Journal of Hydrology 369 (2009) 260–273

Contents lists available at ScienceDirect

Journal of Hydrology

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

Calibration of hydraulic parameters for large-scale vertical flow constructed wetlands

Uli Maier [a,*], Cecilia DeBiase [b], Oliver Baeder-Bederski [c], Peter Bayer [d]

[a] University of Tübingen, Center for Applied Geosciences, Sigwartstr. 10, 72070 Tübingen, Germany
[b] UFZ Leipzig, Helmholtz Center for Environmental Research, Department of Environmental Microbiology, Germany
[c] UFZ Leipzig, Helmholtz Center for Environmental Research, Centre for Environmental Biotechnology (Ubz), Germany
[d] ETH Zürich, Ecological Systems Design, Institute for Environmental Engineering, Swiss Federal Institute for Technology, Switzerland

ARTICLE INFO

Keywords:
Constructed wetlands
Preferential flow
MIN3P
Inverse modeling
Evolution strategy
CMA-ES

SUMMARY

Constructed wetlands for water cleanup have been in use for several years and are promising for cost-efficient remediation of large scale contamination. Within this study, flow conditions in layered vertical soil filters used for remediation of contaminated groundwater were investigated in detail by special discharge experiments and an attuned modeling study. Unsaturated water flow was measured in two vertical flow constructed wetlands for contaminated groundwater treatment at a site in eastern Germany. Numerical simulations were performed using the code MIN3P, in which variably saturated flow is based on the Richards equation. Soil hydraulic functions based on Van Genuchten coefficients and preferential flow characteristics were obtained by calibrating the model to measured data using self-adaptive evolution strategies with covariance matrix adaptation (CMA-ES). The presented inverse modeling procedure not only provides best fit parameterizations for separate and joint model objectives, but also utilizes the information from multiple restarts of the optimization algorithm to determine suitable parameter ranges and reveal potential correlations. The sequential automatic calibration is both straightforward and efficient even if different complex objective functions are considered.

Introduction

Vertical flow constructed wetlands have been in use in domestic wastewater treatment for several years (Baeder-Bederski et al., 2004; Gross et al., 2007; Langergraber, 2008). They are implemented for removal of dissolved organics and NH_4^+, phosphorus (Lantzke et al., 1999), or fecal coliform bacteria (Baeder-Bederski et al., 2005). It has been reported that such vertical flow beds are extremely effective in removing suspended solids and biological oxygen demand (BOD_5), and for nitrification of ammonium during all seasons (e.g. Brix and Arias, 2005). The application of vertical flow constructed wetlands is also under investigation as a promising technology for on-site remediation of groundwater contaminated with aerobically degradable components. The major argument for this method is the low maintenance effort it requires, making it a very low cost water treatment technique. So far, these wetlands have been commonly designed and implemented by using simple rules of thumb such as the Austrian standard ÖNORM B 2505, 1997 or the German standard ATV-A 262, 1997 (Haberl et al., 2003) without considering and quantifying the processes occurring inside such filters in detail. More recently, however, efforts have been made to understand and quantify processes in pilot facilities, including model development, testing, and application (Cooper, 2005; Langergraber and Simunek, 2005; Werner and Kadlec, 2000; Langergraber, 2008).

Vertical soil filters are three phase systems allowing mass exchange among soil, water, and air. The chief goal is to enforce the mutual transfer of the aquatic pollutants to other environmental compartments and thereby accelerate cleanup. Compartment transfer in wetlands may be used for remediation of contaminated water through the release of fugitive gases to the atmosphere (Walter and Heimann, 2000; Whalen, 2005), active or passive uptake by plants (Nowack et al., 2006), precipitation or enhanced biogeochemical transformations (Brix and Arias, 2005).

A great advantage of vertical flow constructed wetlands is the possibility to apply wastewater intermittently. This facilitates dynamic gas exchange with the atmosphere and delivers oxygen to the contaminated water under either permanently unsaturated conditions (transport by diffusion) or periodically changing water saturation (transport by advection). The gas transport of oxygen into the wastewater is crucial for the remediation efficiency of aerobic microbial degradation. To predict the biotransformation capacity of vertical flow constructed wetlands, namely for the capacity of oxygen input and biodegradation, it is necessary to understand the flow processes that influence (i) water retention

* Corresponding author. Tel.: +49 7071 2975041.
E-mail address: uli.maier@uni-tuebingen.de (U. Maier).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.02.032

time within the filter, as well as (ii) air and water saturation over time. A fundamental task is therefore to envision variably saturated flow in high temporal and spatial resolution.

Unfortunately, unsaturated water flow characteristics are not easy to predict, as the general formulation by Richard's equation is non-linear. Within the common Van Genuchten approach, unsaturated water flow is inevitably linked to a set of empirical parameters that are hard to estimate. They can be obtained from the soil water retention measurements of small scale probes or estimated by inverse modeling at larger scales. The latter will be the focus of this paper. In this process, field scale multi-layered soil filters are intermittently supplied with water and the bottom discharge is measured. Based on these measurements and tentative available hydraulic characterization of the filters, soil models are set-up, parameterized, and calibrated. The parameter estimation procedure is adjusted to the specific features of this field case and carried out by an innovative heuristic solver, so-called evolution strategies with covariance matrix adaptation (CMA-ES, Hansen et al., 2003). Special attention is drawn to quantifying and processing measurement uncertainties when fitting the simulated discharge curves to the measured values. First inspection indicates preferential flow conditions in one of the filters, which is dealt with by way of an extended simulation approach according to Mohanty et al. (1997). The developed method of indirect parameterization of hydraulic behavior in vertical flow constructed wetlands will be essential for reliably predicting their water cleanup performance.

Langenreichenbach basins

The research pilot treatment plant Langenreichenbach is located at a field site in Saxony, Germany and has been in operation continuously since 2000. The pilot plant is used for the treatment of domestic wastewater provided by the nearby municipal sewage plant. Pre-treatment begins with a straw filter and continues with the injection of this pre-treated water through 14 filters placed at the site. The aim of the pre-treatment unit is to eliminate the suspended solids and avoid clogging of the vertical and horizontal filters used later on to treat germs (fecal coliforms, helminth eggs) and ammonia content in the water.

The vertical soil filters used in this study are composed of granular material with different grain sizes arranged in layers of varying configurations (Fig. 1). Typically, these soil filters are built up by four layers: cover layer on top, main filter layer, intermediate layer, and bottom drainage layer. The cover layer, composed of coarse material, facilitates water distribution over the entire filter surface area and protects the surface of the main layer from erosion. The intermediate layer, placed between the main layer and the drainage layer, prevents the washing out of fine particles into the drainage layer. To study the influence of two filter materials, the main layers of seven Langenreichenbach basins were filled with washed sand with a grain size of 0–2 mm (Heinrich Niemeyer GmbH & Co KG, Sprotta, Germany), while the other seven filters were filled with a mixture of expanded clay (Fibo Exclay Deutschland GmbH, Lahmstedt, Germany) and sand. The expanded clay is artificially fabricated material, composed of clays baked at very high temperature (~1000 °C) to create light pellets of clay with high specific surface and high porosity (Fig. 2a). The mixture was specifically developed for comparative tests to examine the influence of different types of filter materials. The focus of this study will be exclusively on two basins with different set-ups but mixed main layers of the same material. These basins were selected out of the 14 available filters due to their best credible measurement accuracy. Furthermore, they have the advantage of two moderate, but significantly different, loading intervals. This allows to cover a bandwidth of hydraulic conditions, and so provide the chance to most comprehensively describe flow within the materials they share.

The surface area of each filter is 6.6 m^2 (2.4 m by 2.75 m). Outlet shafts with a cross section of 0.4 m^2 are located at one corner of each basin, reducing the effective area of the soil filters to 6.2 m^2. The two vertical soil filters used in this study for the calibration of soil hydraulic properties are called basin A and basin B (Fig. 1). Both basins are planted with *Phragmites australis*, and plant roots are distributed through the entire soil filter profile. A 4-day measurement period in March 2007 was chosen for evaluation. During this period, controlled intermittent infiltration events were applied that led to distinctly different flow behavior for the two basins. There was no precipitation and minimal evaporation during the considered time period, and thus these processes could be neglected in the simulations.

The set-up of basin A is as follows (Fig. 1): on top there is a 5-cm cover layer composed of gravel (8–16 mm diameter). The underlying 60-cm thick main filter layer is comprised of expanded clay (2–4 mm diameter), which has been crushed and mixed with sand (0–2 mm diameter, Fig. 2a). A 10 cm thick intermediate layer of gravel, 4–8 mm in diameter, intersects the main layer and a

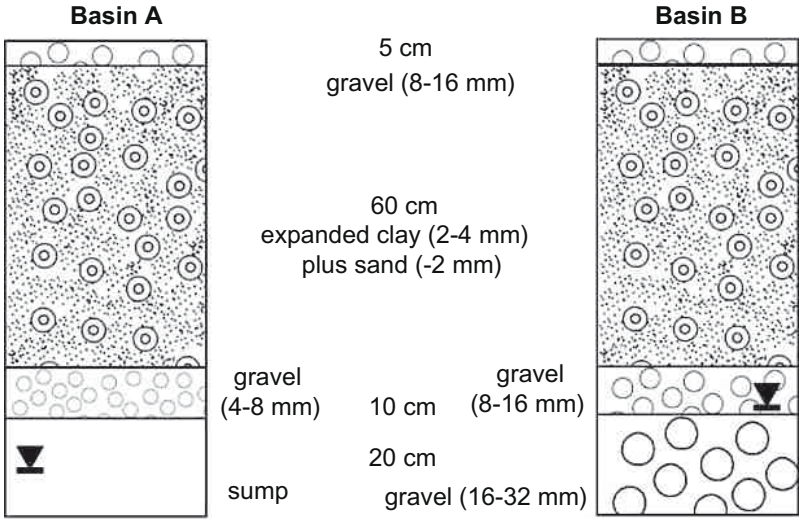

Fig. 1. Schematic set-up of selected vertical flow constructed wetlands at Langenreichenbach.

Fig. 2. Expanded clay pellets (a), sand (b) and gravel 16–32 mm (c) used in the field experiment of the two basins A and B.

20-cm sump on bottom. The intermediate gravel lies on a perforated stainless steel plate, allowing water to drain into the sump. Water was injected from the top through perforated pipes. These pipes are additionally enveloped into 100-mm slit pipes (drainage pipes) which are filled with main filter material, the clay/sand mixture. The water that reached the bottom of the basin is dammed up to a height of 10 cm in the sump, from where it was discharged to a collector pipe in the outlet shaft. A flow rate measurement port was located 5 cm from the distribution system on top. The outflow rate was measured by automatic 5-L pots about 5 m away from the basin. The sump had a minimal water level of about 10 cm. Sump overflow occured when the water height in the sump was higher than 10 cm.

Like basin A, basin B has a 5-cm cover layer on top. The main filter layer is also comparable to that of basin A, with same material and vertical length. Different is the arrangement of the two drainage layers below the main layer: the drainage element in basin B is built up by a 10 cm layer of gravel (8–16 mm diameter) and an underlying 20 cm layer of gravel (16–32 mm diameter, Fig. 2c). Water was injected from the top of the filter through a uniform areal distribution system of perforated pipes (3 mm diameter holes with a density of six holes per m^2) and collected in the drainage layer in a slotted pipe of 100 mm in diameter. The flow rate of the inflow was measured with a magnetic inductive flow meter after the metering pump prior to application to the filter and at the outflow of the basin where the water collected by the collection pipes discharged. The same instrument was used for flow measurements in the two basins. The basin presented a minimal water level of 20 cm from bottom; thus the lowest gravel layer was permanently water-saturated. When the water level exceeded 20 cm, overflow induced the discharge of the basin.

Measurement of water flow

The vertical soil filters were intermittently loaded with repeated waste water pulses, several per day of different temporal distribution. Detailed information about the loading schedules of both basins is listed in Table 1. As depicted in Fig. 3, both inflow and outflow rates were measured. The time periods used for the present study are considered as the most representative of normal operating conditions of the basins, with two distinctly different loading conditions: (i) evenly distributed intervals of infiltration roughly every 6 h in filter A and (ii) repeated, more irregular water application with rapidly succeeding events and longer dry periods according to a daily flow pattern in filter B. The flow rate during waste water injection events was 20 L/min in all cases. The outflow curves show a certain degree of variation although the inflow conditions were controlled and regularly distributed during the period of investigation. Implications will be discussed during the subsequent presentation of the calibration procedure.

Table 1
Water injection strategies for the selected Langenreichenbach basins A and B.

Basin	Volume of injected water per day (L)	Loading pulses per day (–)	Volume of water injected per pulse (L)	Duration of the pulse (min)	Loading distribution scheme
A	248	4	62	3.1	Uniform intervals
B	248	8	31	1.55	Non-uniform intervals

Numerical model

Theory

The simulation of variably saturated water flow using the numerical code MIN3P is described by the Richards equation, which is implemented in MIN3P using the formulation (Mayer et al., 2002):

$$S_a S_s \frac{\partial h}{\partial t} + \phi \frac{\partial S_a}{\partial t} - \nabla \cdot (k_{ra} \mathbf{K} \nabla h) - Q_a = 0 \quad (1)$$

with S_a as the aqueous phase saturation (–), S_S the specific storage coefficient (m^{-1}), h the hydraulic potential (m), which equals pressure head ψ plus gravitational potential z ($h = \psi + z$), ϕ the porosity (–), t time (s), $\mathbf{K}$ the tensor of hydraulic conductivity (m s^{-1}), k_{ra} the relative permeability (–) and Q_a the source/sink-term (m^3/s). Solution of the Richards equation (Eq. (1)) requires the definition of functions linking pressure head ψ to saturation and relative permeability. A well established relationship is provided by the empirical Van Genuchten–Mualem approach (Mualem, 1976; Van Genuchten, 1980), which is given by the system of Eqs. (2)–(5):

$$k_{ra} = S_{ea}^{l}\left[1 - \left(1 - S_{ea}^{1/m}\right)^{m}\right]^{2} \quad (2)$$

$$S_a = S_{ra} + \frac{1 - S_{ra}}{\left[1 + (\alpha\psi_a)^{n}\right]^{m}} \quad (3)$$

$$m = 1 - 1/n \quad (4)$$

$$S_{ea} = \frac{S_a - S_{ra}}{1 - S_{ra}} = \frac{\theta_a - \theta_{ra}}{\theta_{sa} - \theta_{ra}} \quad (5)$$

with S_{ea} (–) as the effective aqueous saturation, S_{ra} (–) the residual saturation, ψ_a matrix potential (m), l (–) a parameter of pore connectivity, m (–), α (m^{-1}) und n (–) empirical parameters, θ_a the actual aqueous, θ_{ra} the residual and θ_{sa} the volumetric water content at saturation for the soil, respectively.

Many approaches for implementation of preferential flow in variably saturated porous media models have been described in recent years (Šimůnek et al., 2003), of which the most advanced require the parallel consideration of two or more domains for the same porous medium. These distinguish between "fast" (macropore) and slow or stagnant pore space, and are coupled by kinetic exchange terms (so-called non-equilibrium flow). These approaches, however, generally come along with disadvantages, such

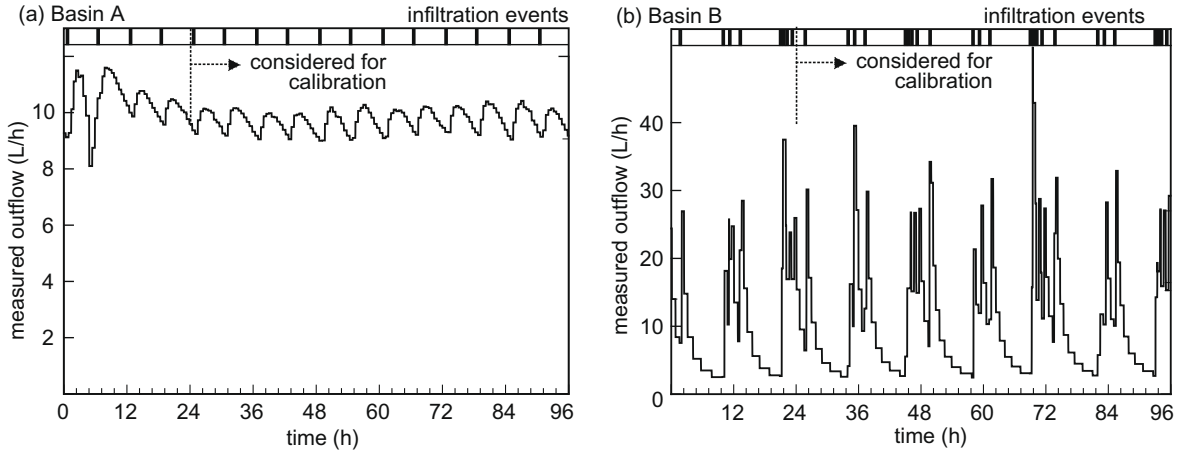

Fig. 3. Infiltration events (top) and measured outflow hydrographs from March 8–11, 2007, with neglected initial phases and later periods which are used for model calibration.

as significantly increased computational demand and a number of additional parameters which are hard to estimate under real world conditions. The alternatives of composite hydraulic functions, which increase hydraulic conductivity above a saturation threshold, do not reproduce lateral non-equilibrium of wetting and flow conditions, but are able to simulate rapid advances of seepage water fronts which are commonly observed in the presence of preferential flow. These approaches require only low computational costs, which is essential if simulations have to be repeated. Such a simplified equilibrium formulation for preferential flow is implemented in the model using the approach of Mohanty et al. (1997). This formulation necessitates only the two additional fitting parameters κ and ψ_{pf}. Hydraulic conductivity K increases exponentially if the pressure head within the unsaturated zone ψ exceeds a certain threshold pressure head ψ_{pf}:

$$K(\psi) = K(\psi) + \kappa \cdot [e^{(\psi - \psi_{pf})} - 1] \quad (6)$$

Parameter κ denotes a scaling factor of dimension (m/s) that determines the magnitude of preferential flow. The formulation used in this study is described in Gérard et al. (2004).

Model set-up and initial parameterization

Water balance error in the measurements is estimated to be ±5%. In order to moderate the impact of this error, the measured outflow data was scaled to fit the 248 L of daily infiltration in both basins. In a preliminary sensitivity analysis, the influence of the initial condition of water saturation within the filter, which is not known in detail, was evaluated by variation of initial pressure heads. The results showed that the outflow behavior depends on the unknown initial condition for approximately the first 24 h of simulation. This early time period is not taken into account for the model fitting (Fig. 3) and thus is neglected in the remainder of this paper.

The lower hydraulic model boundary condition is implemented as a fixed pressure head. It is found to influence the outflow behavior in a way that pressure heads close to saturation yield a relative permeability high enough to drain the lower parts of the filter rapidly. In filter B, however, the water level of the porous medium within the lowest gravel layer is known to be 20 cm because of the outflow device; thus this value can be directly applied to the lower boundary condition. In filter A, the water stage in the sump is below the filter level, providing a constantly unsaturated lower boundary condition. Its pressure head is generally unknown and can hardly be deduced from the soil filter measurements. When pressure head drops, unsaturated hydraulic conductivity drops and the gradient increases in the same manner, providing roughly the same flow rate under the given conditions. If the lower boundary pressure head drops below a certain value, however, the flow behavior becomes independent of the boundary condition. This threshold pressure head corresponds to the unsaturated hydraulic conductivity equal to the Darcy velocity q at the low end of filter discharge. During such temporally rather constant drainage conditions the vertical hydraulic gradient approximates one, thus hydraulic conductivity K will approximate q. This threshold value is found to be roughly $\psi = -0.5$ m for the 8–16 mm gravel (top layer in both filters and drainage layer in filter B) and -0.2 m for the 4–8 mm gravel, of which the latter value is being used as drainage layer in filter A. Consequently, this value is applied to the bottom boundary conditions in filter A.

A relatively fine vertical discretization of 1 cm was chosen. The time step was allowed to vary from a minimum value of 10^{-9} h up to 0.05 h. It is adapted by the model automatically depending on the stiffness of the hydrological problem to solve (Mayer et al., 2002).

The gravel layer (8–16 mm) overlying the main filter material in both basins is quickly passed by the infiltrating water due to its much higher hydraulic conductivity. Hence this layer can be expected to have no significant influence on the overall flow behavior. Nevertheless, it is inspected in more detail during the calibration. The water transition from the expanded clay layer to the underlying gravel may be delayed to some extent due to capillary barrier effects when passing from the finer to the coarser material (e.g. Ross, 1990). This is caused by the lower relative hydraulic conductivity in the coarser material under low water saturation. It should be noted that this delay will explicitly be accounted for in the fitted model.

Soil hydraulic parameters for filter materials used at the Langenreichenbach site were obtained from lab tests performed on material of similar grain size at the Department of Soil Physics (UFZ) and from the literature. These parameters are shown in Table 2. The values for the in situ conditions, however, cannot exactly be deduced from these measurements. For example, the main layer filter material is an aged mixture of expanded clay and sand. Furthermore, to a certain extent all soil layers will be interfused with plant roots. Therefore, an indirect method of parameterization of variably saturated flow is required. The applied automatic calibration procedure will be subsequently presented.

Table 2
Initial soil hydraulic parameters used for model simulation. Values in brackets are initial guesses and denote parameters subject to automatic calibration.

Parameters	Expanded clay[a]	Expanded clay and sand	Sand[a]	Gravel [a]	Gravel	Gravel[b]
Particle size (mm)	2–8	2–4, 0–2	0–2	4–8	8–16	16–32
Van Genuchten parameter α (1/m)	9.76	(3.5)	4.5	8.5	(12)	14.5
Van Genuchten parameter n	1.34	(2.5)	4.59	9.8	(5)	2.68
Saturated hydraulic conductivity K (m/s)	0.00028	(0.00001)	0.000286	0.0013	(0.01)	0.17
Porosity	0.4	0.27	0.3	0.35	0.3	0.3

[a] Values provided by the Soil Physics Department, UFZ Halle, Germany.
[b] Langergraber (2001).

Calibration procedure

The goal of this procedure is to adjust the MIN3P models to the layered filters through the calibration of vital unknown or uncertain parameters. Among these are the hydraulic parameters for the mixture of elapsed clay and sand, as well as the specification of the 8–16 mm gravel. This gravel type is implemented on top of both basins and below the clay/sand core in filter B. Parameter estimation problems of vadose zone models are reported to be complex, potentially having multiple equally probable, non-unique solutions and insensitive parameters (e.g. Hopmans and Šimůnek, 1999; Iden and Durner, 2007). This may be accentuated in case of multiple layers. The fitting criteria, the time series of discharge measurements, are integral signals, and thus it is hardly possible to determine the impact of individual layers. Multiple layers may be even considered as one homogeneous medium, for which a combined Van Genuchten model can be calibrated (Abbaspour et al., 2000; Bayer et al., 2005).

The Langenreichenbach basins, on the contrary, represent an exceptional situation: for two different combinations of the filter materials and under completely different hydraulic boundary conditions, discharge measurements are available. During installation, soil samples from the two different basins were taken and sieved. The sieve analysis delivered the same results. Even if after a longer operation phase hydraulic material properties slightly change, they are expected to remain in the common range of variability and uncertainties when describing unsaturated flow in natural media. Thus the task is to calibrate both filters in a way that the parameterizations are valid for both and converge to one unique solution.

Inverse modeling for estimating soil physical parameters can be done in multiple ways. The most straightforward approach is by simply fitting modeled to measured data with an efficient numerical solver that minimizes the discrepancy, i.e. the relative error or misfit. This has, for example, been demonstrated by Zijlstra and Dane (1996) for layered soils. However, the existence of multiple equally acceptable parameterizations has to be taken into account. Considering a certain degree of conceptual model uncertainty and data inaccuracy, these optimal or close-optimal parameterizations can hardly be sorted based on the fitting error alone. One way out is to use a multi-objective framework while simultaneously minimizing different types of relative errors or of other fitting criteria (e.g. Tang et al., 2006; Mertens et al., 2006). Alternatively, additional information resources may be included in Bayesian frameworks in order to decrease the number of equally probable parameterizations (e.g. Zhang et al., 2006).

Special attention must be given to the solution algorithm, which has to be capable of identifying ranges and correlations of acceptable parameter settings. Local solvers used to minimize a relative error function may be inappropriate to detect all suitable parameterizations. In case of numerous close-optimal or locally optimal solutions, the given parameter space has to be carefully examined. Repeated application of local solvers with different starting points for the optimal search can help here, but despite the computational efficiency of e.g. Multi-start Newton methods, their performance and expressiveness is highly dependent on the specific problem. In contrast to these local methods, however, more general, robust, and preferable global optimization methods have been established, most of these following heuristic concepts. Examples are Ant Colony Algorithms (Abbaspour et al., 2001), the Shuffled Complex Evolution Algorithm (e.g. Vrugt et al., 2003; Wöhling et al., 2008) and numerous singular, hybrid techniques such as those presented by Pan and Wu (1998), Lambot et al. (2002) and Iden and Durner (2007).

Frequently, global search capabilities are achieved by including stochastic search elements. This method yields algorithms potentially less prone to getting stuck in local optima but more computationally intense due to the less directed search. Recently, focus has been set on increasing the robustness of global search algorithms, in particular by including self-adaptive features, which make the algorithm adapt to the specific properties of the calibration problem. For example, by learning the model response and thus the objective function complexity during an iterative optimization procedure, the solution algorithm can more efficiently investigate the parameter space and ideally deduce second-order information and sensitivities. An example is the AMALGAM algorithm by Wöhling et al. (2008). In fact, self-adaptivity is a feature inherent to a major variant of the family of evolutionary algorithms, the evolution strategies. Recently, Fan and Casey (2008) demonstrated the suitability of their stochastic ranking evolution(ary) strategy for soil parameter identification. In our work, we employ the currently most widely used variants, the so-called evolution strategies with covariance matrix adaptation (Hansen et al., 2003). So far, inverse modeling with this algorithm has successfully been conducted for a zero-valent iron reactor model specification (Kouznetsova et al., 2007) and calibration of gene regulatory network models (Hohm and Zitzler, in press).

Evolution strategies are, similar to the more popular genetic algorithms, evolutionary algorithms that simulate natural evolution in an abstract mathematical way. The idea follows a "survival-of-the-fittest" principle, iteratively testing and comparing search points in the decision space (here model parameterizations) based on their objective function values (their fitness). This is carried out by algorithm-specific evolutionary operators that produce in a generation-wise fashion sets of search points, which tend to successively improve and converge to an optimal solution. In the evolution strategy variant used here, the initial population of λ search points is chosen randomly (with uniform probability distribution) and then separately evaluated. The next step, selection, means favoring better candidate solutions, i.e. those $\mu < \lambda$ search points with better objective function values, when creating the next population. This simulates "survival-of-the-fittest" and is an essential element for converging to an optimum. Subsequently, operators such as recombination and mutation are performed to diversify the search, thus escaping from potential local minima traps. They produce a new set, a new population of search points, which is similarly evaluated and processed as the previous one. Each new population forms a new generation, until a certain

termination criterion is met. For the application here, the total number of objective function evaluations is limited to a maximum number.

The unique feature of CMA-ES is the mutation procedure. It is carried out each generation after the best search points are selected and, during the recombination step, their (weighted) centre of mass is computed. By mutation, Gaussian noise is added, which is determined by a correlated sample distribution that is continuously adapted during the optimization procedure. More specifically, CMA-ES learns the pair-wise dependencies of the decision parameters by updating a covariance matrix of the sample distribution (Hansen et al., 2003). This way, it adapts to the structure of the objective function and quasi exploits second-order information. It is implemented in a way that this updating mechanism is independent of the coordinate system, which makes this evolution strategy an efficient and robust solver. For more details, the reader is referred to Hansen and Ostermeier (2001) and Hansen et al. (2003), who provide the entire algorithm and recommendations for setting the algorithm specific control variables. Based on these recommendations we obtain a quasi parameter-free optimization algorithm that is auspicious for heterogeneous application fields with always-unique problems such as those related to parameter estimation in hydrological models. It can handle non-separable and highly miss-scaled functions; applications in hydrology and engineering so far have shown significant global search capabilities even on multi-modal, non-smooth, discontinuous, ill-conditioned and noisy functions (e.g. Bürger et al., 2007; Bayer et al., 2008, 2009).

The measured discharge time series of both basins show a certain degree of variation even under identical inflow conditions (Fig. 3). This noise may be due to measurement errors and other unknown internal or external influences. In order to minimize the influence of noise during model calibration, both outflow curves are cut into similar time periods with equivalent infiltration protocols. For basin A, six sequential intervals with two separate infiltration events each can be distinguished. The intervals are super-positioned to obtain an interpolated average curve. More specifically, the median line was chosen as most appropriate, since it is least sensitive to extremes and outliers (Fig. 4a). This way, reliability of the discharge time series is increased, and we obtain a rather smooth curve that seems free of artifacts. For the basin A case, the root mean squared error (RMSE) is minimized:

$$F_{7b} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\bar{O}_i - P_i)^2} \tag{7}$$

The median curve is discretized into $N = 240$ equal time steps of observations $\bar{O}_i$, which are compared to the simulated discharges P_i at the same time points.

A similar procedure was applied for basin B. However, infiltration events here are not evenly distributed. After intervals of 350 min (5.8 h) and 370 min (6.2 h), respectively, the flow regime shows rather distinct double peaks. For the total measurement timeframe we obtain only three equivalent periods. As in case A, this ensemble is superimposed and the median is computed. As shown in Fig. 4b, the median denotes a heterogeneous discharge trade-off with punctually high variations. The grey area demarcates the discrepancy of discharge values that is spanned by the three curves. The most pronounced discrepancies can be observed at the two major grey area peaks that appear early before the median. This reflects a limited reliability of overlays and median curve, and in particular, apparent uncertainty in the peaks in terms of time and height. It is hardly possible to account for this during an automatic calibration. One way could be to put more weight on the apparently relatively accurate curve declines. However, trial tests showed that in this case the high peaks, which are important features of the discharge time series, could not be satisfactorily reproduced. One alternative option would be fitting to the (grey) range (Fig. 4b) instead of the median. This has been tested and found inappropriate, as the form of the measured curve was not well reproduced visually.

As a compromise, minimization of residuals as for basin A was chosen, with two modifications. First, a time tolerance was included. Slight time shifts between median observations and simulated discharge peaks cause significant errors when using RMSE based fitting. However, Fig. 4b indicates that the exact time points of these peaks are not known. In order to reflect this in the objective function, the observed median is relaxed by 6 min, reflecting a value chosen according to the observed discrepancies. Modeled discharges are compared to the median observed discharges in a range of ±6 min and the minimum residual is taken. The second modification was included after a number of trial calibrations. It turned out that the solutions found with an RMSE based objective function only had underestimated discharge peaks in common. To put more emphasis on the peaks, a penalty is included which deteriorates model parameterizations with no peak higher than 0.7 L. This value reflects the lowest median discharge peak as shown in Fig. 4b. The resulting fitness function reads

$$F_{6b} = \max\left(1, \frac{0.7\ \mathrm{L}}{P_{peak}}\right)\sqrt{\frac{1}{N}\sum_{i=1}^{N}\min(\bar{O}_{i,\pm 6\ \mathrm{min}} - P_i)^2} \tag{8}$$

where the maximization term represents the penalty factor and P_{peak} denotes the highest modeled discharge peak.

Automatic calibration requires an interface between solver program and simulation tool. The latter is iteratively called by the optimization algorithm until reasonable agreement between measured and modeled values is achieved. An automated parameter variation (APV) routine was programed that updates the MIN3P input-files, executes the model, checks performance, and evaluates

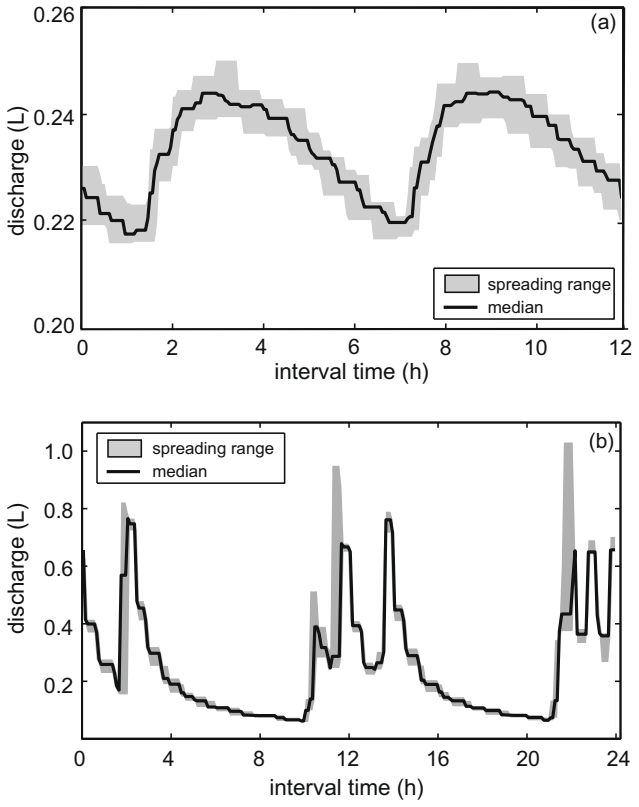

Fig. 4. Spreading range and median discharge curves for case A (a) and case B (b).

the results in one loop. First the number, keyword names of all the variable parameters, and their position in the input-file are provided from a user-defined list. Subsequently, parameter values for each run are defined and written to the input-file, proceeding in a loop as follows: (i) assignment of parameter values, (ii) update of input-file(s), (iii) call of batch file: execution of MIN3P (and/or additional external programs), (iv) check of model performance, (v) evaluation of model output.

This routine can not only be used for sensitivity analysis, but can also be coupled with any kind of parameter estimation tool such as the technique used here. Within this work, APV is coupled with a MATLAB based CMA-ES implementation that also involves the objective function formulation. Interfacing with the APV is particularly useful in case of program instabilities or if MIN3P does not converge for badly chosen parameter combinations. In such cases, the objective function value is assigned a very high value as substitute for processing model output. The selection mechanism of CMA-ES causes the evolutionary search to be deviated away from such unfavorable parameter settings.

The overall calibration procedure is as follows: the model is run for the entire application period, and in order to diminish the role of initial boundary conditions, only the last of the six or, respectively, three modeled periods is taken for the comparison with the measured median. A number of trial runs were performed to test potential calibration outcome, performance of solver and, as discussed above, potential modifications to the objective function. Based upon these trials and own expertise, ranges were defined for the parameters subject to calibration (see Tables 3–5). As demonstrated subsequently, they are modified in the course of the calibration procedure. The basin models were calibrated separately as test cases to assess CMA-ES performance and to reveal the individual best fit parameter configurations. After this, both models were calibrated together to obtain the overall best compromise of parameterization.

Due to the low computation time (a few seconds) for one MIN3P model run, it was possible to conduct several repeated applications of the automatic calibration. This is desirable in view of potential multi-modality of the objective functions. Despite the global search capabilities of the CMA-ES, it can get stuck in local optima, and thus restarts are recommendable to increase the probability of detecting the global optimum (e.g. Bayer and Finkel, 2007). However, in such parameter estimation problems, which commonly are ill-posed and based on inaccurate data, not (only) the best fitting model configuration is of interest. The question is which

Table 3
Initial parameter bounds and results of basin A model calibration. Best solution found as well as parameter ranges for a tolerance of 10% and 50% above best objective function value F_A are listed.

Basin A	Input		Output				
	Initialization		F_A = 0.00124 L (best solution)	F_A = 0.00136 L (+10%)		F_A = 0.00186 L (+50%)	
$K_{clay/sand}$ (m/s)	1.0E−07	1.0E−04	7.01E−07	6.4E−07	7.0E−07	3.9E−07	8.5E−07
$\alpha_{clay/sand}$ (1/m)	0.30	3.20	0.76	0.74	1.11	0.63	1.65
$n_{clay/sand}$	1	10	2.7	2.6	4.0	2.1	6.8
$K_{gravel(8-16)}$ (m/s)	0.005	0.80	0.0395	0.007	0.80	0.005	0.80
$\alpha_{gravel(8-16)}$ (1/m)	1	30	3.0	3.0	3.1	2.3	28.7
$n_{gravel(8-16)}$	1	30	20.8	17.1	22.6	3.0	25.2

Table 4a
Initial parameter bounds and results of basin B model calibration assuming no preferential flow. Objective function values F_B and parameter settings are shown for narrow and broad initial parameter bounds.

Basin B	Narrow parameter bounds			Broad parameter bounds		
	Input Initialization		Output F_B = 0.1507 L	Input Initialization		Output F_B = 0.124 L
$K_{clay/sand}$ (m/s)	3.9E−07	8.5E−07	8.5E−07	1.0E−07	1.0E−05	1.08E−06
$\alpha_{clay/sand}$ (1/m)	0.63	1.65	0.81	0.30	3.20	0.61
$n_{clay/sand}$	2.1	6.8	3.4	1	10	2.1
$K_{gravel(8-16)}$ (m/s)	0.005	0.80	0.8	0.005	0.80	0.005
$\alpha_{gravel(8-16)}$ (1/m)	2.3	28.7	27.3	1	30	23.5
$n_{gravel(8-16)}$	3.0	25.2	5.6	1	30	15.4

Table 4b
Initial parameter bounds and results of basin B model calibration including preferential flow. Best solution F_B and parameter ranges are shown for tolerances of 10% and 17.5% above best solution.

Basin B	Input		Output				
	Initialization		F_B = 0.1039 L (best solution)	F_B = 0.114 L (+10%)		F_B = 0.122 L (+17.5%)	
$K_{clay/sand}$ (m/s)	3.9E−07	8.5E−07	8.22E−07	4.0E−07	8.5E−07	4.0E−07	8.5E−07
$\alpha_{clay/sand}$ (1/m)	0.63	1.65	0.99	0.81	1.29	0.63	1.29
$n_{clay/sand}$	2.1	6.8	4.0	2.9	5.6	2.1	5.6
κ (m/s)	1.0E−10	1.0E−01	1.27E−07	4.54E−08	8.47E−07	2.84E−08	6.16E−05
ψ_{pf} (m)	0.1	5.0	3.8	2.7	4.3	0.1	5.0
$K_{gravel(8-16)}$ (m/s)	0.005	0.80	0.32	0.006	0.40	0.005	0.71
$\alpha_{gravel(8-16)}$ (1/m)	2.3	28.7	11.9	11.4	30.0	2.9	23.0
$n_{gravel(8-16)}$	3.0	25.2	8.8	6.8	13.0	1.0	23.0

Table 5
Initial parameter bounds and results of combined calibration of basin A and B models including preferential flow (in B). Best solution F_{AB} yields individual fits of F_A = 0.00162 L and F_B = 0.124 L. Parameter ranges are listed for tolerance of 5% above best solution.

Basins A & B	Input		Output		
	Initialization		F_{AB} = 2.0187 (best solution)	F_{AB} = 2.1 (+5%)	
$K_{clay/sand}$ (m/s)	4.0E−07	8.5E−07	4.36E−07	4.0E−07	4.5E−07
$\alpha_{clay/sand}$ (1/m)	0.80	1.29	0.82	0.81	0.91
$n_{clay/sand}$	2.4	5.6	2.7	2.4	2.8
κ (m/s)	2.84E−08	6.16E−05	1.1E−06	2.6E−07	9.3E−06
ψ_{pf} (m)	0.1	5.0	3.1	1.3	4.4
$K_{gravel(8-16)}$ (m/s)	0.005	0.71	0.043	0.008	0.59
$\alpha_{gravel(8-16)}$ (1/m)	2.9	23.0	9.7	8.4	22.9
$n_{gravel(8-16)}$	3.0	23.0	8.6	4.5	9.4

assembly of the multiple parameter settings yields acceptable solutions? This has to be ultimately decided by visual inspection, but during the evolutionary search of multiple restarts, numerous model parameter settings were assessed. In the following section, the information obtained from these "evolution paths" will be also exploited for identifying acceptable sub-optimal parameter configurations as well as for potentially existing correlations.

Application and discussion of results

Calibration of basin A model

In the first step, only the model for basin A was calibrated. The measured time series are relatively regular and spawn a uniform median that has a high chance of being reproduced by the model. The governing, unknown parameters for the Van Genuchten equations are hydraulic conductivity, K, and the empirical coefficients, α and n. These are estimated for the main element of the wetland, the clay/sand layer, as well as for the 8–16 mm gravel cover. During the optimization, the value bounds of these parameters are set according to Table 3. They are implemented in the optimization procedure according to the box constraint handling method provided by Hansen et al. (2009). For log-scaled parameters such as hydraulic conductivities, K, the $\log(K)$ values are subject to optimization in order to reflect underlying physical logarithmic relationships also in the mathematical problem formulation.

The CMA-ES is specified according to Hansen and Ostermeier (2001) and Hansen et al. (2003), which recommend a minimum population size of $\lambda = 4 + 3\ \ln(N)$ and a parent size of $\mu = \lambda/2$. Parameter N denotes the problem dimension. Both λ and μ are rounded to the next lower integer. Accordingly, for the $N = 6$ unknowns, a population size of $\lambda = 9$ and a parent size of $\mu = 4$ is used.

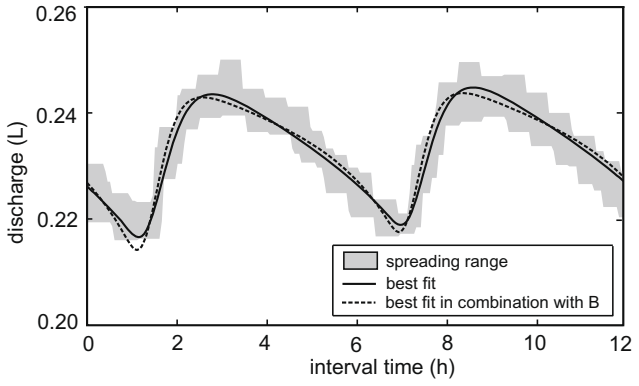

Fig. 5. Discharge curve of best fit for case A (F_A = 0.00124 L). Combined calibration with case B model yields dotted curve (F_A = 0.00162 L, F_{AB} = 2.0187).

In standard terms, the algorithm is specified as (4,9)-CMA-ES. We set a maximum generation number of 70, yielding algorithm termination after 630 objective function evaluations. The CMA-ES was applied 20 times in total, with different random seed initializations of the search points of the first generation.

The discharge curve of the overall best solution is depicted in Fig. 5. It conforms well to the median and also visually reproduces the trend of field measurement time series. Accordingly, the minimum found RMSE is only F_A = 0.00124 L. On the average, already after 330 objective function evaluations a solution of F_A = 0.00124 L + 50% = 0.00186 L was detected during the 20 CMA-ES runs. Exemplary inspection of modeled discharge curves within a range of up to 50% higher than the minimum RMSE revealed that the bulk is still within the grey area of variation. Hence, the corresponding parameter combinations can also be considered acceptable solutions.

When setting a tighter criterion of 10% above the minimum RMSE (F_A = 0.00136 L), only seven of the 20 calibration runs provided solutions, whereas the other runs failed or had not yet converged. Therefore, improvement of the optimization procedure would be possible by formulating a more appropriate termination criterion for the CMA-ES. In fact, due to the stochastic elements of the search, each run takes its own "evolution path", and it is not a priori possible to determine the ideal time to stop the algorithm. One advancement, for example, would be the inclusion of a measure that continuously evaluates the progress and convergence speed. When the fitness is no longer improved or only at a slow rate, then the procedure would be stopped. However, underlying termination criteria would have to be specified, which is delicate without insight into the optimization problem itself and without knowing the true optimal solution. In view of these obstacles, we defined an empirical maximum number of iterations, as is common in related applications. While the limit of 630 iterations may cause premature termination before sufficient convergence, this result may be compensated by applying multiple restarts.

In order to further examine the performance dynamics of CMA-ES, Fig. 6 plots the median convergence curves with 10% and 90%-confidence intervals on a log scale. The confidence intervals span a remarkable range, reflecting the variability of different individual CMA-ES runs applied to the same problem. Overall, the median denotes a fast average convergence with only slight objective function improvement after the first half of the search.

For this parameter estimation problem, it can be assumed that the solution with F_A = 0.086 L is very close to the global optimum. This is substantiated by the respective good visual fit (Fig. 5). Considering that even RMSEs of up to 50% higher yield satisfactory fits, the CMA-ES is very efficient for this type of problem. Only about every third CMA-ES run converged very close (+10%) to the minimum, which may be a sign of multi-modalities of the objective function. For comparison and also for the purpose of more insight, an alternative solution algorithm, the common Nelder–Mead

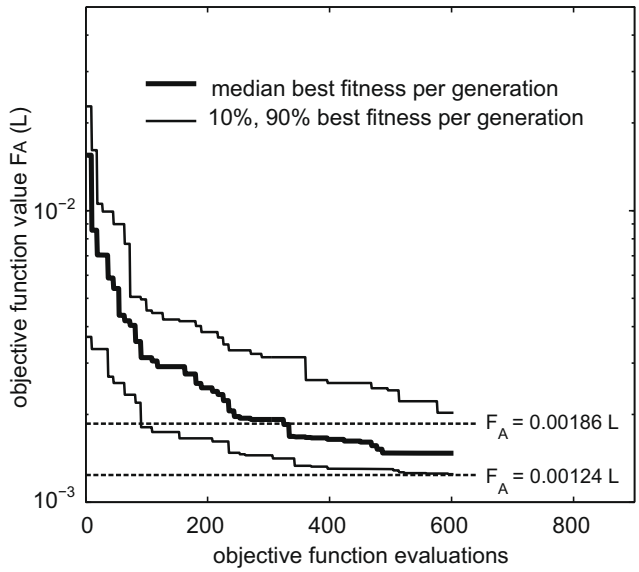

Fig. 6. Median convergence curve for calibration of model for basin A with confidence intervals. As reference, fitness values (calibration errors), $F_A = 0.00186$ L and $F_A = 0.00124$ L are given.

method (Nelder and Mead, 1965), was applied. This method basically represents a local solver and failed to converge to this minimum in most of the trial runs performed (using random initializations). We interpret this as an indication of significant objective function complexity and the existence of local optima.

The best discovered parameter combination is listed in Table 3. The calibrated hydraulic conductivity below 10^{-6} m/s corresponds to the properties of a rather fine or silty sand and is surprisingly low considering the materials used. Its value lies close to the lower boundary of the expected parameter range. This may be explained if one takes into account that the facility looks back on eight years of operation in wastewater cleanup and has experienced the effects of settling of suspended solids or colloids as well as the root growth and decay of the *P. australis* plants and organic matter accumulation, which all may lead to reduction of *K*. The crushed expanded clay pellets and surrounding sand may also form aggregates of high water retention capacity and low permeability, which would be the favorable zones of water percolation at low saturation. If, apart from that, distinct zones of high macroporous permeability do exist which are open to flow only at high saturation, they could, unfortunately, not be elucidated by this conceptual model.

The clay/sand Van Genuchten coefficients are relatively low, as can principally be expected for clay-based materials. In contrast, they are (much) higher for the overlying gravel. The gravel conductivity is estimated to be approximately $K = 0.04$ m/s, representing also a realistic value. Aside from this best solution, we also sampled all parameter configurations of satisfactory fit 10% and 50% higher than the lowest RMSE which had been detected during the evolutionary search. In detail, a total of 132 solutions were found for the 10% criterion and 1816 for the 50% criterion. The individual parameters of these configurations span substantial ranges (Table 3), which are relatively broad for the gravel parameters. This is a sign of low sensitivity of the gravel specifications, which may be attributed to its thin layer width of just 5 cm. In contrast, even if the optimality criterion is relaxed and the 50% threshold is set, the *K* values of the clay/sand mixture remain below 8E−7 m/s.

The set of sub-optimal solutions detected during the calibration procedure can be further exploited for the purpose of examining potential parameter correlations. For both thresholds, we computed the Pearson correlation matrices and inspected those parameter couples with significant correlations (correlation coefficient absolute value above 0.75). From the 15 couples, four were found; the correlation plots can be seen in Fig. 7. Please note that using this statistical inspection method is only a straightforward vehicle for extracting further information and obtaining indications from the numerous parameter settings tested. This potential is commonly overlooked and likely helpful for further interpretation of the parameter estimation problem solution. For exact determination of correlations, random samples from the search space would be necessary. Here biased samples are taken that depend on the algorithm's search history. Nevertheless, statistical representativeness is improved by considering multiple CMA-ES runs. In this sample case, the correlations already evolved after only half (10) of the CMA-ES runs. Although the CMA-ES is ideally suited for this approach, as this search algorithm is based on dynamic learning of parameter correlations, correlations can be analyzed within any other (stochastic) search procedure. Furthermore, the computed correlations are highly dependent on the defined threshold, since the threshold simply determines which local solutions are acceptable. This dependency has to be considered when setting a certain acceptance limit, but its role could also be further inspected. This will be done subsequently when we have a closer look at the four major correlations.

The Van Genuchten coefficients α and n provide good agreement to the observed field data and are highly positively correlated. Within the range of fit values shown in Fig. 7c, they yield high aqueous saturations of the porous medium for the occurring range of pressure heads in combination with either only small changes in relative permeability for the main layer. This weak sensitivity on pressure heads and thus on flowthrough rates seems to be crucial to reproduce the modest changes and smooth curves of discharge rate observed in filter A. In preliminary runs before fitting, generally much higher fluctuations were observed for most parameter combinations, unless they led to much longer delay of discharge. This implies as well that high water contents within the clay/sand mixture seem to be a prerequisite for filter outflow reproduction. Also the Van Genuchten parameters of the gravel seem to be correlated to each other. The other cross-correlations between the parameters of the different materials, between the hydraulic conductivity of the clay/sand mixture and the Van Genuchten parameters of the overlying gravel, are obvious, but can not be interpreted ad hoc. Even so, they appear to be significant, at least within the given tolerance range. However, in case of a more strict tolerance by using the 10% criterion, they tend to disappear.

As to be expected, the fitted soil hydraulic functions lead to only a slight decrease of aqueous saturation within the main layer down to about 70% if pressure head drops to −1 m. In the gravel, the saturation quickly drops to values close to residual saturation if $\psi < -0.5$ m. In the same manner, relative permeability decreases by less than an order of magnitude in the clay/sand mixture for $\psi < -1$ m and by several orders of magnitude for the gravel.

For the parameter ranges that were able to reproduce the measured outflow behavior, water contents within the main filter layer mostly remain close to saturation. Unfortunately, no measured values of water content within the filters are available, so this finding can not be proven. However, the good performance of the filters in oxygen demand removal indicates at least temporarily sufficient aeration of the filter. This suggests that, for example, a non-equilibrium water retention model using dual domain would provide a suitable alternative. Nevertheless, the given findings can be attributed to the slow velocity part of the pore space, indicating that a connected finer part of the pore space exists, which remains at relatively high water saturation during operation of the filter. Apparently, the model concept used here is appropriate for simulating these conditions.

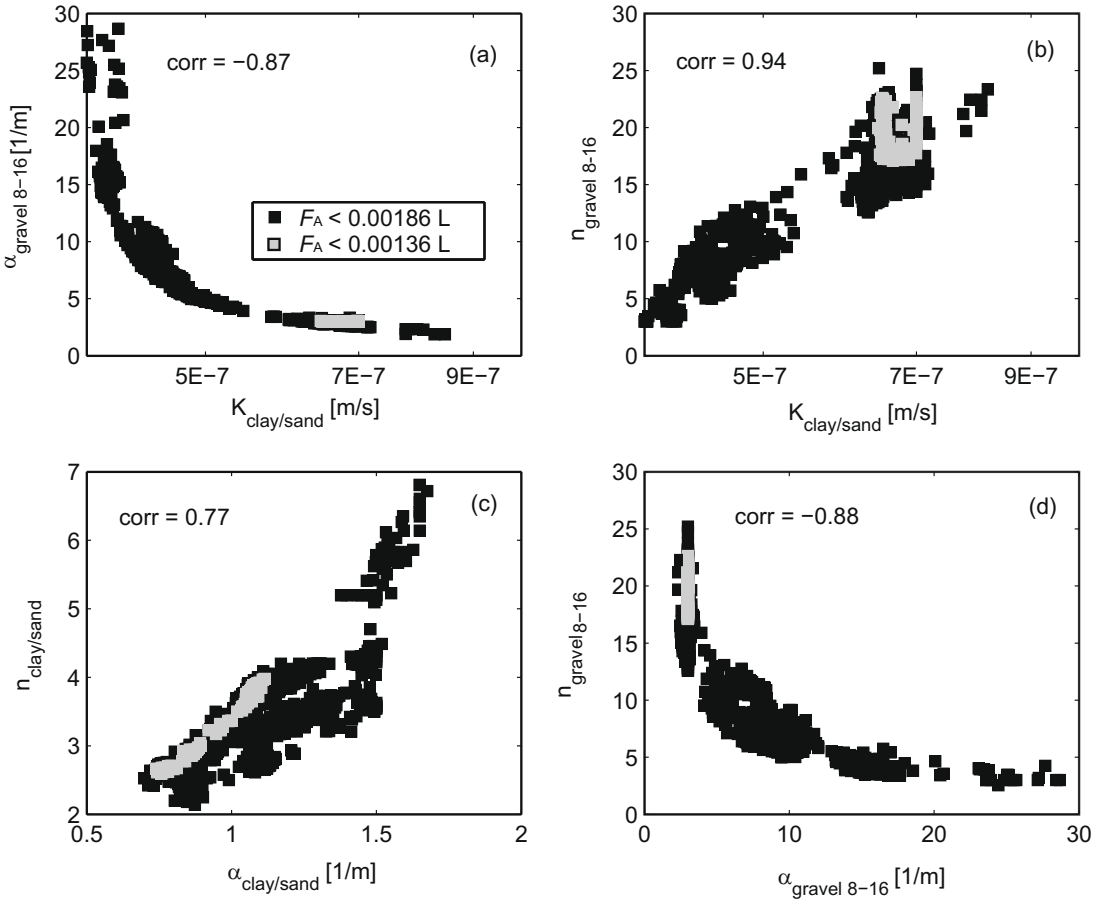

Fig. 7. Parameter correlations for A case by sampling from all parameter inspected during 20 CMA-ES runs. Shown are only those parameter pairs with Pearson correlation coefficients *corr* > 0.75.

Calibration of basin B model

As for basin A, hydraulic conductivities and Van Genuchten parameters of the clay/sand layer and of 8–16 mm gravel were estimated. The gravel here is both on top and bottom of the main clay/sand compartment (Fig. 1). The findings from filter A calibration already provide insight into possible parameter ranges of the same material, and thus the initial parameter bounds for the calibration procedure can be adjusted accordingly (Table 4a and b). The parameter ranges computed for F_A < 0.00186 L were chosen in particular. This threshold is 50% above the minimal error of the best basin A fit and spans a range of parameter configurations still acceptable after visual inspection. As in the previous case, CMA-ES was used for automatic fitting; however, even repeated applications could not deliver satisfactory solutions. In principle, this could indicate that the search algorithm is not suitable. However, numerous test runs, also with other optimization techniques, and manual tuning could not provide better model fit. Therefore the limitation is thought to be the conceptual model that can not fully reflect the actual conditions.

Minimum error assuming no preferential flow turned out to be nearly two orders of magnitude higher than for the A case (F_B = 0.1507 L, Table 4a). The best fit after 20 CMA-ES runs, which produced rather similar results, is depicted in Fig. 8. The discharge time series of the basin B are relatively irregular and show distinctive peaks. Major discrepancies are due to inappropriate reproduction of these peaks even though the objective function is formulated in a way that a 6 min deviation of the median observed values is accepted (Eq. (8)).

As the observed peaks could be a sign of draining macropores, preferential flow conditions are included in the model conceptualization for the B case. Accordingly, two more fitting parameters have to be considered, scaling factor kappa κ and pressure threshold ψ_{pf} (Eq. (6)). Similar to the treatment of hydraulic conductivity, the logarithm of κ is optimized. The population size is increased according to the new problem dimension of N = 8, and the (5,10)-CMA-ES is used for 70 generations (700 objective function evaluations) per application. After 20 randomly initialized applications, the overall best solution gave an error of F_B = 0.1039 L (Table 4b),

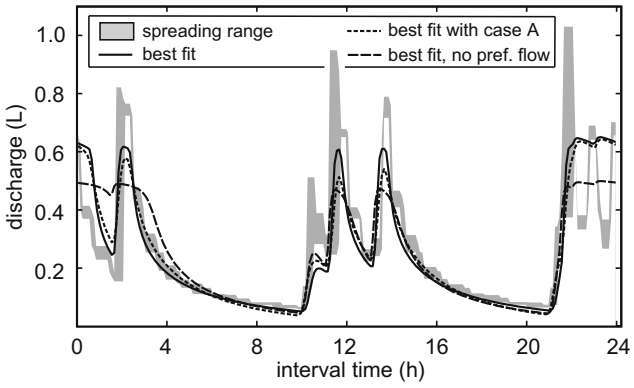

Fig. 8. Discharge curves of best fit for B case with preferential flow (F_B = 0.1039 L) and without (F_B = 0.1507 L). Combined calibration with basin A model yields dotted curve (F_B = 0.124 L, F_{AB} = 2.0187).

which is still high but about 30% lower than for the model set-up without preferential flow. Fig. 8 depicts the best modeled discharge curve, which complies well with the median observations. Although the major curve characteristics are reproduced in a much better way, the peaks are still below the estimates. Please note that none of the solutions with up to 10% worse fit than the best parameterization show peaks of higher than 0.6 L. Even the penalty approach for too low peaks could not force more pronounced inclines.

An additional indication for the two low simulated peaks is given by the high peaks somehow erratically occurring after the prolonged dry periods (discharge declines). These can be hardly explained by pressure head exceeding a threshold, when hydraulic functions in equilibrium distribution over the pore space are assumed. A formulation that accounts for either hysteresis of wetting conditions or a non-equilibrium dual porosity model might be more successful in their reproduction.

Detailed inspection of results so far and from further manipulated objective functions indicated that the most problematic element for appropriate fitting is the last phase after 22 h. Here, multiple abrupt discharge peaks occur that can hardly be simulated. The discharge peaks computed in this phase tend to be anticyclic. Due to the significant error from misfit, automatic calibration tends to favor smoothed trade-offs. More pronounced preferential flow would yield more consistent agreement with the earlier period, but disproportionally augment the error in the late phase. The reason for this difficulty could be a principal limitation of the model conceptual model of preferential flow. However, in view of the temporal discharge variations, which occur despite similar conditions in the underlying sequential experiments, the reliability of observed discharge values may be low. These complications are not observed in case A, where the infiltration is carried out more moderately and where more repeated experiments could be averaged. Discharge uncertainties of case B can be expected to be even higher than those spanned by the minimum/maximum measurements of the three experiments. Nonetheless, a higher tolerance of ± 6 min for the fitting procedure seems to be too speculative. In order to keep the parameter estimation problem as well-posed as possible and for the purpose of expressive results, no further modifications of the problem formulation are considered. Overall, the peaks of calibrated models tend to occur slightly earlier than the median observations. Since also the major measurement peaks are slightly shifted, this does not represent a shortcoming.

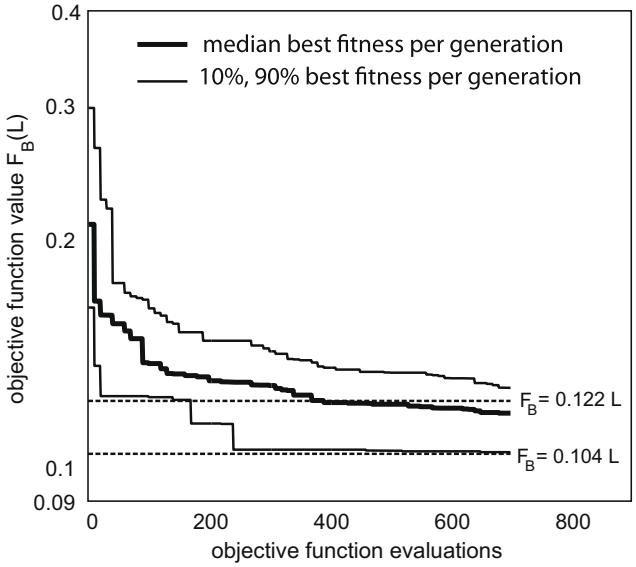

Fig. 9. Median convergence curve for calibration of model for basin B with confidence intervals. As reference, fitness values (calibration errors), F_B = 0.104 L and F_B = 0.122 L are shown.

As for the basin A results, we distinguish a strict and more moderate threshold of maximum error function values (Table 4b), which was set 10% (0.114 L) and respectively 17.5% (0.122 L) higher that the best F_B. These limits are defined after visual inspection of the preferential flow solutions and, due to the worse fit, they are relatively more stringent than for the A case. After 20 repeated runs, the CMA-ES generated heterogeneous results, with several different optimized parameter configurations.

The convergence curves of best RMSE solutions for the 20 CMA trials are examined in Fig. 9. As for the case A results (Fig. 6), the different CMA trials have unique trends, but generally reach a fast improvement in the early stage of the search. On the average, solutions for $F_B \leqslant 0.122$ L are found roughly after the first half of the search (ca. 380 objective function evaluations). Again, the confidence intervals do not widen on the log scale, which indicates that although individual CMA-ES runs may follow different search paths, the solutions found are of (increasingly) similar quality. The trends in Fig. 9 indicate that later termination of the CMA-ES after further objective function evaluations could increase the success rates. However, in this study multiple restarts are preferred to time-consuming fine-tuning in order to explore the variety of equally acceptable parameter configurations.

The optimization algorithm performance was not ideal, with 13 runs that yielded solutions of $\min(F_B) \leqslant 0.122$ L, and only five with fits $\leqslant$0.114 L. This may be interpreted as a sign of a comparatively more complex objective function and with (further) multi-modalities. In particular, the preferential flow conditions introduce additional correlations to the model parameters. This can be seen by visualization of the individual parameter correlations (Fig. 10). The same procedure as in the A case has been applied: all tested parameter configurations with objective function values that fall within the given thresholds are sampled. Pair-wise correlations are only significant between Van Genuchten parameters of the clay/sand mixture and between the preferential flow parameters (Fig. 10a). For the first, the same interpretation as for case A is suggested. Detection of the same type of correlation as in the previous experiment also confirms expressiveness and robustness of the search algorithm.

Correlations to gravel-specific parameters, as found for case A, could not be identified here. This does not necessarily mean that correlations do not exist, since the underlying sampling procedure is only a rough inspection. Moreover, in this case, the higher discrepancies between modeled and observed discharges could hide potentially further existing correlations. When water drains from a fine-grained material into a coarser one, water pressure may be insufficient to wet the larger pores associated with the coarser grains. This capillary barrier effect was observed by distinctly comparing simulations with and without drainage layer below the clay/sand core. This behavior is described by the more rapidly declining relative permeability of coarser material according to the Van Genuchten model. This connection, however present, does not appear as a visible correlation between fit parameter combination for filter B, where the 8–16 mm gravel is used as immediate layer underlying the main filter.

The correlation between the two preferential flow parameters is relatively trivial (Fig. 10b). This can be easily deduced from the Mohanty equation (Eq. (6)) which is composed of a difference of pressure heads within an exponential function. Rearranging leads to factor κ times some extra exponential function of ψ_{pf}, which may be lumped together into just a single parameter. Therefore, the findings from the solution procedure exactly match the theory.

The Van Genuchten parameters of the clay/sand mixture, which appear to be some of the most sensitive parameters, have similar

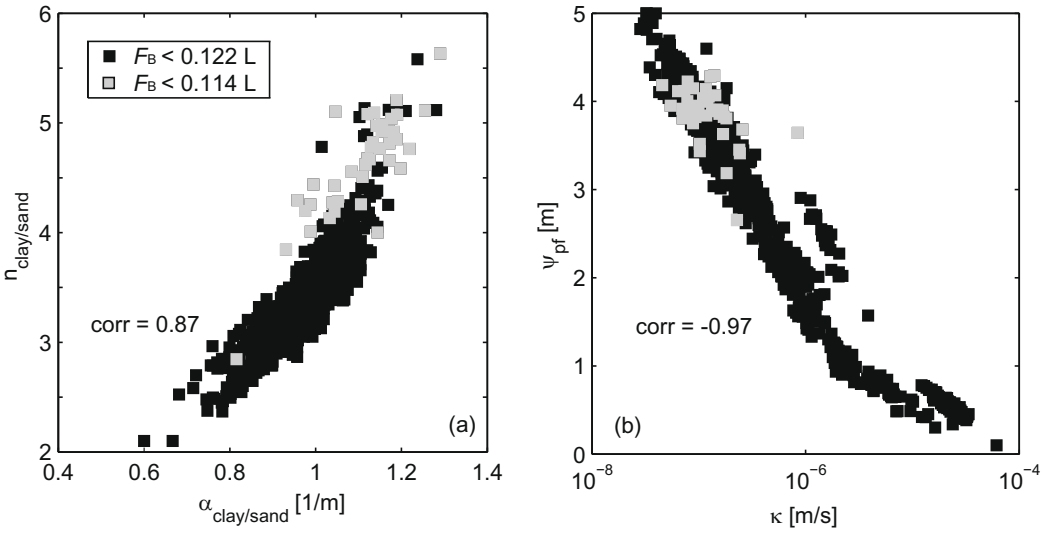

Fig. 10. Parameter correlations for case B by sampling from all parameter inspected during 20 CMA-ES runs. Selected are only those two parameter pairs with Pearson correlation coefficients *corr* > 0.75.

correlations as for case A (see Figs. 6 and 8). Their best fit values tend to be higher for the actual case B. In fact, lower values ($\alpha < 0.8$ m^{-1}, $n < 2.4$) are not favorable due to their effect on the major discharge declines (3 h–10 h, 15 h–21 h). The latter tend to be too steep for small values with significantly underestimated discharge minima. This aspect complicates efforts to derive an overall valid parameterization, although close-optimal α/n pairs seem to exist which comply with both cases. Furthermore, the best fit hydraulic conductivities of the sand/clay mixture are analogous. Highest inconsistencies can be found when comparing the gravel specifications, but the parameter ranges of close-optimal solutions (e.g. $F_A \leqslant 0.00186$ L, $F_B \leqslant 0.122$ L) are broad. This may be interpreted as their relatively low sensitivity, which seems realistic considering the low portion of gravel.

In an extra trial, we extended the initial parameter ranges for the calibration and neglected preferential flow. This was intended to determine whether alternative model parameterizations exist that are capable of reproducing the observations, assuming that no experience from the filter A can be utilized. In fact, as listed in Table 4a, a best fit solution with $F_B = 0.124$ L was found. Main feature is a hydraulic conductivity of $K = 1.08\text{E}{-6}$ m/s, which is slightly higher than that of the best fit when including preferential flow ($K = 8.22\text{E}{-7}$ m/s). This may indicate that preferential flow paths exist, but not extensively, and could be approximated by a slightly increased hydraulic conductivity. A further typical property of this approximation is a rather low value of $\alpha = 0.61$ m^{-1}, which yields relatively steep discharge declines.

Combined calibration of basins A and B models

The findings from basin-specific separate calibration can be combined to derive overall valid ranges for the individual model parameter values. However, particular parameter configurations that are acceptable for both models can only be determined by simultaneously calibrating both models. This can be done by minimizing the weighted sum of both objective functions (Eqs. (7) and (8)). The weights here are chosen based on the individual best fits. This means, the closer an individual model is to its optimal fit, the better:

$$F_{AB} = \max\left(1, \frac{F_A}{0.00186\ \text{L}}\right) + \max\left(1, \frac{F_B}{0.122\ \text{L}}\right) \quad (9)$$

Here we selected as "optimal fit" thresholds those objective function values which had been established as maximal acceptable (50% above best solution for basin A; 17.5% for basin B). The fitness value of F_{AB} is dimensionless and should denote good fits for both models at values of 2. The initial parameter bounds are set according to the overlaps between the model-specific parameter ranges (Table 5). In order to avoid too steep discharge declines of the case B simulations, the lower bound of $\alpha_{clay/sand}$ for the sand/clay mixture was increased to 0.80 m^{-1} and the other Van Genuchten coefficient was limited by $n_{clay/sand} \leqslant 2.4$.

After 20 CMA-ES applications, none of the optimized solutions yielded objective function values of $F_{AB} = 2$. This means that, within the given tolerances for each curve fit, there is a low chance that a solution exists at all. However, these tolerances are arbitrarily defined after visual inspection of the modeled discharge curves. If they are slightly relaxed and e.g. parameterizations found with $F_{AB} < 2.1$ are allowed, then a total of 447 acceptable solutions can be sampled. Such solutions are provided by only six of the 20 CMA-ES runs, indicating a moderate success rate of the optimization algorithm. Overall, the limiting case appears to be basin B, for which none of the 447 parameterizations defer a model-specific fit below the given boundary ($F_B \leqslant 0.122$ L). In contrast, some of the solutions have a case A fit of $F_A = 0.00162$ L, which lies between tolerance $F_A \leqslant 0.00186$ L and best fit found $F_A \leqslant 0.00124$ L.

The best combined solution is very close to the desired minimum and yields $F_{AB} = 2.02$ (see Figs. 5 and 7). It is characterized by a hydraulic conductivity of $K = 4.36\text{E}{-7}$ m/s, which is small and is at the lower boundary of the given value range. The Van Genuchten coefficients ($\alpha = 0.81$ m^{-1}, $n = 2.4$) are also determined close to their minimal values, which originally have been defined to circumvent too steep discharge declines for case B. Below the increased tolerance limit, the sand/clay parameterizations vary only slightly. Furthermore, within these small value ranges, still positive correlations between the Van Genuchten coefficients can be detected. Compared with the sand/clay parameters, the values for the gravel parameters span a wide range and, as expected, appear to be much less sensitive. The combined calibration narrows in particular the value range of coefficient n to its lower part ($4.5 \leqslant n \leqslant 9.4$), which has been also prioritized during case B calibration. Here, the negative correlation to the clay/sand mixture K and the gravel, as previously perceived for basin A alone, are

conserved. Apparently, the final gravel specification is constrained by interplay between both models.

Combined calibration confines also the preferential flow parameters, which are only estimated for the basin B model. Their value ranges are $1.1E{-}6 \leqslant \kappa \leqslant 2.6E{-}7$ and $1.3\ m \leqslant \psi_{pf} \leqslant 4.4\ m$, which are significant. However, as expected and revealed for case B alone, they are still strongly correlated.

Conclusions

Simulating the processes in constructed wetlands is a delicate task, since appropriate model parameterization is not straightforward. Even if their design is well known and individual material properties are assessable, the always unique field conditions and noisy measurements hamper a perfect process-based reproduction of observations in wetlands. Thus, as is common in vadose zone models, related parameter estimation problems can be ill-posed and involve insensitive parameters. Examining and interpreting these relationships is crucial to avoid a deterministic and stiff black box model. However, no general methodology has been available so far. The solution procedure for fitting simulated to observed data must be emphasized, and a sequential approach solving separated sub-problems must be preferred to a crude all at-once tactic. The study site of this work, the Langenreichenbach basins, is ideally suited for the development of a strategic parameter estimation concept. The prevailing conditions are well examined, despite common measurement errors and the inability to exactly reconstruct the entire hydrological system. The presented procedure capitalizes on a robust optimization algorithm, namely the evolution strategy, which acts as unspecific and efficient global solver for any objective function type and modification considered.

By applying two different experiments with two soil filters that are composed of the same materials, it was possible to substantially narrow the crucial material parameter values. MIN3P models were set-up and case-specific fitness functions were formulated in order to maximize the fit between observed and simulated bottom discharges of the filters. Separate calibration of each filter model yielded insight into potential parameter ranges and existing correlations. The methodology used is straightforward: the optimization routine is restarted several times and, from each run, close-optimal solutions are sampled. Evolution strategies are stochastic global solvers that explore the parameter space, with highest exploration close to the solution they converge. They certainly do not deliver a statistical inspection of sample solutions. However, statistical expressiveness can be easily improved by combining the information from multiple restarts. For the exemplary field study, several parameter correlations have been detected, of which some are clear, but others certainly not intuitive. In fact, knowledge about potentially existing correlations is crucial for handling ill-posed problems such as those common in vadose zone model parameter estimation. We used the lessons learnt from separate filter model calibration to constrain the combined calibration. Objective functions were adjusted to stress calibration objectives and to respect specific measurement conditions. The final material parameters work for both models, and even if valid in only narrow ranges, the still carry mutual correlations, like the Van Genuchten coefficients of the core clay/sand layer.

A main contribution of this work is the introduction of the CMA-ES optimization algorithm into vadose zone inverse modeling. There is definitely potential for further applications, especially for complex and multidimensional objective functions. The number of hydraulic model runs per optimization can be kept moderate, and we demonstrated that multiple restarts do not only increase the success rate, but can also be exploited for a robust statistical interpretation. For the sake of advancement, population size could be increased each new CMA-ES run, as done by Hansen et al. (2009), in order to improve the success rate on the expense of more hydraulic model runs. This could be particularly favorable if more focus is set on finding one global optimum instead of a set of potential close-optimal solutions.

Room for improvement definitely exists for the model concept when simulating the preferential flow conditions in filter B. The considered Mohanty approach is computationally efficient and a clear improvement compared to ignoring draining macropores. However, the outflow peaks were only partly reproduced and continuously predicted below the observations. Generally, the success of the calibration appears to be constrained by the equilibrium (single phase) conceptual model of preferential flow. The results of the calibration indicate that a more sophisticated preferential flow model – such as a dual permeability approach – is likely to provide an even better description of the flow response to the loadings. This will be subject to future work.

Acknowledgements

The authors acknowledge the financial support from the BMBF Helmholtz project SAFIRA II. This work was supported through the AQUATERRA project within the European Commission the 6th framework program (Contract No. GOCE 505428) and the GW-LCA project within the 7th framework program (Contract No. PIEF-GA-2008-220620). The authors also thank Margaret Hass for her invaluable help when preparing the manuscript.

References

Abbaspour, K.C., Kasteel, R., Schulin, R., 2000. Inverse parameter estimation in a layered unsaturated field soil. Soil Science 165, 109–123.

Abbaspour, K.C., Schulin, R., Van Genuchten, M.Th., 2001. Estimating unsaturated soil hydraulic parameters using ant colony optimization. Advances in Water Resources 24, 827–841.

Baeder-Bederski, O., Kuschk, P., Mosig, P., Müller, R.A., Borneff-Lipp, M., Dürr, M., 2004. Reducing faecal germs in municipal sewage using planted soil filters: initial results of a pilot plant system. Acta Horticulturae (ISHS) 643, 257–263.

Baeder-Bederski, O., Dürr, M., Borneff-Lipp, M., Kuschk, P., Netter, R., Daeschlein, G., Mosig, P., Müller, R.A., 2005. Retention of *Escherichia coli* in municipal sewage by means of planted soil filters in two-stage pilot plant systems. Water Science and Technology 51 (9), 205–212.

Bayer, P., Finkel, M., 2007. Optimization of concentration control by evolution strategies: formulation, application, and assessment of remedial solutions. Water Resources Research 43, W02410. doi:10.1029/2005WR004753.

Bayer, A., Vogel, H.-J., Ippisch, O., Roth, K., 2005. Do effective properties for unsaturated weakly layered porous media exist? An experimental study. Hydrology and Earth System Sciences 9, 517–522.

Bayer, P., Bürger, C.M., Finkel, M., 2008. Computationally efficient stochastic optimization using multiple realizations. Advances in Water Resources 31 (2), 399–417. doi:10.1016/j.advwatres.2007.09.004.

Bayer, P., Duran, E., Baumann, R., Finkel, M., 2009. Optimized groundwater drawdown in a subsiding urban mining area. Journal of Hydrology 365 (1–2), 95–104. doi:10.1016/j.jhydrol.2008.11.028.

Brix, H., Arias, C.A., 2005. The use of vertical flow constructed wetlands for on-site treatment of domestic wastewater: New Danish guidelines. Ecological Engineering 25 (5), 491–500.

Bürger, C.M., Bayer, P., Finkel, M., 2007. Algorithmic funnel-and-gate system design optimization. Water Resources Research 43, W08426. doi:10.1029/2006WR005058.

Cooper, P.F., 2005. The performance of vertical flow constructed wetland systems with special reference to the significance of oxygen transfer and hydraulic loading rates. Water Science and Technology 51 (9), 81–90.

Fan, Z., Casey, F.X.M., 2008. Estimating solute transport parameters using stochastic ranking evolutionary strategy. Vadose Zone Journal 7, 124–130.

Gérard, F., Tinsley, M., Mayer, K.U., 2004. Preferential flow revealed by hydrologic modeling based on predicted hydraulic properties. Soil Science Society of America Journal 68 (5), 1526–1538.

Gross, A., Shmueli, O., Ronen, Z., Raveh, E., 2007. Recycled vertical flow constructed wetland (RVFCW) – a novel method of recycling greywater for irrigation in small communities and households. Chemosphere 66 (5), 916–923.

Haberl, R., Grego, S., Langergraber, G., Kadlec, R., Cicalini, A., Dias, S., Novais, J., Aubert, S., Gerth, A., Thomas, H., Hebner, A., 2003. Constructed wetlands for the treatment of organic pollutants. Journal of Soils and Sediments 3 (2), 109–124.

Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9 (2), 159–195.

Hansen, N., Müller, S.D., Koumoutsakos, P., 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evolutionary Computation 11 (1), 1–18.

Hansen, N., Niederberger, A.S.P., Guzzella, L., Koumoutsakos, P., 2009. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. IEEE Transactions on Evolutionary Computation 13 (1), 180–197.

Hohm, T., Zitzler, E., in press. Multicellular pattern formation: parameter estimation for ode-based gene regulatory network models. IEEE Engineering in Medicine and Biology Magazine.

Hopmans, J.W., Šimůnek, J., 1999. Review of inverse estimation of soil hydraulic properties. In: Van Genuchten, M.Th., et al. (Eds.), Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media. Proceedings of the International Workshop, Riverside, CA, 22–24 October, 1997. U.S. Salinity Laboratory, Riverside, CA, pp. 643–659.

Iden, S.C., Durner, W., 2007. Free-form estimation of the unsaturated soil hydraulic properties by inverse modelling using global optimization. Water Resources Research 43. doi:10.1029/2006WR005845.

Kouznetsova, I., Bayer, P., Ebert, M., Finkel, M., 2007. Modelling the long-term performance of zero-valent iron using a spatio-temporal approach for iron aging. Journal of Contaminant Hydrology 90 (1–2), 58–80.

Lambot, S., Javaux, M., Hupet, F., Vanclooster, M., 2002. A global multilevel coordinate search procedure for estimating the unsaturated soil hydraulic properties. Water Resources Research 38 (11), 1224. doi:10.1029/2001WR001224.

Langergraber, G., 2001. Development of a simulation tool for subsurface flow constructed wetlands. Wiener Mitteilungen. Band 169. Dissertationsarbeit zur Erlangung des Doktorgrades der Universität für Bodenkultur Wien.

Langergraber, G., 2008. Modeling of processes in subsurface flow constructed wetlands: a review. Vadose Zone Journal 7 (2), 830–842.

Langergraber, G., Simunek, J., 2005. Modeling variably saturated water flow and multicomponent reactive transport in constructed wetlands. Vadose Zone Journal 4 (4), 924–938.

Lantzke, I.R., Mitchell, D.S., Heritage, A.D., Sharma, K.P., 1999. A model of factors controlling orthophosphate removal in planted vertical flow wetlands. Ecological Engineering 12 (1–2), 93–105.

Mayer, K.U., Frind, E.O., Blowes, D.W., 2002. Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions. Water Resources Research 38 (9), 1174–1195.

Mertens, J., Stenger, R., Barkle, G.F., 2006. Multiobjective inverse modeling for soil parameter estimation and model verification. Vadose Zone Journal 5, 917–933.

Mohanty, B.P., Van Genuchten, M.T., Bowman, R.S., Hendrickx, J.M.H., 1997. New piecewise-continuous hydraulic functions for modeling preferential flow in an intermittent-flood-irrigated field. Water Resources Research 33 (9), 2049–2063.

Mualem, Y., 1976. A new model predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research 12, 513–522.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Computer Journal 7, 308–313.

Nowack, B., Mayer, K.U., Oswald, S.E., van Beinum, W., Appelo, C.A.J., Jacques, D., Seuntjens, P., Gerard, F., Jaillard, B., Schnepf, A., Roose, T., 2006. Verification and intercomparison of reactive transport codes to describe root-uptake. Plant and Soil 285 (1–2), 305–321.

Pan, L.H., Wu, L.S., 1998. A hybrid global optimization method for inverse estimation of hydraulic parameters: annealing-simplex method. Water Resources Research 34, 2261–2269.

Ross, B., 1990. The diversion capacity of capillary barriers. Water Resources Research 26 (10), 2625–2629.

Šimůnek, J., Jarvis, N.J., Van Genuchten, M.Th., Gärdenäs, A., 2003. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. Journal of Hydrology 272, 14–35.

Tang, Y., Reed, P.M., Wagener, T., 2006. How efficient and effective are multiobjective evolutionary algorithms for hydrologic model calibration? Hydrology and Earth System Sciences 10, 289.

Van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44, 892–898.

Vrugt, J.A., Gupta, H.V., Bouten, W., Bastidas, L.A., Sorooshian, S., 2003. Effective and efficient algorithm for multi-objective optimization of hydrologic models. Water Resources Research 39 (8), 1214. doi:10.1029/2002WR001746.

Walter, B.P., Heimann, M., 2000. A process-based, climate-sensitive model to derive methane emissions from natural wetlands: application to five wetland sites, sensitivity to model parameters, and climate. Global Biogeochemical Cycles 14 (3), 745–765.

Werner, T.M., Kadlec, R.H., 2000. Wetland residence time modeling. Ecological Engineering 15, 77–90.

Whalen, S.C., 2005. Biogeochemistry of methane exchange between natural wetlands and the atmosphere. Environmental Engineering Science 22 (1), 73–94.

Wöhling, Th., Vrugt, J.A., Barkle, G.F., 2008. Comparison of three multiobjective optimization algorithms for inverse modeling of vadose zone hydraulic parameters. Soil Science Society America Journal 72, 305–319.

Zhang, D., Beven, K.J., Mermoud, A., 2006. A comparison of non-linear least square and GLUE for model calibration and uncertainty estimation for pesticide transport in soils. Advances in Water Resources 29, 1924–1933.

Zijlstra, J., Dane, J.H., 1996. Identification of hydraulic parameters in layered soils based on a quasi-Newton method. Journal of Hydrology 181, 233–250.