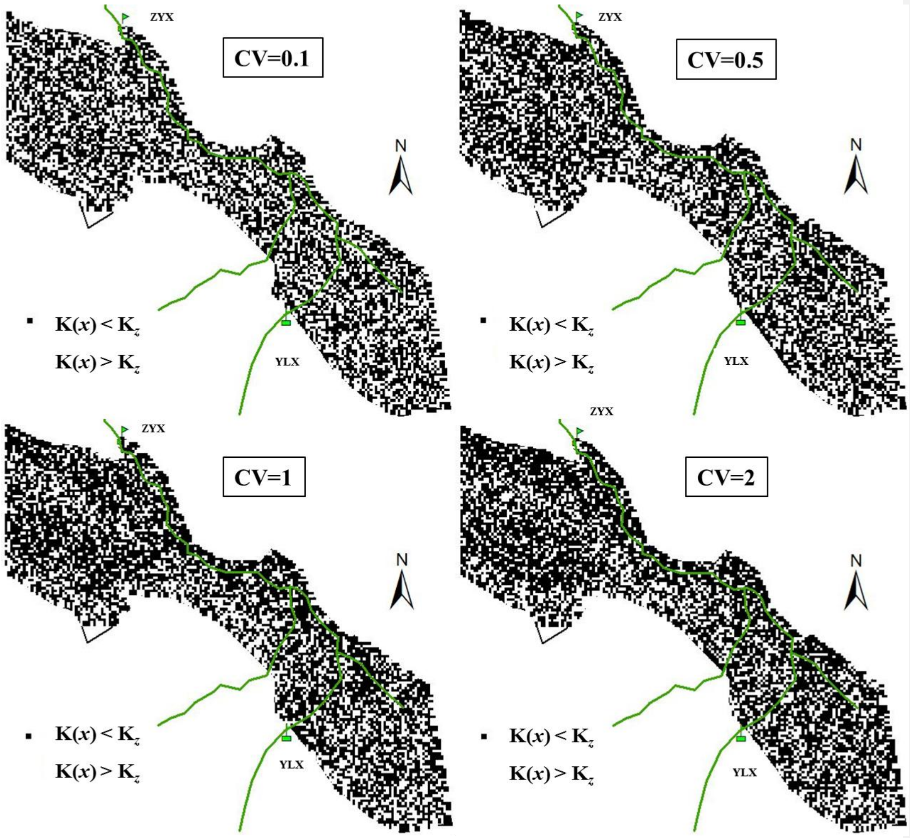


Fig. 9. Grid-scale $K(x)$ field map for CV equaling to 0.1, 0.5, 1 and 2

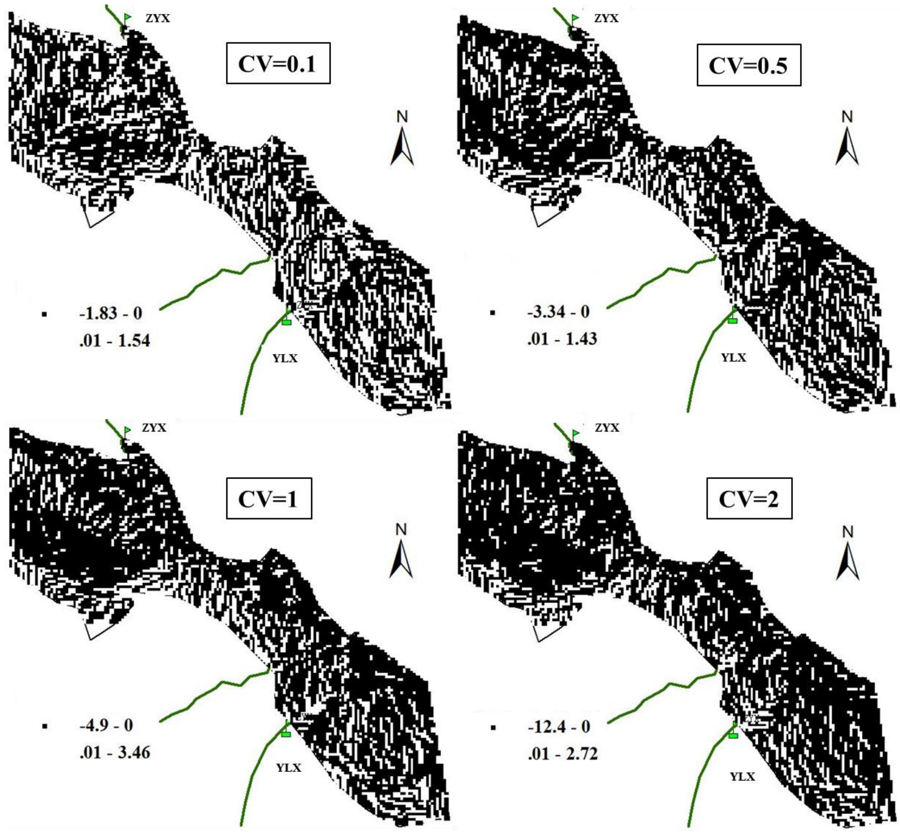


Fig. 10. The field of normalized groundwater flow differences $\Delta q(x)$ with grid-scale $K(x)$