Journal of Hydrology 554 (2017) 383–396

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

The influence of riverbed heterogeneity patterns on river-aquifer exchange fluxes under different connection regimes

Q. Tang [a,b,*,1], W. Kurtz [a,b], O.S. Schilling [c,d], P. Brunner [c], H. Vereecken [a,b], H.-J. Hendricks Franssen [a,b]

[a] Forschungszentrum Jülich GmbH, Institute for Bio- and Geosciences: Agrosphere (IBG-3), 52425 Jülich, Germany
[b] Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, Geoverbund ABC/J, Jülich, Germany
[c] University of Neuchâtel, Centre for Hydrogeology and Geothermics, 2000 Neuchâtel, Switzerland
[d] Department of Geology and Geological Engineering, Université Laval, Pavillon Adrien-Pouliot, Québec, QC G1V 0A6, Canada

ARTICLE INFO

Article history:
Received 24 April 2017
Received in revised form 13 September 2017
Accepted 15 September 2017
Available online 18 September 2017
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Renato Morbidelli, Associate Editor

Keywords:
Data assimilation
Ensemble Kalman filter
Riverbed characterization
Patterns
Fully-coupled physically-based flow modelling
HydroGeoSphere

ABSTRACT

Riverbed hydraulic conductivity (K) is a critical parameter for the prediction of exchange fluxes between a river and an aquifer. In this study, the role of heterogeneity patterns was explored using the fully integrated hydrological model HydroGeoSphere simulating complex, variably saturated subsurface flow. A synthetic 3-D river-aquifer reference model was constructed with a heterogeneous riverbed using non-multi-Gaussian patterns in the form of meandering channels. Data assimilation was used to test the ability of different riverbed K patterns to reproduce hydraulic heads, riverbed K and river-aquifer exchange fluxes. Both fully saturated as well as variably saturated conditions underneath the riverbed were tested. The data assimilation experiments with the ensemble Kalman filter (EnKF) were carried out for four types of geostatistical models of riverbed K fields: (i) spatially homogeneous, (ii) heterogeneous with multi-Gaussian distribution, (iii) heterogeneous with non-multi-Gaussian distribution (channelized structures) and (iv) heterogeneous with non-multi-Gaussian distribution (elliptic structures). For all data assimilation experiments, state variables and riverbed K were updated by assimilating hydraulic heads. For saturated conditions, heterogeneous geostatistical models allowed a better characterization of net exchange fluxes than a homogeneous approximation. Among the three heterogeneous models, the performance of non-multi-Gaussian models was superior to the performance of the multi-Gaussian model, but the two tested non-multi-Gaussian models showed only small differences in performance from one another. For the variably saturated conditions both the multi-Gaussian model and the homogeneous model performed clearly worse than the two non-multi-Gaussian models. The two non-multi-Gaussian models did not show much difference in performance. This clearly shows that characterizing heterogeneity of riverbed K is important. Moreover, particularly under variably saturated flow conditions the mean and the variance of riverbed K do not provide enough information for exchange flux characterization and additional histogram information of riverbed K provides crucial information for the reproduction of exchange fluxes.
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Quantifying river-aquifer interactions is important for understanding flow and transport mechanisms between rivers and aquifers. It is also important for assessing the impact of climate change on water resources (Goderniaux et al., 2009) and balancing human water demands and ecosystem base flow maintenance in arid and semi-arid regions (Zhou et al., 2013). An important requirement for the estimation of river-aquifer exchange fluxes is a proper representation of riverbed properties, e.g. riverbed hydraulic conductivity (K) (Brunner et al., 2017). In early studies, riverbeds were often not simulated explicitly, or strongly simplified and modeled as homogeneous layers (e.g. Fox and Durnford, 2003). However, field measurements and laboratory analysis showed that riverbed properties can be different from the properties of the underlying aquifer, and that riverbed K can vary over several orders of magnitude within a single reach (Calver, 2001; Leek et al., 2009). Studies also indicated a large impact of riverbed heterogeneity on the prediction of river-aquifer exchange fluxes. It is therefore important to assess the impact of riverbed heterogeneities in numerical flow models (Kalbus et al., 2009; Irvine et al., 2012; Lackey et al., 2015; Schilling et al., 2017). Irvine et al. (2012) found that a

* Corresponding author at: Forschungszentrum Jülich GmbH, Institute for Bio- and Geosciences: Agrosphere (IBG-3), 52425 Jülich, Germany.
E-mail addresses: q.tang@fz-juelich.de, qi.tang@awi.de (Q. Tang).
[1] Currently work at Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven, Germany.

https://doi.org/10.1016/j.jhydrol.2017.09.031
0022-1694/© 2017 Elsevier B.V. All rights reserved.

heterogeneous riverbed could be replaced by a homogeneous equivalent while maintaining the accuracy of the prediction of infiltration fluxes in a losing stream system as long as the hydraulic connectivity between river and aquifer is not different between the calibration period and the prediction period. Based on the investigations of Irvine et al. (2012), Schilling et al. (2017) developed an efficient method to rapidly assess the potential for unsaturated flow conditions to appear in heterogeneous riverbed-heterogeneous aquifer systems using key statistical variables. Lackey et al. (2015) compared the estimation of stream depletion using both homogeneous and heterogeneous riverbed K fields, which were varying over two orders of magnitude along the streambed. They found that a homogeneous conceptualization of the riverbed properties always led to errors in the estimate of stream depletion.

In several studies the impact of riverbed heterogeneity was assessed by data assimilation (Hendricks Franssen et al., 2011; Kurtz et al., 2012; Kurtz et al., 2013; Kurtz et al., 2014). Hendricks Franssen et al. (2011) successfully estimated spatially variable leakage coefficients (together with spatially variable aquifer K) with data assimilation for the river-groundwater flow system in the upper Limmat valley in Switzerland. Kurtz et al. (2013) explored whether high-resolution characterization of spatially heterogeneous riverbeds was required in a variably saturated groundwater system with strong river-aquifer interactions and found that it is important to represent at least the basic zones of heterogeneity. Kurtz et al. (2014) also implemented a data assimilation scheme for joint assimilation of groundwater temperature data and piezometric head data in the Upper Limmat Valley near Zurich (Switzerland) and tested cases with heterogeneous riverbeds and found that both model states and parameters can be better predicted by updating aquifer and riverbed hydraulic parameters. None of those earlier studies accounted for more complicated heterogeneous patterns of riverbed K, such as multimodal non-Gaussian distributions, which are expected to occur in practice (Cheng et al., 2011). To fill this gap, we investigated in a previous study the impact of different simple and complex patterns of heterogeneous riverbed K on the characterization of hydraulic heads, riverbed properties, and river-aquifer exchange fluxes (Tang et al., 2015).

1.1. Limitations of previous research

In order to estimate river-aquifer exchange fluxes correctly, it is important that the selected hydrological model calculates these in a physically based manner. The majority of numerical groundwater models (e.g., MODFLOW (McDonald and Harbaugh, 1988), SPRING (Delta h Ingenieugesellschaft mbH, 2006)) are based on a conductance approach. The disadvantage of the conductance approaches is that the non-linear unsaturated flow behavior underneath the river is usually not taken into account, and lateral flow within the riverbed is also neglected (Brunner et al., 2010). Our previous study (Tang et al., 2015) used the model SPRING and was also based on a conductance approach. Fully-integrated physically-based surface-subsurface models are increasingly applied to provide a physically more consistent description of surface water-groundwater exchange processes. Examples are provided by VanderKwaak and Sudicky (2000) (InHM), Panday and Huyakorn (2004) (MODHMS), Kollet and Maxwell (2006) (Parflow), Camporese et al. (2010) (CATHY), Therrien et al. (2010) and Brunner and Simmons (2012) (HydroGeoSphere). Such integrated hydrological models can better capture the dynamics of surface water-groundwater interactions but require additional input parameters (e.g. parameters for simulation of surface water flow) compared to the conductance based approach, where only aquifer states are calculated.

Only a few data assimilation studies were carried out with fully coupled surface water-groundwater models. Examples are provided by Camporese et al. (2009) and Rasmussen et al. (2015), amongst others. Both groundwater head and stream discharge observations were used for assimilation in those works and Rasmussen et al. (2015) also considered the estimation of spatially homogeneous leakage coefficients. Kurtz et al. (2016) developed a data assimilation scheme for the terrestrial system modelling platform TerrSysMP (Shrestha et al., 2014). It allows joint updating of both model states and parameters of the CLM (land surface) and Parflow (subsurface including overland flow) modules. However, until now no studies updated heterogeneous riverbeds in fully coupled surface water-groundwater models.

1.2. Contribution of this paper

Tang et al. (2015) used a conductance based model to conclude that complex non-multi-Gaussian riverbed patterns did not result in significantly different river-aquifer exchange fluxes compared to multi-Gaussian riverbed patterns (with the same overall mean riverbed K and variance as the non-multi-Gaussian riverbed). In this work, we employ a fully coupled surface water-groundwater models and evaluate the impact of complex heterogeneous patterns of riverbed K on river-aquifer exchange fluxes. The heterogeneous patterns of riverbed K are estimated by the ensemble Kalman filter (Evensen, 1994), a data assimilation approach.

The experiments consist of two different cases. In one case, saturated conditions for the riverbed and the underlying aquifer prevail and full hydraulic connection between the river and aquifer is thus maintained. In the second case, variably saturated conditions occur within and underneath the riverbed and, due to the unsaturated conditions the estimation problem becomes non-linear and it is expected that it will be more challenging to infer river-aquifer exchanges fluxes and model states and parameters. The role of different riverbed K patterns for this configuration is therefore much unclearer and results are expected to show more variation. This will provide further insights to what extent representing the geological complexity of a riverbed is required for the characterization of states and fluxes under both fully saturated and variably saturated conditions underneath the riverbed in the river-aquifer system. The analysis was made for four different geostatistical models of riverbed K including non-multi-Gaussian fields with and without channelized structures, multi-Gaussian fields and homogeneous fields.

2. Theory and methods

2.1. Coupled surface-subsurface flow simulations with HydroGeoSphere

If river-aquifer interactions are modelled by the conductance approach, river stages are needed as a predefined input and are conceptualized as fixed head dependent boundary conditions in the river-aquifer model (Knapton, 2009). The river-aquifer exchange fluxes are then calculated as a linear function of the gradient of hydraulic heads between groundwater and surface water (Brunner et al., 2010). However, if the groundwater level is lowered, an unsaturated zone can develop and once the river and the aquifer are hydraulically disconnected exchange fluxes are no longer changing linearly as a function of head differences between the surface water body and the aquifer (Brunner et al., 2009). Compared to conductance-based models, a fully-coupled surface water-subsurface flow model allows a dynamic, two-way coupling when accounting for the river-aquifer interaction. Exchange fluxes are then calculated in a fully coupled manner. HydroGeoSphere

(Therrien et al., 2010; Brunner and Simmons, 2012; Aquanty Inc, 2016) is an example of such fully coupled groundwater-surface water models and used for the work presented in this paper.

In HGS, variably saturated subsurface flow is simulated with the three-dimensional Richards equation and surface water flow is solved using a one- or two-dimensional approximation of the Saint Venant equations. The surface flow and subsurface flow equations are solved simultaneously for each time step at the interface nodes in a joint equation system, and complete water balance and solute budgets are calculated. HGS has successfully been used to simulate the interactions between groundwater and surface water (Goderniaux et al., 2009; Partington et al., 2013), further surface water and groundwater systems interactions with vegetation (Banks et al., 2011; Schilling et al., 2014), micro-topographic wetland runoff (Frei et al., 2010) and also large scale solute transport (Blessent et al., 2011).

The equation describing unsteady shallow water flow derived from the two dimensional Saint Venant equations using the diffusive wave approximation is:

$$\frac{\partial \varphi_0 h_0}{\partial t} = -\nabla \cdot d_0 \mathbf{q_0} - d_0 \Gamma_0 \pm Q_0 \tag{1}$$

where t is time [T], φ_0 is the surface flow domain porosity [-], h_0 is the water surface elevation [L], d_0 is the depth of the flow [L] and Γ_0 is the fluid exchange rate [$L^3\ L^{-3}\ T^{-1}$] between the surface domain and the subsurface domain, which will be described in detail in the following paragraph. Q_0 stands for external sources and sinks [$L\ T^{-1}$], and $\mathbf{q_0}$ is the fluid flux [$L\ T^{-1}$] given by

$$\mathbf{q_0} = -\mathbf{K_0} \cdot k_{r0} \nabla (d_0 + z) \tag{2}$$

where $\mathbf{K_0}$ is surface conductance [$L\ T^{-1}$] dependent on the Manning's equation, k_{r0} is a dimensionless factor varying between 0 and 1, accounting for the additional resistance in horizontal conductance caused by obstruction storage, and z is the elevation above sea level [L]. If a critical depth (the depth of flow where the specific energy is at a minimum) boundary condition is applied at the downstream of the surface flow nodes, stream discharge is calculated based on the upstream water level and the channel geometry:

$$Q_0 = \sqrt{g d_0} \tag{3}$$

where Q_0 is the discharge, g is gravitational acceleration [$L\ T^{-2}$] and d_0 is the flow depth.

Richards equation is used for modelling three-dimensional transient groundwater flow in a variably saturated porous medium:

$$\frac{\partial}{\partial t}(\theta_s s_w) = -\nabla \cdot \mathbf{q} + \Gamma_0 \pm Q \tag{4}$$

where θ_s is the saturated soil water content [–] which is assumed equal to the porosity φ [–], s_w is the degree of water saturation [–], Q are sinks and sources [$L^3\ L^{-3}\ T^{-1}$], and $\mathbf{q}$ is the fluid flux [$L\ T^{-1}$] given by

$$\mathbf{q} = -\mathbf{K} \cdot k_r \nabla (\psi + z) \tag{5}$$

where k_r is relative permeability of the porous medium [-], ψ is the pressure head [L], z is the elevation defined in equation (2) and $\mathbf{K}$ is hydraulic conductivity tensor [$L\ T^{-1}$] given by

$$\mathbf{K} = \frac{\rho g}{\mu} \mathbf{k} \tag{6}$$

where ρ is the density of water [$M\ L^{-3}$], g is gravitational acceleration [$L\ T^{-2}$], μ is the viscosity of water [$M\ L^{-1}\ T^{-1}$], and $\mathbf{k}$ is the permeability tensor of the porous medium [L^2]. This equation is solved using the control volume finite element method. The relationship between ψ, θ and k_r is described by the van Genuchten functions (Van Genuchten, 1980):

$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha|\psi|)^n]^m} \tag{7}$$

$$k_r = \theta^{0.5}\left[1 - \left(1 - \theta^{\frac{1}{m}}\right)^m\right]^2, m = 1 - \frac{1}{n} \tag{8}$$

where θ_r is the residual water content [–], α is the inverse of the air entry pressure head [L^{-1}], and n is the pore size distribution index [-].

In HGS, two approaches can be used for flow coupling between subsurface and surface domains. In the first approach, the top layer of nodes represents both surface and subsurface domains, assuming that the same heads are assigned to the surface and the uppermost subsurface nodes. The surface and subsurface flow equations are solved simultaneously for those surface-subsurface interface nodes. The second approach uses Darcy's law for flow relations between the surface nodes and the first layer of subsurface nodes, with an assumption that they are separated by a thin layer where the leakage occurs. A first-order exchange coefficient is introduced and exchange fluxes are calculated linearly dependent on the pressure difference between the surface and subsurface domain. A detailed comparison between the two approaches is given in Liggett et al. (2012). In both of these approaches subsurface and surface flow equations are fully coupled. In this work, the second approach (called the dual node approach) is selected to represent the relation between surface domain and subsurface domain. Therefore, fluid exchange rates between groundwater and surface water are given by

$$d_0 \Gamma_0 = \frac{k_{re} \mathbf{K_{zz}}}{l_{exch}} (h - h_0) \tag{9}$$

where $\mathbf{K_{zz}}$ is the vertical saturated hydraulic conductivity of the underlying porous medium, h is the subsurface water head [L], h_0 is the surface water head [L], and l_{exch} is the coupling length [L]. The relative permeability term k_{re} for the exchange flux is the same as the relative permeability of the medium k_r in Eq. (5) during the exfiltration process (water flows from subsurface to surface). When infiltration (flow from surface to subsurface) occurs, k_{re} is defined as below:

$$k_{re} = \begin{cases} S_{exch}^{2(1-S_{exch})} & when\ d_0 < H_s \\ 1 & when\ d_0 > H_s \end{cases} \tag{10}$$

where $S_{exch} = \frac{d_0}{H_s}$, H_s is the obstruction height [L].

While the dual node approach follows a similar conceptualization as the leakage principle in conductance based models, the representation of exchange fluxes between river and aquifer in HGS in the unsaturated zone is more realistic because it is limited by the relative permeability. Moreover, if the coupling length l_{exch} of the dual node approach is set to a very small value, the effects of the dual node approach on simulation results are minimal and results are comparable to the common node approach, while providing a higher numerical stability compared to the common node approach (Liggett et al., 2012).

2.2. Data assimilation with ensemble Kalman filter

Data assimilation allows constraining states and parameters of numerical flow models by incorporating observations to correct the model evolution in state space. It is a powerful tool for uncertainty quantification and can also be used to estimate unknown riverbed K. In the present study, the ensemble Kalman filter (EnKF) (Evensen, 1994) is used as the data assimilation technique for simultaneously updating state variables and parameters of the integrated hydrological model HGS. In EnKF, model uncertainties

are taken into account by using multiple realizations of input files for initial conditions, forcing functions and parameters. The corresponding predicted state vectors, called ensemble members or stochastic realizations, are used to represent the statistical distribution of the system state. The data assimilation procedure of this study can be described by the following steps:

1) For each realization of riverbed K generated according to section 3.3, the model states at the current time step are forecasted using the solution from the previous time step as initial condition with HGS:

$$\mathbf{h}_{t,i} = M(\mathbf{h}_{t-1,i}, \mathbf{p}_{t-1,i}, \mathbf{q}_{t-1,i}) \tag{11}$$

where $\mathbf{h}$ is the state vector, $\mathbf{p}$ a vector with the model parameters, $\mathbf{q}$ a vector with model forcings, M the numerical flow model (HGS in this study), t the time step counter and i the realization number.

2) After every time step of the simulation, EnKF is applied to update both model states and parameters according to:

$$\mathbf{x}^a_{t,i} = \mathbf{x}^f_{t,i} + \alpha \mathbf{G}(\mathbf{y}_{t,i} - \mathbf{H}\mathbf{h}_{t,i}) \tag{12}$$

where $\mathbf{x}^a_{t,i}$ is the updated augmented state vector (analysis) containing both states (hydraulic heads) and parameters ($\log_{10}(K)$), $\mathbf{x}^f_{t,i}$ is the forecasted equivalent, α is a damping factor used only for the parameter update which can take values between 0 and 1, $\mathbf{G}$ is the Kalman gain, and $\mathbf{y}_{t,i}$ is a vector with perturbed measurements. According to Burgers et al. (1998), measurements need to be perturbed to account for the measurement uncertainty in order to guarantee the correct estimation of the posterior variance of the updated ensemble with EnKF, which can be written as:

$$\mathbf{y}_{t,i} = \mathbf{y}_t + \varepsilon_{t,i} \tag{13}$$

where $\mathbf{y}_t$ is a vector with the original measurements at time step t, and $\varepsilon_{t,i}$ is a vector with normally distributed measurement errors with zero mean and specified variance accounting for the measurement uncertainty.

The Kalman gain $\mathbf{G}$, which determines the weight assigned to the model simulation on the one hand and the measurements on the other hand, is given by

$$\mathbf{G} = \mathbf{C}\mathbf{H}^T(\mathbf{H}\mathbf{C}\mathbf{H}^T + \mathbf{R})^{-1} \tag{14}$$

where $\mathbf{H}$ is the measurement operator which maps the simulated results on the observation locations, $\mathbf{R}$ is the measurement error covariance matrix, and $\mathbf{C}$ is the covariance matrix with covariances between states, between states and parameters, and between parameters according to:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{hh} & \mathbf{C}_{h,\log_{10}(K)} \\ \mathbf{C}_{\log_{10}(K),h} & \mathbf{C}_{\log_{10}(K),\log_{10}(K)} \end{pmatrix} \tag{15}$$

Once the updating step (step (2)) is finished, the updated states and parameters are used as inputs for the simulation of the next time step, the algorithm returns to step (1), and the procedure is repeated until the end of the simulation. The coupling of EnKF and HGS is described in Kurtz et al. (2017).

3. Synthetic flow modeling experiments

3.1. Overview

To investigate the importance of riverbed K patterns for river-aquifer interactions, simulations with complex, non-multi-Gaussian reference riverbed K fields are compared to simulations with multiple geostatistical models for riverbed K. The model simulation period is two years in total. Data assimilation experiments were performed in the first year, and the second year served as a verification period. Before the data assimilation runs, a spin up run of one year with the same transient model forcings as for the data assimilation experiments was carried out. The modeling experiments consist of the following steps:

1) Ten reference riverbed K fields are generated, and the corresponding river-aquifer interactions are simulated for a two year period with HGS. The first year is considered as assimilation period and second year as verification period. The riverbed K fields, the simulated hydraulic heads and the simulated (net) exchange fluxes serve as the 'truth'.
2) A large number of initial riverbed K fields are generated for each of the different geostatistical models, using geostatistical simulation algorithms.
3) Data assimilation experiments are carried out for a one year period for each of the four geostatistical models, using the initial riverbed K fields as input. The updated riverbed K fields and the simulated heads and fluxes are subsequently evaluated against the reference fields.
4) The estimated fields of riverbed K parameters are used as input for verification experiments for the following one year period with different hydrological conditions and without data assimilation, for each of the four geostatistical models. The simulated heads and fluxes are then evaluated against the reference fields.

3.2. Three dimensional river-aquifer model setup

3.2.1. Model domain

In this study, a three-dimensional synthetic river-aquifer model is simulated. The subsurface domain (i.e., the aquifer) has a spatial extent of 500 m × 250 m × 10.5 m, while the surface domain (i.e., the river) is defined as a channel centered on top of the aquifer, with spatial dimensions of 500 m × 80 m × 0.5 m. Both the surface and subsurface domain are discretized by 3 D-blocks of size 10 m × 10 m × 0.1 m resulting in a grid of 131,250 cells in total. The numerical model is outlined in Fig. 1. The model is inclined along the x-direction with a slope of 0.01 m/m. The riverbed (including the riverbanks) is conceptualized by a rectangular cross-section with eight rows of elements. This constrains the surface water to flow within the channel.

3.2.2. Twin cases set up: saturated vs. variably saturated

The importance of riverbed K patterns for river-aquifer interactions is evaluated for two cases using the same model domain but different states of connection between the river and the aquifer. The 'saturated' case indicates that in the simulated model, riverbed and the subsurface below the riverbed are fully saturated and the river and the aquifer are hydraulically connected. Alternatively, the 'variably saturated' case refers to the model which is partially unsaturated under the river and the system is either fully hydraulically disconnected (i.e., an unsaturated zone is present under the entire streambed) or in a transitional state with partly saturated and unsaturated conditions. This variably saturated case is obtained by adjusting the ratio of K in the aquifer and riverbed according to Brunner et al. (2009), where, in order for an unsaturated zone to occur, the following equation needs to hold true:

$$\frac{K_c}{K_a} \leqslant \frac{h_c}{h_c + d} \tag{16}$$

where K_c refers to the K of the riverbed, K_a is the K of the aquifer, h_c is the thickness of the riverbed, and d is the ponded water depth.

As an example, the saturation map of the variably saturated case for the reference riverbed K field No. 6 is displayed in Fig. 2.

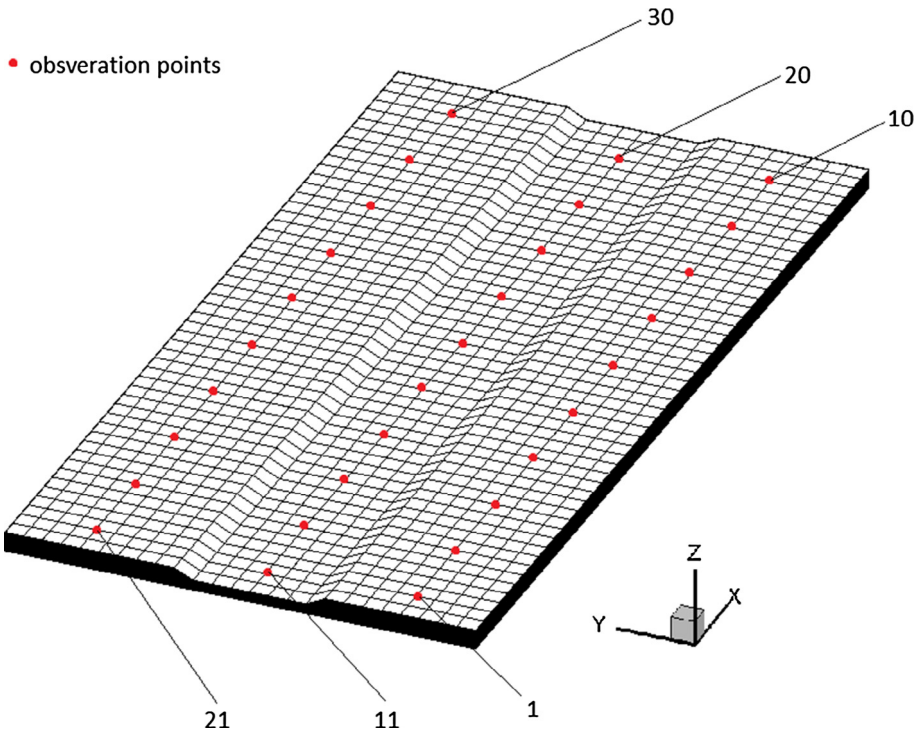

Fig. 1. The synthetic 3-D river-aquifer model domain with locations of synthetic hydraulic head observations (red solid points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

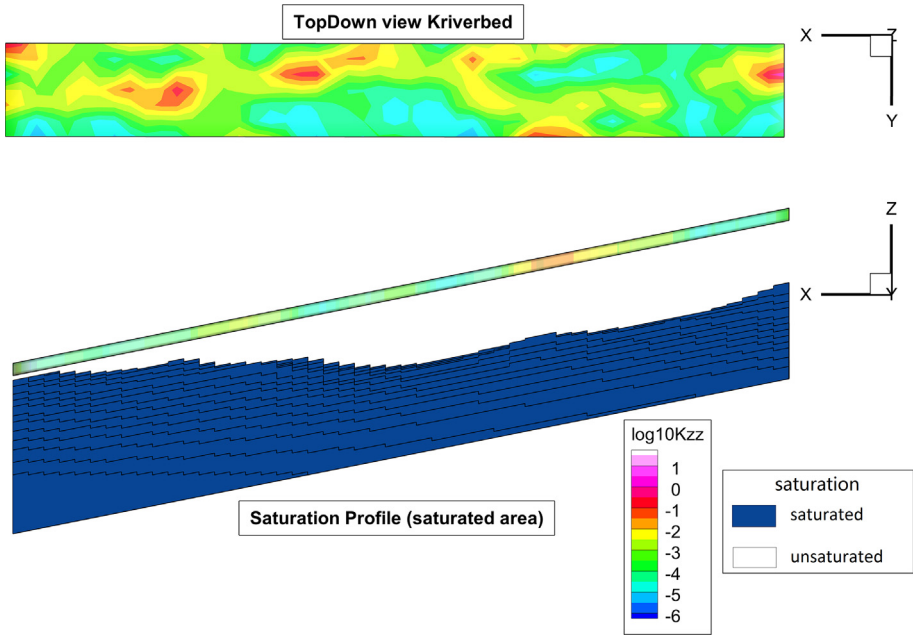

Fig. 2. 2-D view of the $\log_{10}(K)$ map (top row) and connectivity conditions under the river (bottom row) for the reference K field No. 6 of the variably saturated case. Displayed is the initial condition of the assimilation period. The $\log_{10}(K)$ field is colored according to the magnitude of K, and the saturated regions underneath the river are shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.3. Model parameterization

The aquifer and riverbed are conceptualized as gravel-sand material; van Genuchten parameters for aquifer and riverbed are $\alpha = 3.48\ m^{-1}$ and $n = 1.75$, according to Li et al. (2008). Specific storage is set to $10^{-4}\ m^{-1}$ and porosity is 0.25. The coupling length is set to a very small value of 10^{-3} m for calculating the exchange fluxes between the surface and the subsurface domain. Like in the study of Tang et al. (2015), the aquifer is homogeneous with a hydraulic conductivity value of 10^{-3} m/s and only riverbeds are heterogeneous. For the saturated flow case the geometrical mean value of K riverbed is equal to -6 $\log_{10}$ m/s. Under this condition, the aquifer underneath the riverbed is saturated and the relationship between exchange fluxes and head difference from surface water to groundwater is approximately linear. For the variably

saturated case, the geometrical mean value of riverbed K was lowered by two magnitudes.

3.2.4. Boundary conditions

Surface water flow is generated through transient fixed head boundary conditions defined for the upstream nodes along the river cross-section (x = 0, y = 80–160). A critical depth boundary condition is defined at the downstream part (x = 500, y = 80–160). In the subsurface domain, transient prescribed head boundary conditions are defined for the two planes y = 0 and y = 250. No flow boundaries are applied to the aquifer nodes along x = 0 and x = 500. Fig. 3 summarizes the boundary conditions used for the saturated and the variably saturated cases over the complete simulation period. For each of the two cases, the same transient boundary conditions are used for the spin-up runs and the assimilation experiments. For the saturated case, in the assimilation period the transient fixed heads defining the inflow at the river inlet are taken from Kurtz et al. (2014). Transient prescribed heads for the subsurface domain are set between z = 404.5 m and 405.5 m. In the verification period, lower prescribed head boundary values are used for the subsurface domain, but otherwise the setup is similar to the assimilation period. For the variably saturated case, in the assimilation period, river stages are the same as for the saturated case and prescribed heads in the subsurface domain are 5 m lower than for the saturated case. The verification period for the variably saturated case has the same subsurface prescribed heads as the assimilation period but river stages have much larger fluctuations than in the assimilation period.

3.3. Hydraulic conductivity fields

3.3.1. Reference riverbed K

A non-multi-Gaussian riverbed K field with channelized structures was used as the reference riverbed K. For this purpose, a training image reflecting channelized structures is generated using the software SGeMS (Remy et al., 2009), see Fig. 4(a). The training image consists of two facies: channels occupy a proportion of 0.4 of the whole image, and the rest symbolizes background material. The two facies represent two materials with high and low permeability (e.g. sand and clay). In order to avoid random effects in the results of statistical analysis, which might be induced if only one particular reference field is used, ten different K reference fields are generated. The ten realizations of facies distributions are generated using the direct sampling method (Mariethoz et al., 2010). Within each facies, $\log_{10}(K)$ values are generated independently by sequential multi-Gaussian simulation using GCOSIM3D (Gómez-Hernández and Journel, 1993). The corresponding geostatistical parameters are provided in Table 1.

3.3.2. Initial riverbed K used for the data assimilation experiments

As outlined in section 3.1, four different geostatistical models were used to generate initial riverbed K fields for the assimilation experiments: (1) non-multi-Gaussian distributed fields with channelized structures, (2) non-multi-Gaussian fields with elliptic structures, (3) multi-Gaussian fields and (4) homogeneous fields. As for the reference K fields, non-multi-Gaussian fields are simulated with the direct sampling method. The multi-Gaussian fields are generated by sequential multi-Gaussian simulation.

The two types of non-multi-Gaussian heterogeneous K fields are generated as follows:

1) Two training images (for each of the two types of non-multi-Gaussian fields independently, see Fig. 4) are generated by SGeMS. The training image for the non-multi-Gaussian model with channelized structures is the same as used for the generation of the ten reference fields. Each of the two training images is composed of two different facies with a proportion of 0.4 for channels/ellipses.
2) In total 200 stochastic realizations of the spatial distribution of facies are generated with help of the direct sampling method, using the training images from step 1).
3) Each facies is independently populated with multi-Gaussian distributed $\log_{10}(K)$ values, using the sequential Gaussian simulation technique. The geostatistical parameters used for defining the variograms for each facies are given in Table 1.

For the multi-Gaussian heterogeneous K fields, a similar mean and variance of riverbed $\log_{10}(K)$ are used as for the non-multi-Gaussian random fields by sequential multi-Gaussian simulation. The mean and variance for the multi-Gaussian random fields is calculated from the mean and variance of each of the two facies of the non-multi-Gaussian fields, using the geometric mean. A more detailed description of the generation technique of non-multi-Gaussian and multi-Gaussian riverbed fields is given in Tang et al. (2015).

The stochastic realizations of $\log_{10}(K)$ values for the equivalent homogeneous case are generated by taking the geometric mean of each of the 200 random fields for the non-multi-Gaussian case with channelized structures. Fig. 5 shows examples of stochastic realizations of riverbed K for each of the four geostatistical models. Fig. 6 displays the histograms of riverbed K, calculated over all 200 stochastic realizations for each of the three geostatistical models with heterogeneous K. The figure illustrates that the histogram is bimodal for the two non-multi-Gaussian cases, but Gaussian for the multi-Gaussian case.

3.4. Modelling strategy

3.4.1. Spin-up run

The spin-up process includes two parts: first a quasi-steady state simulation that runs for 10,000 days with constant boundary conditions corresponding to the forcings of the first time step of the assimilation period, followed by an additional one year exit spin-up run with the 200 stochastic realizations of riverbed K fields, using the same transient boundary conditions as for the assimilation experiments. The one year quasi-steady state simulation departs from dry initial conditions. The following one year exit spin-up runs are made for each of the four geostatistical models and two cases (fully saturated and variably saturated) and result in different initial heads for each of the different realizations of riverbed K. These differing initial conditions are necessary to reflect the different interactions resulting from the different K fields, and allow generating an adequate ensemble spread. In summary, the spin-up runs provide the initial hydraulic heads needed for the data assimilation experiments.

3.4.2. Observations obtained from the reference runs

For the data assimilation experiments, 30 piezometric head observations are taken from the sixth layer (saturated case) or bottom layer (variably saturated case) and used as virtual observations. 10 of the 30 measurements are beneath the riverbed and the other 20 in the aquifer north and south of the river. These virtual observations are taken from each of the ten reference runs. A measurement error of 5 cm is imposed.

3.4.3. Data assimilation experiments

Data assimilation experiments are performed for each of the four geostatistical models, using in total 200 stochastic realizations per model. The damping factor for the parameter update is set to

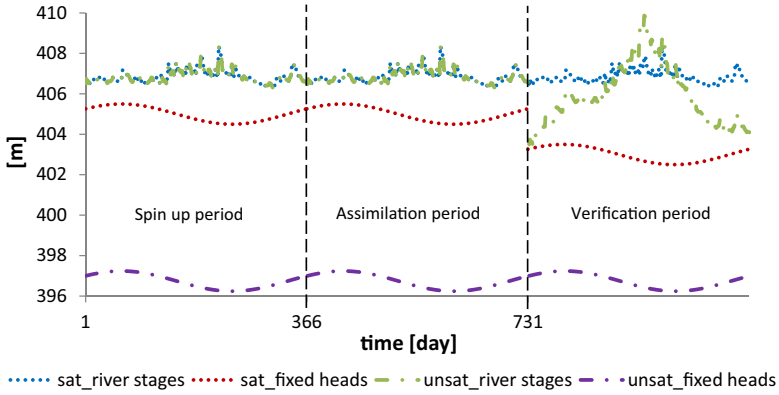

Fig. 3. Fixed head boundary conditions for the subsurface domain for the saturated case (red round dots; abbr. sat_fixed heads) and for the variably saturated case (purple long dash dots; abbr. unsat_fixed heads). Transient river stages for the upstream river nodes are shown for the saturated case with blue round dots (abbr. sat_river stages), and for the variably saturated case with green long dash dots (abbr. unsat_river stages). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

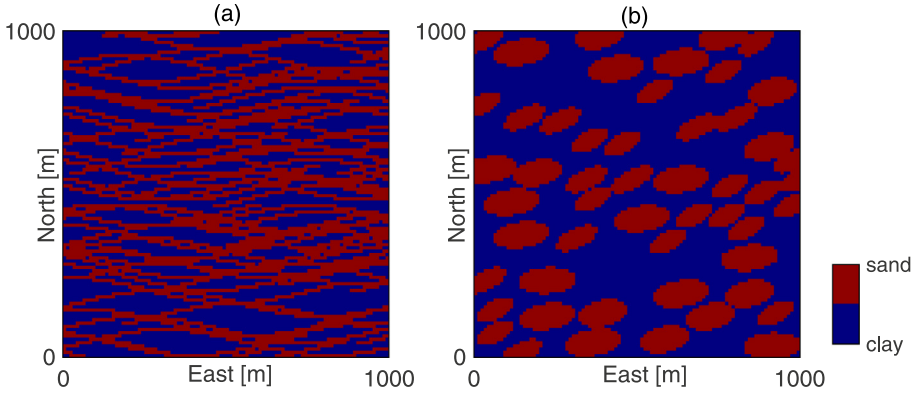

Fig. 4. Two training images for the generation of stochastic realizations for non-multi-Gaussian *K* fields: (a) with channelized structures and (b) with elliptic structures. (a) is also used for generating ten reference riverbed *K* fields.

Table 1
Geostatistical parameters for the two facies in the non-multi-Gaussian distributed *K* fields and geostatistical parameters for the multi-Gaussian fields.

Facies	Variogram type	Mean ($\log_{10}$(m/s))	Range (m)	Sill ($\log_{10}$(m/s))
channel/ellipse	Spherical	−4.3	100	0.5
background	Spherical	−7.3	100	0.5
multi-Gaussian	Spherical	−6.1	100	2.7

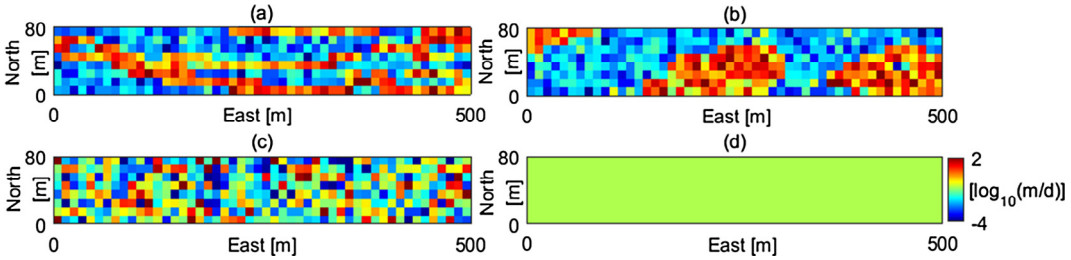

Fig. 5. Examples of stochastic realizations of initial riverbed *K* fields: (a) non-multi-Gaussian field with channelized structures; (b) non-multi-Gaussian field with elliptic structures; (c) multi-Gaussian field; (d) homogeneous field.

0.1. In the data assimilation experiments, hydraulic heads and riverbed *K* are updated daily by assimilating the 30 virtual piezometric head data. Open loop simulations, without data assimilation, are also performed for comparison purposes. Table 2 summarizes the assimilation scenarios.

3.4.4. Verification experiments

The estimated hydraulic head and riverbed *K* fields at the end of the assimilation period are used as input fields for the verification experiments. This is done for both the data assimilation runs with update of heads and $\log_{10}$ (*K*) and for the open loop simulations,

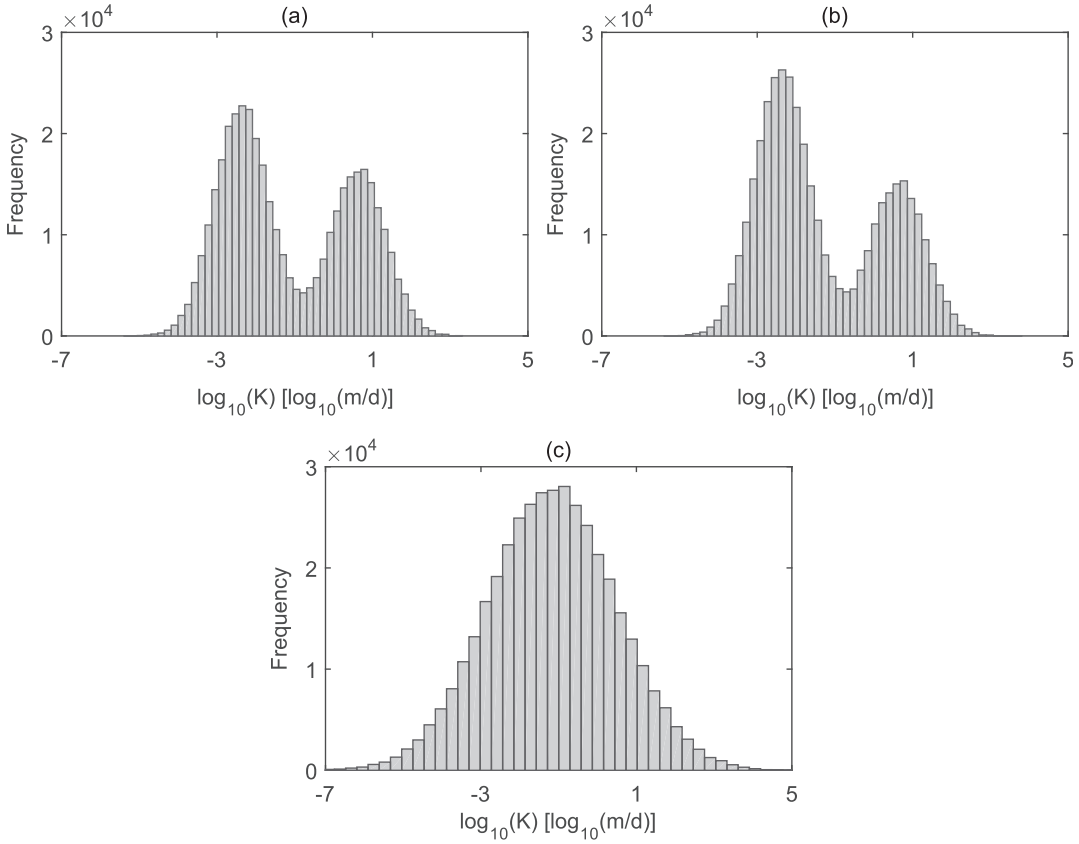

Fig. 6. Histograms of riverbed K for three geostatistical models, calculated over all the 200 realizations: (a) non-multi-Gaussian fields with channelized structures; (b) non-multi-Gaussian fields with elliptic structures; (c) multi-Gaussian fields.

Table 2
Simulation scenarios for data assimilation experiments. The symbol '√' represents the update of h/$\log_{10}(K)$ is done, while '×' represents not.

Riverbed patterns	Scenarios	Update h	Update $\log_{10}(K)$
non-multi-Gaussian (channel)	channel_open loop	×	×
	channel_hK	√	√
non-multi-Gaussian (elliptic)	ellip_open loop	×	×
	ellip_hK	√	√
multi-Gaussian	multi_open loop	×	×
	multi_hK	√	√
homogeneous	homo_open loop	×	×
	homo_hK	√	√

and for each of the four geostatistical models of riverbed K. In the one year verification period the parameter fields are not modified and no data assimilation is done.

3.4.5. Performance measures

The Root Mean Square Error (RMSE) is used to evaluate the characterization of model states, riverbed K and river-aquifer exchange fluxes. RMSE is evaluated separately both for the data assimilation as well as the verification periods. The RMSE for hydraulic heads is calculated for all model nodes (including both surface nodes and subsurface nodes) over all simulation time steps:

$$\mathrm{RMSE}(h) = \sqrt{\frac{1}{n_t n_{nodes\_all}} \sum_{i=1}^{n_t} \sum_{j=1}^{n_{nodes\_all}} \left(\bar{h}_{i,j}^{f} - h_{i,j}^{ref}\right)^2} \tag{17}$$

where n_t is the number of simulation time steps, n_{nodes_all} the total number of model nodes, the overbar indicates ensemble average, h hydraulic heads, the superscript f indicates model simulations and the superscript ref the reference field.

The RMSE for the updated riverbed K is calculated over all riverbed elements at the final simulation time step:

$$\mathrm{RMSE}(K) = \sqrt{\frac{1}{n_{river\_ele}} \sum_{j=1}^{n_{river\_ele}} \left(\overline{\log_{10}(K_j^a)} - \log_{10}(K_j^{ref})\right)^2} \tag{18}$$

where $\overline{\log_{10}(K_{j,n_t}^a)}$ is the updated ensemble mean value of $\log_{10}(K)$ for the j^{th} element of the riverbed at the final time step of the assimilation period and n_{river_ele} is the number of river elements.

The RMSE for exchange fluxes is calculated over all the river nodes for all time steps:

$$\mathrm{RMSE}(Q) = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} \left(\overline{Q_i} - Q_i^{ref}\right)^2} \tag{19}$$

$$Q_i = \sum_{k=1}^{n_{nodes\_river}} Q_k \tag{20}$$

where n_{nodes_river} is the number of the river nodes.

For an individual time step, the RMSE for hydraulic heads and net exchange fluxes is calculated according to:

$$\mathrm{RMSE}(h,t) = \sqrt{\frac{1}{n_{nodes\_all}} \sum_{j=1}^{n_{nodes\_all}} \left(\bar{h}_{i,j,t}^{f} - h_{i,j,t}^{ref}\right)^2} \tag{21}$$

$$\mathrm{RMSE}(Q,t) = |\overline{Q_{i,t}} - Q_{i,t}^{ref}| \tag{22}$$

4. Results

4.1. Saturated case

4.1.1. Assimilation period

Compared to the open loop simulations, data assimilation improved the characterization of model states and parameters, especially when the models included heterogeneous riverbed structures. Since the simulation of heads and exchange fluxes is influenced by model dynamics, average RMSE (h,t) and RMSE (Q, t) calculated over ten references are shown in Fig. 7(a, c). Displayed are results for the four geostatistical models. From Fig. 7(a) we see that only at the beginning of the data assimilation period RMSE (h, t) is clearly higher for ellip_hK and homo_hK compared to the other two models, but later RMSE shows a substantial reduction. After a simulation period of 50 days, channel_hK and ellip_hK have very similar errors, with lower RMSE than for multi_hK and homo_hK. Within the assimilation period, the maximum difference in RMSE (h,t) among the four geostatistical models is 7 cm.

The boxplot of RMSE (K) calculated over ten reference models is displayed in Fig. 8(a). The final updated ensemble mean riverbed $\log_{10}$ (K) field for the four different geostatistical models is shown for one of the reference fields in Fig. 9(a–d). The non-multi-Gaussian models with channelized structures resemble the true K field best, while, non-surprisingly, the homogeneous model deviates most. Nevertheless, the differences between the two non-multi-Gaussian models are small.

The large differences in the characterization of riverbed K patterns between the non-multi-Gaussian models and the other models also manifest themselves in terms of exchange fluxes: Compared to the non-multi-Gaussian model with channelized structures, the average RMSE (Q) over ten references is 61.3% higher for the non-multi-Gaussian model with elliptic structures and 32.6% higher for the multi-Gaussian model. The homogeneous model performs considerably worse than the others, which resulted in the highest errors: the RMSE (Q) is 271.3% higher than for the non-multi-Gaussian model with channelized structures.

4.1.2. Verification experiment

The results of the verification experiments are evaluated in terms of RMSE (h) and RMSE (Q), shown in Table 3. Values for the ten individual reference fields and their average are provided. The non-multi-Gaussian model with channelized structures results in the smallest RMSE (h) in four out of ten cases, the non-multi-Gaussian model with elliptic structures in five out of ten, and the multi-Gaussian model in one. The non-multi-Gaussian models thus clearly outperform the other models. The average RMSE (h) is also lower for the heterogeneous models (2.9–3.2 cm) than for the homogeneous model (6 cm). Among the three heterogeneous models, the differences are minor. Slightly smaller RMSE (h) are found for two non-multi-Gaussian models (average RMSE (h) 2.9–3.0 cm) than the multi-Gaussian model (average RMSE (h) 3.2 cm).

Also in terms of characterization of river-aquifer exchange fluxes the heterogeneous models outperform the homogeneous models. The average RMSE (Q) is 3.17 m^3/d for the homogeneous models compared to 1.7–2.3 m^3/d for the heterogeneous models; the RMSE (Q) is about 86.5% larger for the homogeneous model than for the non-multi-Gaussian model with channelized structures. Among the three heterogeneous models, like for the RMSE (h), similar performance is observed for the two non-multi-Gaussian models (average RMSE (Q) between 1.7 and 1.8 m^3/d), while the RMSE (Q) for the multi-Gaussian model is 41.2% larger than the non-multi-Gaussian model with channelized structures.

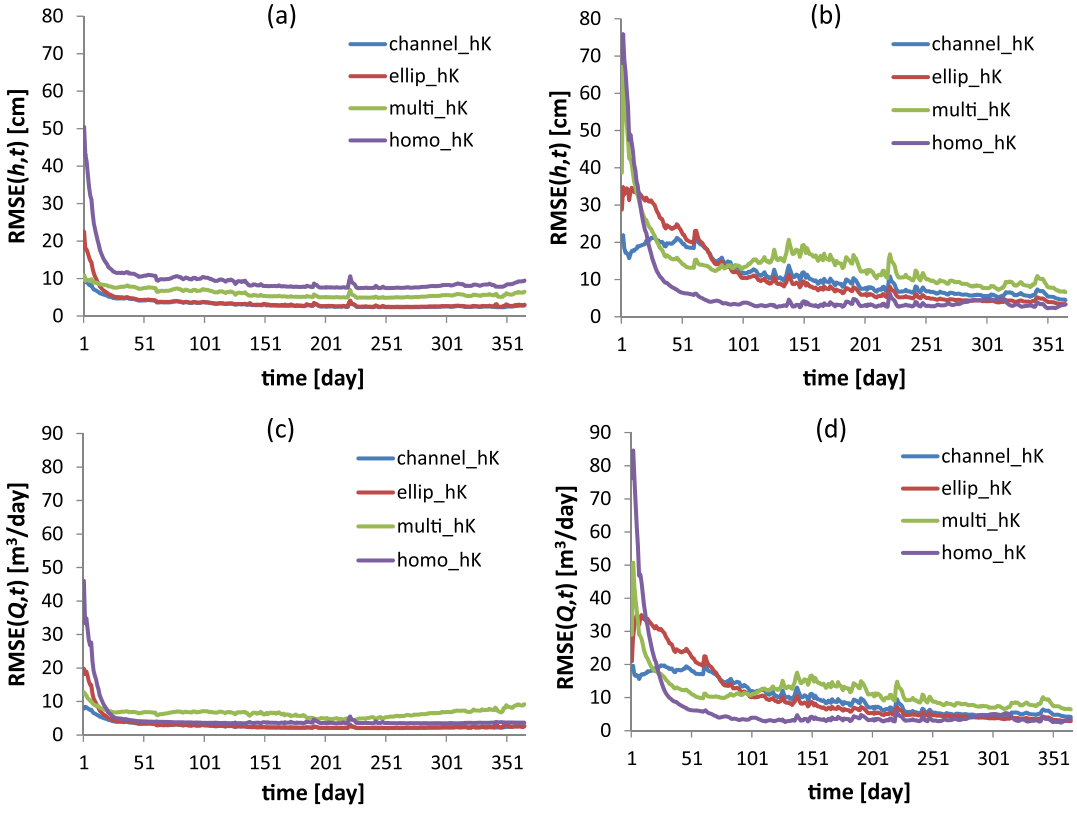

Fig. 7. Average RMSE (h,t) for (a) the saturated case and (b) the variably saturated case; average RMSE (Q,t) for (c) the saturated case and (d) the variably saturated case. These RMSEs were calculated over ten references and for the four geostatistical models, the saturated case and assimilation period.

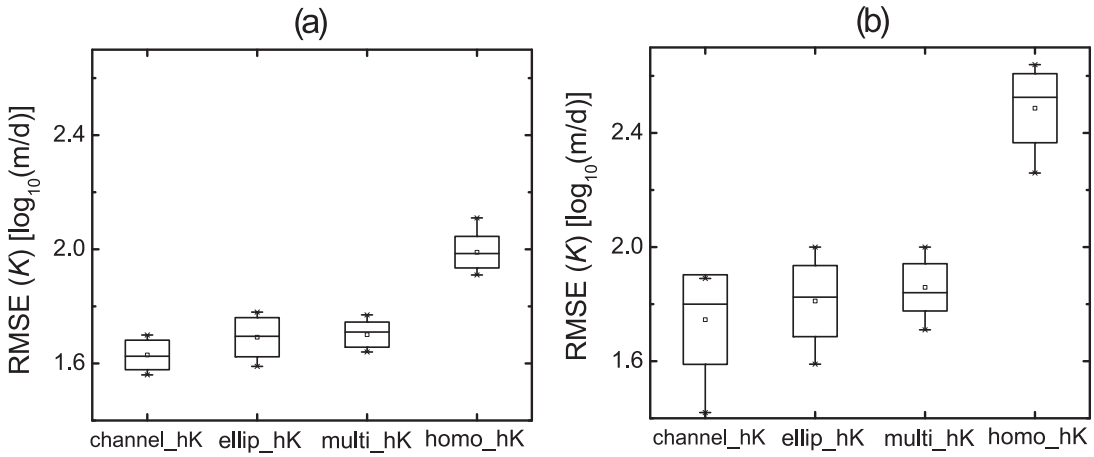

Fig. 8. Boxplots of RMSE (K) calculated over ten references (K is evaluated at the end of the assimilation period) for the four geostatistical models for (a) the saturated case and (b) the variably saturated case.

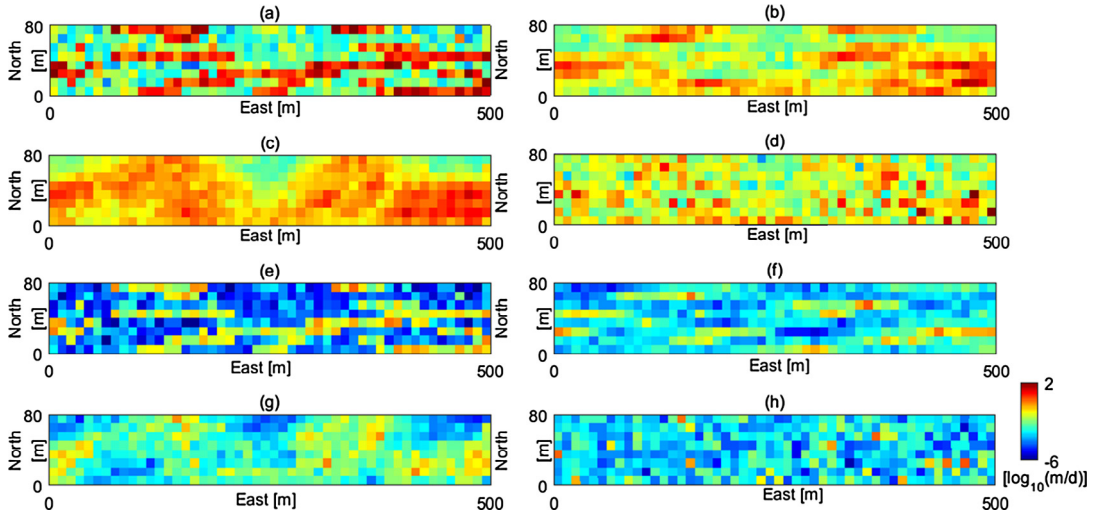

Fig. 9. Ensemble mean riverbed $\log_{10}(K)$ fields at the end of the assimilation period for the saturated case (a–d) and the variably saturated case (e–h). Shown are the reference (a, e) and different geostatistical models: (b, f) channel_hK; (c, g) ellip_hK; (d, h) multi_hK.

Table 3
RMSE (h) (cm) and RMSE (Q) (m^3/d) for the saturated case in the verification period. Shown are results for the ten reference fields and the four geostatistical models: the non-multi-Gaussian model with channelized structures (channel), the non-multi-Gaussian model with elliptic structures (ellip), the multi-Gaussian model (multi) and the homogeneous model (homo).

	channel		ellip		multi		homo	
	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)
Ref_1	3.16	2.07	2.72	2.00	3.24	1.86	4.75	3.19
Ref_2	2.78	1.60	3.06	1.97	3.05	2.22	3.97	0.95
Ref_3	2.65	2.55	2.91	3.11	2.97	3.19	7.97	5.87
Ref_4	3.38	2.55	3.10	2.63	4.71	1.69	7.88	4.53
Ref_5	2.94	3.61	2.68	3.86	2.65	3.77	5.02	3.57
Ref_6	3.20	0.95	3.26	0.69	3.32	0.55	7.80	2.72
Ref_7	2.79	0.32	2.73	0.25	2.73	0.39	4.65	3.04
Ref_8	3.46	1.04	3.19	1.34	3.37	0.65	4.60	0.60
Ref_9	2.67	1.69	2.96	1.53	3.43	0.70	8.83	5.33
Ref_10	2.93	0.60	2.60	0.47	2.87	8.45	4.47	1.88
Average	**3.00**	**1.70**	**2.92**	**1.79**	**3.23**	**2.35**	**6.00**	**3.17**

4.2. Variably saturated case

4.2.1. Assimilation period

The average RMSE (h,t) and RMSE (Q,t) calculated over all ten references are shown in Fig. 7(b, d). The errors are larger in the beginning of the assimilation period and vary strongly between the different models due to the different initial conditions. However, towards the end of the assimilation period, the RMSE (h,t) and RMSE (Q,t) for the four models become smaller and more similar. Within the first 100 days, data assimilation allows reducing

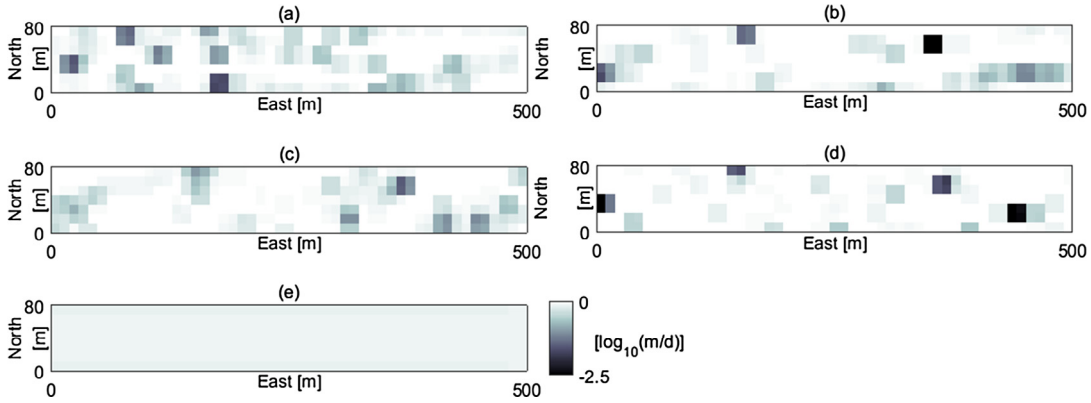

Fig. 10. Maps with exchange fluxes for the variably saturated case at the end of the assimilation period. Shown are the reference (a) and results for the different geostatistical models: (b) channel_hK; (c) ellip_hK; (d) multi_hK; (e) homo_hK. Negative fluxes indicate infiltration from the river into the aquifer.

Table 4
RMSE (h) (cm) and RMSE (Q) (m^3/d) for the variably saturated case in the verification period. Shown are results for the ten reference fields and the four geostatistical models.

	channel		Ellip		multi		homo	
	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)	RMSE (h)	RMSE (Q)
Ref_1	14.48	2.45	13.67	0.75	19.54	6.08	42.92	7.16
Ref_2	18.69	3.31	20.87	2.02	21.73	5.40	44.85	2.86
Ref_3	18.12	2.69	19.05	1.52	18.39	3.40	59.61	14.01
Ref_4	18.90	4.28	17.72	2.11	17.62	3.78	37.96	2.49
Ref_5	21.08	4.87	17.64	2.86	21.44	6.92	32.21	4.05
Ref_6	21.53	3.55	18.66	4.46	22.84	6.38	41.22	4.04
Ref_7	14.37	5.26	15.50	6.38	17.14	7.14	41.21	8.36
Ref_8	18.58	2.72	14.28	2.23	19.24	3.76	50.48	4.36
Ref_9	15.83	3.17	15.65	0.79	17.67	4.43	23.55	3.94
Ref_10	20.81	2.19	17.45	1.79	20.28	4.86	29.94	5.91
Average	**18.24**	**3.45**	**17.05**	**2.49**	**19.59**	**5.22**	**40.40**	**5.72**

the RMSE (h,t) by more than 80% for all four geostatistical models of riverbed K, especially for the homogeneous model which produces the smallest RMSE (h) at the end of the assimilation period. Fig. 8(b) shows the boxplot of RMSE (K) for the updated riverbed K fields. In six out of ten cases the non-multi-Gaussian case with channelized structures shows the best characterization of K in terms of RMSE (K). Fig. 9(e–h) indicates that only the non-multi-Gaussian model with channelized structures could capture some of the channels present in the reference K fields; the multi-Gaussian model deviates strongly from that reference pattern. Overall, heterogeneous fields allow better capturing the spatial distribution of exchange fluxes, although their exact spatial position is often not correct. This is illustrated in Fig. 10. The relatively good performance of the homogeneous model in reproducing exchange fluxes compared to the heterogeneous models is related to very large flux errors for a few grid cells in the heterogeneous cases, which have an important influence on the performance statistics, even though we focus on the net fluxes.

4.2.2. Verification experiment

The verification experiment with variably saturated conditions was carried out with stronger variations in river stage than during the assimilation phase. The RMSE (h) and RMSE (Q) scores of the verification period are provided in Table 4. As in the assimilation period, the RMSE (h) for the three heterogeneous models are smaller than for the homogeneous model. In total, the non-multi-Gaussian model with channelized structures results in the smallest RMSE (h) for three out of ten reference cases, the non-multi-Gaussian model with elliptic structures six out of ten, and the multi-Gaussian model in one out of ten cases. The two non-multi-Gaussian models result in the lowest mean RMSE (h). The average RMSE (h) for the non-multi-Gaussian model with channelized structures is 18.2 cm and for the non-multi-Gaussian model with elliptic structures it is 17.1 cm. The multi-Gaussian model shows only a slightly larger mean RMSE (h) (19.6 cm). The homogeneous model leads to a RMSE (h) of 40.4 cm, which is 121% higher than for the non-multi-Gaussian model with channelized structures.

The two non-multi-Gaussian models result in the best characterization of net exchange fluxes: in two out of ten cases the non-multi-Gaussian model with channelized structures outperforms all other geostatistical models, and in eight out of ten cases the non-multi-Gaussian model with elliptic structures. The average RMSE (Q) of the non-multi-Gaussian model with channelized structures and the non-multi-Gaussian model with elliptic structures are 3.45 m^3/d and 2.49 m^3/d, respectively. Although the homogeneous riverbed has clearly the largest RMSE (h) and RMSE (K), for the characterization of net exchange fluxes the homogeneous model performs only slightly worse (RMSE (Q) = 5.72 m^3/d) compared to the multi-Gaussian model (5.22 m^3/d). The mean RMSE (Q) for the homogeneous model is 65.8% larger than for the non-multi-Gaussian model with channelized structures.

4.3. Correlation analysis

In a final step, linear correlations were calculated between different observations and simulated heads for all riverbed nodes, all references and all simulation scenarios, with the aim to find out whether head observations in the assimilation process provided necessary and sufficient information or not. The percentage of grid cells for which the absolute value of the linear correlation coefficient was >0.5 and >0.2 was calculated, for three different time

Table 5
The percentage of grid cells for which the absolute value of the linear correlation coefficient was >0.5 and >0.2, for three different time steps, the four different geostatistical models and the two different considered cases. The percentage of grid cells is an average value calculated over all reference fields.

model	*K* type	t = 50 days		t = 180 days		t = 365 days	
		>0.5	>0.2	>0.5	>0.2	>0.5	>0.2
Saturated case	channel	2.60%	21.66%	2.44%	21.39%	1.60%	15.04%
	ellip	2.40%	18.43%	2.45%	19.88%	1.42%	10.31%
	multi	1.32%	8.47%	1.15%	10.43%	1.13%	8.16%
	homo	99.01%	99.56%	99.78%	100.00%	98.69%	99.78%
Variably saturated case	channel	0.87%	15.91%	1.06%	16.51%	1.16%	16.43%
	ellip	0.85%	11.53%	1.09%	19.06%	2.77%	32.53%
	multi	0.59%	9.64%	0.81%	9.91%	2.11%	24.90%
	homo	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

steps (after 50, 180 and 365 days), all four different geostatistical models and the two different saturation cases. The percentage of grid cells represents an average value calculated over all reference fields. Table 5 summarizes these scores. The correlations for the homogeneous model are much stronger than for the heterogeneous models for all the displayed time steps and for both saturated and variably saturated cases. For the homogeneous model 99.6%–100% of the grid cells show a correlation coefficient >0.2, while for the heterogeneous model this percentage is between 8.5% and 32.5%.

The high correlation between the observations and the simulated hydraulic heads for the homogeneous model likely relates to the fact that the different observations all spatially coincide with the one homogeneous riverbed *K*, while for heterogeneous riverbeds observations provide more local information to the surrounding grid elements, and not directly relate to the entire riverbed. This highlights that for homogeneous riverbeds measurement data can be more informative to update the single (and unknown) riverbed *K*. As in the homogeneous case the representation of the riverbed is strongly simplified, the measurement data allow the identification of this unknown riverbed *K* value.

5. Discussion

The results of this study revealed that for all four geostatistical models and both the saturated and variably saturated cases, data assimilation improved the characterization of hydraulic heads. In terms of characterizing riverbed *K*, data assimilation provided an improvement only when heterogeneous models of riverbed *K* were used. Finally, in most simulation scenarios the estimation of net river-aquifer exchange fluxes was improved through data assimilation.

One important question is whether an erroneous heterogeneous/homogeneous approximation of a heterogeneous, non-multi-Gaussian riverbed with channelized structures allows characterizing the river-aquifer exchange fluxes adequately. Performance is evaluated based on the results in the verification period. Comparison is first made between an erroneous heterogeneous non-multi-Gaussian model (in this work the non-multi-Gaussian model with elliptic structures) and a correct heterogeneous non-multi-Gaussian model (here the non-multi-Gaussian model with channelized structures): Results provided in sections 4.1.2 and 4.2.2 illustrate that both for the saturated and the variably saturated case differences between the two non-multi-Gaussian models were minor, especially for the variably saturated case, as the erroneous heterogeneous non-multi-Gaussian model outperformed even the correct heterogeneous non-multi-Gaussian model. Next, the performance of the multi-Gaussian model and the (correct) non-multi-Gaussian model with channelized structures is compared: For the saturated case the average RMSE (Q) of the multi-Gaussian model is slightly larger, although five out of ten times it outperforms the correct non-multi-Gaussian model. For the variably saturated case the multi-Gaussian model performs clearly worse than the (correct) non-multi-Gaussian model with channelized structures.

These comparisons highlight that for the saturated, hydraulically connected river-aquifer systems, the error in the characterization of net river-aquifer exchange fluxes does not significantly differ between a non-multi-Gaussian model with channelized structures, a non-multi-Gaussian model with elliptic structures and a multi-Gaussian model. This is a similar conclusion as for the river-aquifer model analyzed by Tang et al. (2015), where the pattern of the heterogeneous riverbed *K* had little influence on the characterization of river-aquifer exchange fluxes. However, for variably saturated river-aquifer systems which are not hydraulically connected everywhere, the error in reproducing net river-aquifer exchange fluxes is smaller for the non-multi-Gaussian models than for a multi-Gaussian model. This implies that knowing only the mean and the variance of the riverbed *K* contains not sufficient information for estimating the net river-aquifer exchange fluxes, and knowing the histogram of the riverbed *K* provides valuable additional information. On the other hand, the small differences between the two non-multi-Gaussian models imply that information on exact patterns and connectivity is not important for improving the estimation of river-aquifer exchange fluxes. It is possible that the estimation of solute fluxes between river and aquifer would be more affected by the riverbed *K* pattern.

The homogeneous model results in larger errors than the correct non-multi-Gaussian model in terms of average RMSE (Q), although the homogeneous model outperforms the correct model three out of ten times in both the saturated and the variably saturated cases. It is somewhat surprising that for the variably saturated case the characterization of net exchange fluxes with the homogeneous model was not much worse than for the heterogeneous models, given that both hydraulic heads and riverbed *K* were significantly better characterized using the heterogeneous models. It was also expected that the heterogeneous models would outperform the homogeneous model stronger for the variably saturated case than for the saturated case (Irvine et al., 2012). Different reasons can be postulated for the relatively good performance of the homogeneous model (albeit with larger RMSE values than for the heterogeneous non-multi-Gaussian models):

1) Although a homogeneous approximation of a heterogeneous riverbed induces errors, in this study the heterogeneous approximation also induced errors, which is related to a lack of information from hydraulic head observations (both in terms of the limited information content of hydraulic heads towards reproducing fluxes, as well as in terms of the limited number of available head observations). This is one limitation of this study. In the study of Irvine et al. (2012) the homogeneous approximation of the heterogeneous riverbed was based on measured net river-aquifer exchange flux over

the studied stream reach. This is the best information possible over the studied stream reach. In this work, both for the homogeneous and the heterogeneous approximations of the true heterogeneous riverbed, only a limited amount of hydraulic head data is available. This is a key difference between the problem considered and the upscaling problem analyzed by Irvine et al. (2012).

2) The absolute correlations between measured hydraulic heads and modelled riverbed hydraulic heads were larger for the homogeneous case than for the heterogeneous cases, illustrating that for the homogeneous case measurements were more informative. Although for both the homogeneous approximation of the heterogeneous riverbed and the characterization of the heterogeneous riverbed only a limited number of point measurements was used, those point measurements were more informative to estimate a single unknown homogeneous value than to estimate spatially heterogeneous fields.
3) For the variably saturated case, for which the homogeneous approximation should be more problematic according to the findings of Irvine et al. (2012), for a large part of the riverbed the saturation condition (i.e., saturated or unsaturated) did not change over time. This implies that the sketched saturation condition in Fig. 2 was not very dynamic over time, and under such conditions the homogeneous approximation is less problematic than for highly dynamic variably saturated conditions. However, this is controlled by both the dynamic model forcings and the ratio of riverbed hydraulic conductivity and aquifer hydraulic conductivity. This is another limitation of this study.

Altogether points (1)–(3) show that for approximating a heterogeneous riverbed with a homogeneous value it is important to distinguish between the upscaling problem as outlined by Irvine et al. (2012) and the inverse problem, where typically a limited number of measurement data is available (as was the case in this study). Moreover, in this study the aquifer is simplified and assumed to be known and homogeneous over the whole simulation period. The homogeneity assumption affects the river-aquifer exchange fluxes and simplifies the parameter estimation problem. In reality, a more complex, heterogeneous aquifer is common. Therefore, further study should focus on the role of a spatially heterogeneous aquifer on the estimation of riverbed properties with different geostatistical models for those riverbed properties. In addition, further evaluation to inversely estimate riverbed properties by data assimilation is important and is currently carried out for the Emme catchment in Switzerland. This will provide further insight into understanding the role of complex heterogeneous patterns within a more dynamic and uncertain river-aquifer system. In a real-world case study, additional complicating factors are heterogeneity of the aquifer (neglected in this study), additional sources of uncertainty (e.g., boundary conditions, temporal changes in riverbed properties) and the geostatistical model which is adopted for the riverbed properties. In addition, measurements are probably not only affected by random errors, which can be larger than assumed in this study, but also by systematic errors. In general, additional sources of uncertainty might make it more complicated to estimate heterogeneous riverbed properties by data assimilation, and could also reduce further the differences between the different geostatistical models applied in this work.

6. Conclusions

We compared four geostatistical models of riverbed K for simulating river-aquifer interactions with the integrated surface water-groundwater model HydroGeoSphere. HydroGeoSphere calculates a fully coupled feedback between surface water and groundwater, and river-aquifer exchange fluxes are better approximated when an unsaturated zone is present between surface water and aquifer. In this work, the reference ("true") K field was drawn from a non-multi-Gaussian distribution with channelized structures. Four different geostatistical models were compared: a homogeneous model, a multi-Gaussian model, a non-multi-Gaussian model with elliptic structures and a non-multi-Gaussian model with channelized structures. 200 stochastic realizations of K fields were generated for each of the four geostatistical models, and served as input parameter files for model simulations and data assimilation experiments. Model simulations were done for (1) a setup with saturated conditions in and below the riverbed and (2) another setup with variably saturated conditions. Piezometric head measurement data were used for updating model states (heads) and parameters (riverbed K) with the ensemble Kalman filter. The performance was evaluated in terms of its ability to characterize hydraulic heads, riverbed K and net river-aquifer exchange fluxes. From the analysis above, we conclude:

1) Assimilation of hydraulic head data improved the characterization of hydraulic heads, riverbed K and river-aquifer exchange fluxes for all four geostatistical models, even though in three out of four cases the prior geostatistical models did not coincide with the reference geostatistical model.
2) In the case of a fully saturated river-aquifer system, both the non-multi-Gaussian and the multi-Gaussian geostatistical models outperform the homogeneous model in terms of characterization of model states, parameters and exchange fluxes. However, the differences in performance between the three heterogeneous models are minor, indicating that the spatial pattern of riverbed K has, in most cases, only a limited influence on river-aquifer exchange fluxes, similar to what has been shown for a one-way coupled model by Tang et al. (2015).
3) In the case of variably saturated conditions in and under the riverbed, conclusions are similar for the characterization of model states and parameters. However, concerning the characterization of net river-aquifer exchange fluxes results are slightly different. The performance of a multi-Gaussian K field and a homogeneous equivalent K are clearly worse than of the two non-multi-Gaussian models. The two different non-multi-Gaussian models, however, performed similarly. It can thus be concluded that for a variably saturated river-aquifer system the exact spatial pattern of the heterogeneous riverbed K is not important for characterizing river-aquifer exchange fluxes, but besides mean and variance also histogram information of the riverbed K is valuable for selecting an appropriate geostatistical model. However, this might not be true for simulating solute transport process which can be strongly affected by connectivity and complex spatial patterns of riverbed K.

Acknowledgements

The authors gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium and VSR commission on the supercomputer JURECA (Krause and Thornig, 2016) at Forschungszentrum Jülich. The first author, Qi Tang, also acknowledges the financial support from China Scholarship Council (CSC). W. Kurtz gratefully acknowledges the funding provided by the German Research Foundation (DFG) for project SFB-TR32 "Patterns in soil-vegetation-atmosphere systems: monitoring, modelling and

data assimilation". O.S. Schilling is grateful for the funding received through the Swiss National Science Foundation fellowship grant P2NEP2_171985. The authors would like to acknowledge the editors and reviewers for their time and efforts for improving the manuscript.

References

Aquanty Inc, 2016, HydroGeoSphere. A Three-dimensional Numerical Model Describing Fully Integrated Subsurface and Surface Flow and Solute Transport: Waterloo, ON, Canada.

Banks, E., Brunner, P., Simmons, C.T., 2011. Vegetation controls on variably saturated processes between surface water and groundwater and their impact on the state of connection. Water Resour. Res. 47 (11).

Blessent, D., Therrien, R., Gable, C.W., 2011. Large-scale numerical simulation of groundwater flow and solute transport in discretely-fractured crystalline bedrock. Adv. Water Resour. 34 (12), 1539–1552.

Brunner, P., Cook, P.G., Simmons, C.T., 2009. Hydrogeologic controls on disconnection between surface water and groundwater. Water Resour. Res. 45 (1).

Brunner, P., Simmons, C.T., 2012. HydroGeoSphere: a fully integrated, physically based hydrological model. Ground Water 50 (2), 170–176.

Brunner, P., Simmons, C.T., Cook, P.G., Therrien, R., 2010. Modeling surface water-groundwater interaction with MODFLOW: some considerations. Groundwater 48 (2), 174–180.

Brunner, P., Therrien, R., Renard, P., Simmons, C.T., Hendricks Franssen, H.-J., 2017. Advances in understanding river – groundwater interactions. Rev. Geophys. 55. https://doi.org/10.1002/2017RG000556.

Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Rev. 126 (6), 1719–1724.

Calver, A., 2001. Riverbed permeabilities: information from pooled data. Ground Water 39 (4), 546–553.

Camporese, M., Paniconi, C., Putti, M., Orlandini, S., 2010. Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. Water Resour. Res. 46 (2).

Camporese, M., Paniconi, C., Putti, M., Salandin, P., 2009. Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. Water Resour. Res. 45 (10), W10421.

Cheng, C., Song, J.X., Chen, X.H., Wang, D.M., 2011. Statistical Distribution of Streambed Vertical Hydraulic Conductivity along the Platte River, Nebraska. Water Resour. Manage. 25 (1), 265–285.

Delta h Ingenieugesellschaft mbH, 2006, Spring 3.3: Witten, Germany.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99 (C5), 10143–10162.

Fox, G.A., Durnford, D.S., 2003. Unsaturated hyporheic zone flow in stream/aquifer conjunctive systems. Adv. Water Resour. 26 (9), 989–1000.

Frei, S., Lischeid, G., Fleckenstein, J., 2010. Effects of micro-topography on surface–subsurface exchange and runoff generation in a virtual riparian wetland—A modeling study. Adv. Water Resour. 33 (11), 1388–1401.

Goderniaux, P., Brouyère, S., Fowler, H.J., Blenkinsop, S., Therrien, R., Orban, P., Dassargues, A., 2009. Large scale surface–subsurface hydrological model to assess climate change impacts on groundwater reserves. J. Hydrol. 373 (1), 122–138.

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multiGaussian fields, Geostatistics Troia'92. Springer, pp. 85–94.

Hendricks Franssen, H., Kaiser, H., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. Water Resour. Res. 47 (2).

Irvine, D.J., Brunner, P., Franssen, H.-J.H., Simmons, C.T., 2012. Heterogeneous or homogeneous? Implications of simplifying heterogeneous streambeds in models of losing streams. J. Hydrol. 424, 16–23.

Kalbus, E., Schmidt, C., Molson, J., Reinstorf, F., Schirmer, M., 2009. Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge. Hydrol. Earth Syst. Sci. 13 (1), 69–77.

Knapton, A., 2009. Gulf Water Study: An Integrated Surface-groundwater Model of the Roper River Catchment, Northern Territory. Water Resources Branch, Department of Natural Resources, Environment, The Arts and Sports, NT, Australia.

Kollet, S.J., Maxwell, R.M., 2006. Integrated surface–groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. Adv. Water Resour. 29 (7), 945–958.

Krause, D., Thörnig, P., 2016. JURECA: general-purpose supercomputer at Jülich supercomputing centre. J. Large-scale Res. Facilities JLSRF 2, 62.

Kurtz, W., He, G., Kollet, S.J., Maxwell, R.M., Vereecken, H., Franssen, H.-J.H., 2016. TerrSysMP-PDAF version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model. Geosci. Model Dev. 9 (4), 1341–1360.

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

Kurtz, W., Hendricks Franssen, H.J., Kaiser, H.P., Vereecken, H., 2014. Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. Water Resour. Res. 50 (2), 1665–1688.

Kurtz, W., Hendricks Franssen, H.J., Vereecken, H., 2012. Identification of time-variant river bed properties with the ensemble Kalman filter. Water Resour. Res. 48 (10).

Kurtz, W., Lapin, A., Schilling, O.S., Tang, Q., Schiller, E., Braun, T., Hunkeler, D., Vereecken, H., Sudicky, E., Kropf, P., Hendricks Franssen, H.-J., Brunner, P., 2017. Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. Environ. Modell. Software 93, 418–435.

Lackey, G., Neupauer, R.M., Pitlick, J., 2015. Effects of streambed conductance on stream depletion. Water 7 (1), 271–287.

Leek, R., Wu, J.Q., Wang, L., Hanrahan, T.P., Barbet, M.E., Qiu, H.X., 2009. Heterogeneous characteristics of streambed saturated hydraulic conductivity of the Touchet River, south eastern Washington, USA. Hydrol. Processes 23 (8), 1236–1246.

Li, Q., Unger, A., Sudicky, E., Kassenaar, D., Wexler, E., Shikaze, S., 2008. Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. J. Hydrol. 357 (3), 317–336.

Liggett, J.E., Werner, A.D., Simmons, C.T., 2012. Influence of the first-order exchange coefficient on simulation of coupled surface–subsurface flow. J. Hydrol. 414, 503–515.

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46 (11).

McDonald, M. G., Harbaugh, A. W., 1988, A Modular Three-Dimensional Finite-difference Ground-water Flow Model.

Panday, S., Huyakorn, P.S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. Adv. Water Resour. 27 (4), 361–382.

Partington, D., Brunner, P., Frei, S., Simmons, C., Werner, A., Therrien, R., Maier, H., Dandy, G., Fleckenstein, J., 2013. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resour. Res. 49 (9), 5501–5519.

Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrol. Earth Syst. Sci. 19 (7), 2999–3013.

Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics With SGeMS: A User's Guide. Cambridge University Press.

Schilling, O., Doherty, J., Kinzelbach, W., Wang, H., Yang, P., Brunner, P., 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. J. Hydrol. 519, 2258–2271.

Schilling, O.S., Gerber, C., Partington, D.J., Purtschert, R., Brennwald, M.S., Kipfer, R., Hunkeler, D., Brunner, P., 2017. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel 37Ar tracer method. Water Resour. Res. (under review)

Shrestha, P., Sulis, M., Masbou, M., Kollet, S., Simmer, C., 2014. A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. Monthly Weather Rev. 142 (9), 3466–3483.

Tang, Q., Kurtz, W., Brunner, P., Vereecken, H., Franssen, H.-J.H., 2015. Characterisation of river–aquifer exchange fluxes: the role of spatial patterns of riverbed hydraulic conductivities. J. Hydrol. 531, 111–123.

Therrien, R., McLaren, R., Sudicky, E., Panday, S., 2010. HydroGeoSphere: A Three-Dimensional Numerical Model Describing Fully-Integrated Subsurface and Surface Flow and Solute Transport: Groundwater Simulations Group. University of Waterloo, Waterloo, ON.

Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44 (5), 892–898.

VanderKwaak, J., Sudicky, E., 2000. Application of a Physically-based Numerical Model of Surface and Subsurface Water Flow and Solute Transport. IAHS Publication, pp. 515–523.

Zhou, Y., Wenninger, J., Yang, Z., Yin, L., Huang, J., Hou, L., Wang, X., Zhang, D., Uhlenbrook, S., 2013. 2013, Groundwater–surface water interactions, vegetation dependencies and implications for water resources management in the semi-arid Hailiutu River catchment, China–a synthesis. Hydrol. Earth Syst. Sci. 17 (7).