Optimization of concentration control by evolution strategies: Formulation, application, and assessment of remedial solutions

Peter Bayer[1] and Michael Finkel[1]

Received 22 November 2005; revised 5 September 2006; accepted 21 September 2006; published 10 February 2007.

[1] The uniqueness and mathematical complexity of typical groundwater remediation or control problems involving numerical models necessitate appropriate solvers that find optimal solutions reliably and within reasonable computational time. The aim of this paper is to introduce an innovative evolutionary algorithm, the evolution strategies with covariance matrix adaptation and rank μ update (CMA-ES), used as an external solver in combination with groundwater transport models. A broad range of hypothetical pump-and-treat design problems is set up to derive recommendations for a robust CMA-ES configuration. One- and five-well cases are distinguished, for which both extraction rates and well positions have to be optimized in order to partly capture a contaminant plume. Intrinsic natural attenuation of the contaminant that emanates from a continuous source is considered. During this study, it is revealed that when pumping rates are minimized, well positions should be focused close to the source. In contrast, when contaminant mass extraction is minimized, positions distant from the source close to the plume fringes prove to be preferable.

Citation: Bayer, P., and M. Finkel (2007), Optimization of concentration control by evolution strategies: Formulation, application, and assessment of remedial solutions, *Water Resour. Res.*, *43*, W02410, doi:10.1029/2005WR004753.

1. Introduction

[2] The development of combined simulation and optimization methods to detect ideal groundwater management options has evolved as an important field in water resources research, particularly as it directly links physicochemical process description and economic planning. Simulation tools, such as groundwater models, are used to determine and predict the response of groundwater systems to technical measures. Optimization algorithms automate the search for optimal managerial strategies regarding technical, economic, or other objectives, such as ecological or social purposes expressed by decision makers. In spite of the work that has been done in this area and the huge number of theoretical studies available, its acceptance is, in practice, rather low. Still, scientific objectives are at the forefront of most real world applications [e.g., *Marryott et al.*, 1993; *Gailey and Gorelick*, 1993; *Zheng and Wang*, 2002]. However, the use of mathematical optimization methods, rather than common trial-and-error approaches, for the selection of preferable management solutions has demonstrated an enormous potential to save costs by facilitating better technical system design. In their demonstration projects on hydraulic optimization of existing pump-and-treat systems (PTS) at three contaminant sites, the *U.S. Environmental Protection Agency* [1999] identified possible savings of millions of dollars. Technical decision criteria included the positions and pumping rates of wells, which were redesigned to reach low-cost solutions while achieving hydraulic control objectives (i.e., capture of groundwater contaminations).

[3] The U.S. Environmental Protection Agency and Department of Defense (US DoD) [*Zhang et al.*, 2004] successfully initiated the application of both flow and transport optimization codes to three real world PTS cases, with the purpose of testing the capabilities of selected methods for practical use. Compared only with hydraulic-based design, the use of transport simulation enables the prediction of the spatial and temporal effects of remedial actions on the distribution of contaminant mass and concentrations in the subsurface. Simultaneously, the computational time for evaluating one technological variant (e.g., a PTS pumping scheme) is higher. Modern optimization algorithms require iterative model calls, which commonly increase with problem complexity and dimensionality. However, compared to trial and error, the most benefits from mathematical optimization are observed for complex and high-dimensional problems [*Zhang et al.*, 2004]. This aspect substantiates the need for concepts that are especially suited for these kinds of problems.

[4] Each groundwater management problem is unique in that it typically involves individual boundary conditions, and in some cases, multiple objectives. Though the majority of studies deal with the design of well systems for contaminated site management or water supply [*Freeze and Gorelick*, 1999; *Mulligan and Ahlfeld*, 1999; *Cheng et al.*, 2000; *Vink and Schot*, 2002; *Mayer et al.*, 2002; *McPhee and Yeh*, 2004], several related research fields have evolved in which mathematical optimization is utilized to determine

[1]Center for Applied Geosciences, University of Tübingen, Tübingen, Germany.

0043-1397/07/2005WR004753

ideal policies, such as for watershed management [*Srivastava et al.*, 2002; *Veith et al.*, 2003; *Feyen and Gorelick*, 2005], groundwater monitoring [*Cieniawski et al.*, 1995; *Reed et al.*, 2000], or the assessment of alternative remediation technologies [*Marryott*, 1996; *Bayer et al.*, 2001; *Liu and Minsker*, 2004; *Shieh and Peralta*, 2005]. Among the many optimization algorithms available, there is a clear trend toward soft computing heuristic search techniques, which are regarded as superior to classic gradient-based methods for nonconvex problems with several local optima and for which a reduced time complexity is deduced. In fact, nonconvexity of objective functions is common for groundwater management problems [*Ahlfeld and Sprong*, 1998; *Zheng and Wang*, 1999; *Aly and Peralta*, 1999; *Bayer and Finkel*, 2004]. Heuristic solvers have also proven to be robust in tackling problems involving objective function discontinuities. Their performance is less sensitive to the nature of decision variables, type of constraints and noise, as has been reported in relation to several groundwater management formulations [*Rogers and Dowla*, 1994; *Guan and Aral*, 1999; *Smalley et al.*, 2000; *Orr and Meystel*, 2005].

[5] From a practical perspective, the universal and flexible applicability of heuristics is appealing. One optimization algorithm can be used for manifold problems, independent of hydraulic characteristics of a system, the relevant contaminant transport processes at a site, and simulation and evaluation models used [*Marryott*, 1996; *Zheng and Wang*, 1999; *Ranjithan*, 2005]. After a problem-specific objective function is set up and the decision variables are selected, groundwater models are iteratively called to evaluate a number of management system design alternatives. The decision space is searched in a guided manner, instead of enumerating all possible solutions and, by doing so, the computational burden of running demanding flow and/or transport models can be significantly reduced. However, the number of model runs may still be high, especially for poorly scaled and high-dimensional problems with a large number of decision variables. If we define the performance measure as the number of model calls required in finding an optimum, or at least near-optimum solution, which fulfills a predefined condition (e.g., reduction of total pumping rate of PTS to a certain level), then significant differences can be observed with respect to this criterion among the different solution methods. Despite the fact that there is no optimization algorithm that is universally most efficient [*Wolpert and Macready*, 1997], the uniqueness of real world problems must not result in the application of unique optimization procedures to achieve a higher level of acceptance in practice. Instead, suitable unspecific solution methods should be provided, with a predictable and satisfactory performance. This means that they should be reliable in finding, or at least sufficiently approaching, an optimal solution at appreciable computational expense.

[6] In recent papers [*Bayer and Finkel*, 2003, 2004], we demonstrated the applicability and robustness of derandomized evolution strategies (CMA-ES) [*Hansen and Ostermeier*, 2001] for the capture zone adaptation of extraction wells. Especially for higher problem dimension, that is, if more than one well is to be optimized, evolution strategies were found to be preferable to commonly used simple genetic algorithms (SGA). A significantly lower number of model runs were required in finding optimal pumping well schemes to capture a prespecified zone in a hypothetical, highly heterogeneous aquifer. In the presented study, CMA-ES with rank μ update [*Hansen et al.*, 2003] will be employed for PTS optimization by using concentration-based transport models. The purpose is not only to present a tool with which to solve such a "concentration control" [*Mulligan and Ahlfeld*, 1999] problem, but to devise recommendations for the ideal CMA-ES configuration, which are valid for typical groundwater management problems. Focusing on PTS is aligned with the majority of previous studies, which adopted this remediation technology as a benchmark as it allows for the testing of optimization methods for their applicability to problems with archetypal features for groundwater management problems.

[7] In *Bayer and Finkel* [2004], a moving well formulation was presented and solved by utilizing a numerical groundwater model, particle tracking and a former CMA-ES implementation by *Hansen and Ostermeier* [2001]. It was shown that solutions for problems with one to eight wells can be obtained with up to one tenth of the function evaluations required in cases where a common GA variant is applied. Though only one hypothetical example site was considered, the results of this study indicated that nonconvex and high-dimensional problems are particularly tractable at a reasonable number of groundwater model runs. Compared to the other heuristic solver, CMA-ES could more reliably identify global optima and proved to be a distinctively robust algorithm [*Bayer and Finkel*, 2004].

[8] This paper is organized as follows: First, the CMA-ES with rank μ update is introduced and the algorithm-specific procedural components are explained. Then a hypothetical contaminated site is set up and implemented in a finite element model. A continuous source feeds the groundwater with a constant contaminant concentration. The naturally occurring or synthetically accelerated degradation capacity of the aquifer is assumed insufficient in ensuring that contaminant concentrations are below the maximum allowable level at a certain distance from the source. The optimization problem is formulated in order to identify a pumping well scheme for a PTS that ideally prevents the plume from reaching the downgradient compliance line, with the plume fringe defined as the isoline of the desired concentration threshold. By variation of the natural attenuation capacity and the spatial conductivity distribution of the aquifer, scenarios (one homogeneous and the ensemble of 100 different heterogeneous aquifers) are defined. The effect of optimization criteria, minimization of pumping rates and minimization of mass extraction rate are analyzed: Hundreds of different CMA-ES optimization runs are applied to derive generally valid recommendations for an appropriate configuration of this optimization algorithm. The obtained well schemes are examined to extract typical characteristics of optimal solutions. These characteristics or trends may be important for the reduction of the decision space or even in anticipating proper well schemes without mathematical optimization when information from the site is scarce. Because of the enormous computational effort required in this analysis, the focus is set on configuring one-well extraction systems within a predefined well placement

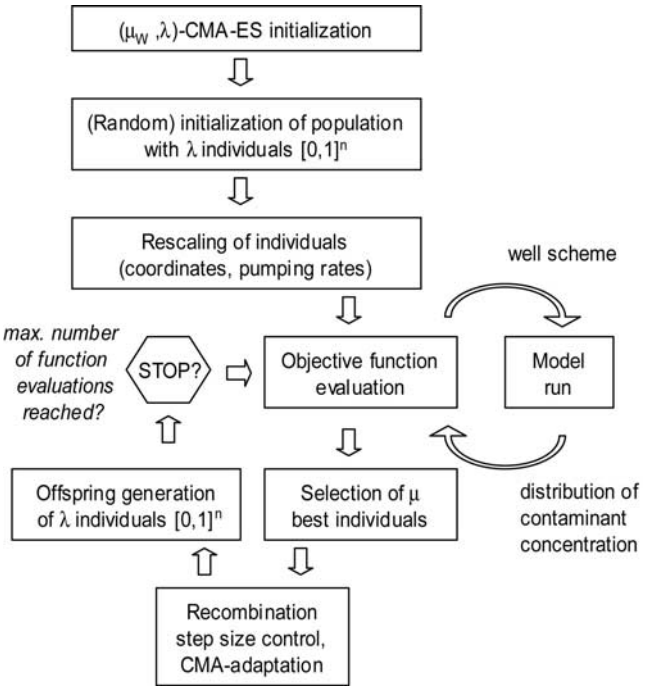

Figure 1. Application of CMA-ES for one optimization attempt (OR): After initialization of the optimization algorithm by means of a randomly generated first population, offspring generations are produced until a prespecified stopping criterion is met (here: Is the maximum number of function evaluations reached?).

area. Additionally, as an example of multiple well problems, the CMA-ES is used to optimize a five-well PTS for five heterogeneous aquifer realizations to obtain insights into the algorithm's performance for these kinds of high-dimensional concentration control problems.

2. CMA-ES With Rank μ Update and Weighted Recombination

[9] Evolutionary algorithms such as GA and ES are iterative, direct and stochastic search algorithms, which have been developed parallel to one another since the 1960s [*Schwefel*, 1995]. As indicated by their names, these methods are oriented toward natural evolution, mimicking characteristic features of Darwin's evolution theory in order to solve control and optimization problems. Similar to common canonical GA, (μ,λ)-CMA-ES [*Hansen and Ostermeier*, 2001; *Hansen et al.*, 2003] and ES generally work on the basis of populations. A population represents a number (μ) of n-dimensional, real-valued, decision variable vectors, so-called individuals, which in our case are well schemes. After the individuals of a population are independently assessed by running the model and evaluating the objective (fitness) function, the μ best individuals are selected as the origin (the parents) for a new population (the offspring). Every population denotes a generation, which is iteratively created until a stopping criterion, for example, a maximum number of model runs, a certain objective value limit or stagnation, is reached.

[10] The immense quantity of research in the field of evolutionary computation continuously produces new variants of evolutionary optimization algorithms. Many of these are hybrid algorithms, which hinder a clear categorization. A general characteristic of CMA-ES, in contrast to the binary representation inherent to most GAs, is that it operates on real-valued decision parameters. Therefore, compared to the most common GA variant in our area, the simple GA [e.g., *Rogers and Dowla*, 1994; *Aly and Peralta*, 1999], different evolutionary operators are implemented to generate offspring. Recently, however, promising real-valued GA variants have also evolved, such as the Improved Real-Coded GA by *Yoon and Shoemaker* [2001] or the Generalized Generation Gap (G3) GA by *Deb et al.* [2002].

[11] As for evolution strategies in general, the main procedure for the CMA-ES is mutation. This involves adding mutation steps—realizations of a random vector from a Gaussian distribution around the parents—to create new individuals. CMA-ES utilizes a covariance matrix to rotate and scale the mutation distribution. This procedure is controlled by several CMA-ES specific strategy parameters, which can evolve during the evolutionary search, yielding a self-adaptive algorithm. By iteratively updating the covariance matrix, the probability of reproducing the mutation steps that yielded the actual population is increased. Repeated selection of the best individuals then forces the algorithm to converge at an optimum. Several ES schemes have been developed with partly stochastic control of the mutation procedure [e.g., *Schwefel*, 1995]. The derandomized ES have a deterministic implementation and utilize information over a number of generations (the evolution path), instead of only using the previous one. Therefore an iterative improvement of search points is not only attained by processing the information of the best λ individuals per generation, but also by analyzing the successful search path, which codetermines the mutation steps for producing the subsequent offspring.

[12] The CMA-ES is initialized by an arbitrary or user-specified first population of $k = 1, .., \lambda$ individuals (i.e., search points) $\mathbf{x}_{k,1}$ representing generation $t = 1$ ($t = 1, .., t_{\max}$). After evaluating the objective function, the μ best are chosen as parental vectors ("selection", Figure 1). In one option, their center of mass $\langle \mathbf{x} \rangle_{\mu,t+1}$ is computed as the origin for the offspring. This is termed intermediate recombination, which corresponds to an equal weighting of all selected individuals, irrespective of their (relative) fitness (expressed by "(μ_I,λ)-CMA-ES"). Here, as a modification, we used weighted recombination ("(μ_W,λ)-CMA-ES"), which slightly shifts the origin in favor of better individuals. Alternative recombination operators that particularly differ in the determination of the origin for the offspring are discussed by *Deb et al.* [2002]. For example, the G3-specific parent-centric recombination (PCX) uses individually selected parents, rather than their center of mass, to produce the next generation.

[13] After the selection and recombination step in the CMA-ES, mutation is carried out by adding a normally distributed random vector with zero mean and adaptive covariance. Specifically, the transition from one generation t to the next $t + 1$, (i.e., offspring generation) (Figure 1), is formally described by

$$\mathbf{x}_{k,t+1} = \langle \mathbf{x} \rangle_{\mu,t} + \sigma_t \mathbf{B}_t \mathbf{D}_t \mathbf{z}_{k,t+1} \quad (1)$$

With weighted recombination, $\langle \mathbf{x} \rangle_{\mu,t}$ is computed by the weights $\omega_{k'}$. These add up to 1 or are normalized by their sum to obtain total effective weights of 1:

$$\langle \mathbf{x} \rangle_{\mu,t} = \frac{\sum_{k'=1}^{\mu} \omega_{k'} \mathbf{x}_{k',t}}{\sum_{k'=1}^{\mu} \omega_{k'}} \quad (2)$$

given that

$$\omega_{k'} = \ln(\mu + 1) - \ln(k') \quad (3)$$

and $k' = 1, \ldots, \mu$ is the index of the best individuals per generation, sorted in ascending order with respect to their fitness (for minimization). According to equation (1), λ base points, $\mathbf{z}_{k,t+1} \sim N(0, \mathbf{I})$, are created for the next generation. These are independent realizations of an n-dimensional normally distributed random vector with a covariance matrix equal to the identity matrix $\mathbf{I}$ and with a zero mean, $N(0, \mathbf{I})$. The base points are rotated and scaled by the eigenvectors $\mathbf{B}_t$ and the square root of the eigenvalues $\mathbf{D}_t$ of the covariance $\mathbf{C}_t$. Utilizing the covariance that is adapted during the evolutionary search (see below) yields a strategy that is invariant against any linear transformation of the search space. Strategy parameter σ_t represents the global step size, which is also adapted during each generation and scales the distribution. The obtained points are then grouped around the new origin, $\langle \mathbf{x} \rangle_{\mu,t}$. The step size control rule for σ_{t+1} reads

$$\sigma_{t+1} = \sigma_t \exp\left[\frac{c_\sigma}{d_\sigma}\left(\frac{\|\mathbf{p}_{\sigma,t+1}\|}{\mathrm{E}\,\|N(0,\mathbf{I})\|} - 1\right)\right] \quad (4)$$

$$\mathbf{p}_{\sigma,t+1} = (1 - c_\sigma)\mathbf{p}_{\sigma,t} + \sqrt{c_\sigma\,(2 - c_\sigma)\mu_{eff}}\;\mathbf{C}_t^{-\frac{1}{2}}\;\frac{\langle \mathbf{x} \rangle_{\mu,t+1} - \langle \mathbf{x} \rangle_{\mu,t}}{\sigma_t} \quad (5)$$

Equation (5) indicates that the global step size σ_t is updated to σ_{t+1} by referring to $\mathbf{p}_{\sigma,t+1}$, which represents the "conjugate evolution path." Utilizing the evolution path means that the steps of previous generations are collected in order to obtain information about the overall search direction. *Hansen and Ostermeier* [2001] call this procedure cumulation, which is an efficient adaptive method of increasing or reducing the global step size according to the examined fitness landscape. The evolution path, $\mathbf{p}_{\sigma,t+1}$, is calculated as the exponentially smoothed sum of consecutive steps $\frac{\langle \mathbf{x} \rangle_{\mu,t+1} - \langle \mathbf{x} \rangle_{\mu,t}}{\sigma_t}$. Smoothing is achieved by assigning relative weights to the previous steps according to equation (5), whereas strategy parameter c_σ denotes the backward time horizon of the evolution path. The higher this constant is set ($c_\sigma \leq 1$), the less information from the past is preserved. Constant $\mu_{eff} = 1/\sum_{i=1}^{\mu} w_i^2$ is introduced into the information decay function of $\mathbf{p}_{\sigma,t+1}$ for obtaining normalization by the square root term [*Hansen et al.*, 2003].

[14] The length of the evolution path, $\|\mathbf{p}_{\sigma,t}\|$, is compared to its (expected) length under random selection, $\mathrm{E}\|N(\mathbf{0}, \mathbf{I})\|$ (approximated by $\sqrt{n}(1 - 1/(4n) + 1/(21n^2))$) in order to modify the global step size (equation (4)). If the evolution path is longer than expected, the step size increases. In this case, we can anticipate that mutation steps, which are longer than the actual mutation steps, will speed up the convergence to an optimum by replacing a greater number of small steps. Vice versa, σ_t is decreased when it is shorter than expected, and so it is able to achieve a near optimum. Please note that the argument of the exponential function in equation (4) has to be scaled by c_σ/d_σ, where d_σ is an empirically determined, CMA-ES-specific damping parameter (see Notation).

[15] The covariance $\mathbf{Z}_{t+1}$ of the μ best individuals per generation is exploited to describe the fitness landscape around their weighted mean:

$$\mathbf{Z}_{t+1} = \frac{1}{\sum_{k'=1}^{\mu} \omega_{k'}} \sum_{k'=1}^{\mu} \frac{\omega_{k'}}{\sigma_t}\left(x_{k',t} - \langle \mathbf{x} \rangle_t\right)\left(x_{k',t} - \langle \mathbf{x} \rangle_t\right)^{\mathrm{T}} \quad (6)$$

The new covariance matrix $\mathbf{C}_{t+1}$ is obtained by utilizing both the information of the evolution path (covariance $\mathbf{p}_{c,t+1}\mathbf{p}_{c,t+1}^T$), as well as the covariance matrices of the current and previous generations ("rank μ update"):

$$\mathbf{C}_{t+1} = (1 - c_{\mathrm{cov}})\mathbf{C}_t + c_{\mathrm{cov}} \cdot \left[\frac{1}{\mu_{eff}}\mathbf{p}_{c,t+1}\,\mathbf{p}_{c,t+1}^{\mathrm{T}} + \left(1 - \frac{1}{\mu_{eff}}\right)\mathbf{Z}_{t+1}\right] \quad (7)$$

where c_{cov} and μ_{eff} are used as weighting constants. The learning rate of the covariance matrix is given by c_{cov}, which determines the quota of new information that is added to $\mathbf{C}_t$ during each generation.

[16] Equivalent to the cumulative step size adaptation procedure (4)–(5), the evolution path $\mathbf{p}_{c,t+1}$ (initialization $\mathbf{p}_{c,1} = 0$) is computed by

$$\mathbf{p}_{c,t+1} = (1 - c_c)\mathbf{p}_{c,t} + \sqrt{c_c\,(2 - c_c)\mu_{eff}}\;\frac{\langle \mathbf{x} \rangle_{\mu,t+1} - \langle \mathbf{x} \rangle_{\mu,t}}{\sigma_t} \quad (8)$$

We used the default values given by *Hansen et al.* [2003] for the strategy parameters c_σ, d_σ, c_{cov}, c_{c} and μ_{eff} to obtain a robust solver (see Notation). The only "free" parameters are population size λ and number of parents μ. The former is set to the recommended lower limit of $\lambda = 4 + \lfloor 3 \ln n \rfloor$ and $\mu = \lfloor \lambda/2 \rfloor$. *Hansen et al.* [2003] show that an increase of λ can potentially raise the global search capacity. This is, however, at the expense of a higher number of function evaluations needed to converge to an optimum. To keep the number of the model runs low, λ is not enlarged in this study.

[17] To obtain a uniform scaling for initialization, all decision parameters are scaled at an interval of [0,1] (cf. Figure 1). Constraints on the decision parameter search spaces are implemented according to *Hansen* [2006]. If infeasible search points are sampled, the objective function value is computed as the sum of the median of the previously sampled feasible points with a sufficiently large penalty. The latter is calculated on the basis of the distance from the current infeasible point to the next feasible one. For the initialization of the CMA-ES, $\mathbf{C}_1$ is represented by the identity matrix $\mathbf{I}$. The initial step size, σ_1, has to be large

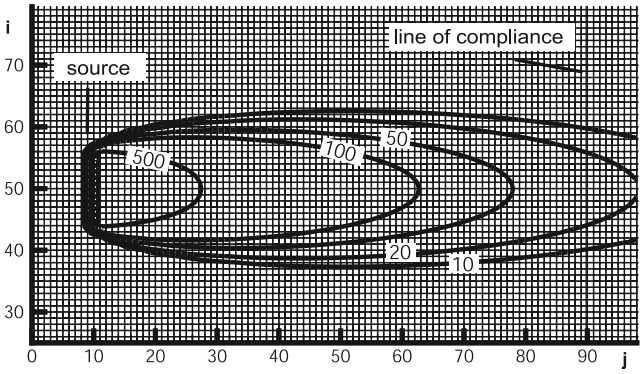

Figure 2. Section of the finite element model grid with concentration contours (μg/L) for a homogeneous aquifer and an intrinsic degradation rate of 25% less than required to ensure compliance with threshold of 10 μg/L at control plane ($k_c = 0.75\ k_0$).

enough to avoid an increase of σ_t during the first generations, which may result in inefficient exploitation of the initial function evaluations for adapting the CMA-ES [*Hansen and Ostermeier*, 2001]. Here it is set at 0.5, which can be regarded as an upper limit, compared to the settings suggested for CMA-ES applications on other multimodal (test) functions [e.g., *Kern et al.*, 2004, *Bayer and Finkel*, 2004]. For further recommendations for the implementation of the CMA-ES and the comprehensive development of the optimization algorithm, the reader is referred to *Hansen* [2006].

3. Flow and Transport Modeling

3.1. Governing Equations

[18] The objective of designing a PTS here is to prevent a contaminant plume from evolving beyond a critical length. For the purpose of this study, steady state, saturated and confined flow is assumed, with a contaminant plume emanating from a continuous source. To approximate the solutions from partial differential equations describing groundwater flow and contaminant mass transport, we set up a basic finite element model using the object oriented program system GeoSys/RockFlow (GS/RF) [*Kolditz and Bauer*, 2004]. Within GS/RF, an operator splitting technique is used, first to calculate flow, and then, assuming equilibrium reactions, to calculate conservative transport. The governing equation for the two-dimensional groundwater flow is [e.g., *Bear*, 1972]

$$\nabla\,(\mathbf{K}\cdot\nabla h)\;-\;\sum_{w} Q'_w\,\delta(i-i_w,j-j_w)=0 \qquad (9)$$

Parameter $h(i, j)$ (L) is the hydraulic head at spatial model coordinates i and j, $\mathbf{K}$ (L/T) is the hydraulic conductivity tensor, Q'_w(L^2/T) represents the well extraction rate per unit aquifer thickness from well w located at spatial coordinates i_w and j_w, and $\delta(i - i_w, j - j_w)$ is the Dirac delta (1/L^2) evaluated at well w. Standard Galerkin finite elements are used to discretize the model area, solve the flow equation and compute the groundwater flow velocity v (L/T) on the basis of Darcy's law. The model is used to simulate the transport of one dissolved, nonsorbing species c originating from the source of contamination. Reactive transport in the saturated model aquifer with constant effective porosity θ is described by

$$\frac{\delta C_c}{\delta t} \;=\; \nabla\,(\mathbf{D}_c\nabla C_c)\;-\nabla(vC_c)-\;k_cC_c \;-\sum_{w\in W}\frac{C_w}{\theta}Q'_w\,\delta(i-i_w,j-j_w) \qquad (10)$$

where $C_c(i, j, t_m)$ (M/L^3) is the solute concentration, k_c (1/T) is the first-order degradation rate constant, C_w (M/L^3) is the concentration in the water abstracted by a well w. Parameter $\mathbf{D}_c$ is the hydrodynamic dispersion tensor (L^2/T) with

$$\mathbf{D}_c \;=\; (\alpha_T\,|v|+D^*)\delta_{ij}\mathbf{I}+\;(\alpha_L-\alpha_T)\frac{v_iv_j}{|v|} \qquad (11)$$

Parameter α_T denotes the transversal, α_L the longitudinal dispersivity (L), D^* the molecular diffusion coefficient (L^2/T), δ_{ij} the Kronecker delta and $|v|$ the magnitude of v. For the application of the optimization algorithm, only steady state plumes, without temporal changes are considered, such that $\delta C_c/\delta t_m = 0$ (equation (10)). This sets the focus on the "worst case," that is, the maximum extension of the plume.

[19] Assuming a first-order kinetic degradation mechanism, which is quantified by a spatially stable degradation constant, is a rough simplification. It has been selected as a straightforward approximation of the typically complex biogeochemical multispecies reactions of contaminants when degraded in the subsurface [*Stenback et al.*, 2004; *Bauer et al.*, 2006]. However, as the main focus is on the application of the CMA-ES, it seems acceptable to streamline the modeling concept in order to show how the considered type of control problems can be solved. By this approximation, the simulation time can be reduced and the different attenuation capacities of an aquifer are only specified by the values for one variable. The degradation rate constant implicitly describes the relevant processes, except sorption, that remove contamination from the aqueous phase and that are expected to be mostly controlled by the in situ microbial degradation activity [*Schäfer*, 2001]. Principally, first-order kinetics are a valid assumption when the degradation rate is mainly a function of the contaminant concentration, availability of required substrate is not a limiting factor and the microbial mass can be approximated as temporally constant within the region of interest [*Clement et al.*, 2000].

3.2. Groundwater Model Setup

[20] The finite element model represents an area of 100 $\times$ 100 m, which is discretized by a uniform base grid of quadratic cells, each with a side length of 1 m and with a constant thickness of 1 m (Figure 2). No flow boundaries in the north and south, together with constant head boundaries along the western and eastern margins, impose a regional hydraulic gradient of 2.0e-03 from west to east in the confined aquifer. Close to the western margin, a north-south oriented linear source, placed along column 10 from row 45 to row 54, continuously releases some contaminant of concern at a constant concentration of 1000 μg/L. Dispersion is specified by $\alpha_T = 0.2$ m, $\alpha_L = 0.5$ m, $D^* = 1.0\text{e-}8$ m/s^2, and the effective porosity is set at $\theta = 0.15$. Around the line

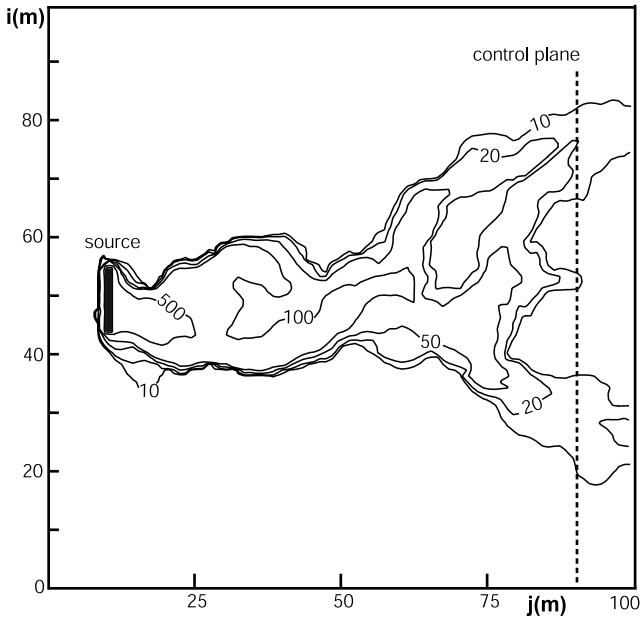

Figure 3. Concentration contours (μg/L) for a heterogeneous aquifer example (realization A, intrinsic degradation rate of $k_c = 0.75\ k_0$).

at the source, the grid is refined to a radius of 1 m around each of the corresponding nodes of the base grid. The grid refinement of the source zone increases the number of nodes by 176. For the simulation of the pumping wells, an automatic grid refinement routine [*Kaiser et al.*, 2000] raises the local discretization level at a radius of 2.1 m in all directions from the well node. Refining the grid around one pumping well yields 118 additional nodes. A more detailed description of this approach is given by *Bayer et al.* [2005]. This procedure is extremely computationally efficient, and is utilized each time the model is called by the optimization routine. Locally, it improves the accuracy of the numerical model, while maintaining the comparatively coarse base grid to hold simulation time low. Though optimization procedures exist which exploit surrogate models of coarse grid resolution as approximations [e.g., *Liu and Minsker*, 2004], the use of autoadaptive grids is, in fact, not common for optimization of groundwater management problems.

[21] Homogeneous, as well as deterministic heterogeneous conductivity distributions are considered. The latter are represented by an ensemble of one hundred different realizations of lognormally distributed **K**, produced randomly by unconditional Gauss simulation using GSTAT [*Pebesma*, 2004]. A lognormally distributed conductivity $\mathbf{Y} = \ln(\mathbf{K})$ with log mean of -4.0 (related to m/s) was considered. In order to explore different degrees of heterogeneity, five subsets, each of 20 realizations, were produced. Production was based on an exponential variogram with the following combinations of variance σ_Y^2 and correlation range I_Y: $\sigma_Y^2 = 1/I_Y = 3$ m, $\sigma_Y^2 = 1/I_Y = 2$ m, $\sigma_Y^2 = 1/I_Y = 15$ m, $\sigma_Y^2 = 2/I_Y = 7$ m, $\sigma_Y^2 = 2/I_Y = 15$ m. Because of the small number of 20 realizations representative of statistically equivalent heterogeneities, the results of optimization will not be compared among the subsets. Instead, an inspection of CMA-ES performance is given for the homogeneous case, one arbitrarily selected heterogeneous realization ($\sigma_Y^2 = 2/I_Y = 7$ m) and for the bulk ensemble as a whole.

[22] A virtual control plane, perpendicular to the imposed regional hydraulic gradient, is located at a distance of 80 m downgradient from the source (at eastern column 90 over all rows). It serves as the line of compliance (loc) where the efficiency of PTS policies is assessed. Along this loc and further downgradient, the contaminant concentration must not exceed 10 μg/L. For each hydraulic conductivity realization the minimum degradation rate constant k_0 was determined as a reference, which ensures that the concentration threshold at the control plane is not surmounted. This specific value was decreased to $k_c = 90\%$, 75% and 50% of k_0, to simulate three analogous cases of natural attenuation intensity for each realization. In doing so, 303 scenarios with different degradation constants and conductivity distributions were generated, which were subject to optimization. Figure 2 depicts the model section with source zone, steady state contaminant plume and control plane assuming a homogeneous aquifer and $k_c = 0.75\ k_0$. Figure 3 depicts the concentration contour lines for an exemplary heterogeneous realization, at degradation rate $k_c = 0.75\ k_0$.

4. Development of Optimization Problem and Results

4.1. Formulation of Objective Functions

[23] One typical economic optimization criterion when adopting PTS for long-term hydraulic control is the minimization of the total pumping rate $Q = \Sigma Q_w$ [e.g., *Wang and Ahlfeld*, 1994; *Aly and Peralta*, 1999; *Bayer and Finkel*, 2004; *Guan and Aral*, 2004]. The reason is that costs for both treatment and operation of pumps are usually strongly correlated to the rate of extracted contaminant water. If the number of wells is subject to optimization or well installation costs differ depending on variable well locations [*Zheng and Wang*, 1999; *Chang and Hsiao*, 2002], these costs may also be included in the objective function. For the purposes of our study, these differences in expenditures for well installation are presumed to be of minor relevance, especially as PTS costs are expected to be dominated by operational costs. A further determinant of costs for PTS is the (flow averaged) concentration $\overline{C}_w$ of the contaminant in the water to be treated on site. A prime example of this is, as described by sorption isotherms, the specific consumption of granular activated carbon typically increases with $\overline{C}_w$ (and Q). Cost functions may become much more complex, especially when PTS is considered as a restoration rather than a static control measure, because of time dependencies of these costs, scheduled planning concepts, and the incorporation of further aspects such as charges for site planning, closure and labor [e.g., *McKinney and Lin*, 1996, *Huang and Mayer*, 1997; *Zhang et al.*, 2004; *Bayer et al.*, 2005].

[24] In this study, we do not present a detailed cost calculation, but rather, focus on the two key factors controlling operational and installation costs of a PTS. These are (1) the pumping rate and (2) the mass extraction rate $Q\overline{C}_w$. While the importance of the first factor has already been dealt with, the role of the second should be explained further. As mentioned above, typical cost functions involve a product of Q with some nonlinear function of $\overline{C}_w$, which is scaled by specific cost coefficients. The nonlinear function

of $\overline{C}_w$ may be approximated by $\overline{C}_w^{\varphi}$ ($\varphi < 1$) [e.g., *Chan Hilton and Culver*, 2005]. Assuming that either the exponent φ is close to 1, or that a linear approximation is sufficient for the range inspected, then the objective is to minimize $Q\overline{C}_w$, that is, the mass extraction rate of contaminant. This is somehow contradictory to the objectives set when PTS is used for site cleanup, especially when minimization of restoration time is considered. In this case, a maximal mass extraction rate could be desirable. Owing to the fact that short time frames of remediation by PTS are rather unusual, it is preferable to use PTS solely for hydraulic control in many situations. Consequently, the concentrations of contaminants in the extracted water can govern the operational cost. In the framework of our study, the objective of minimizing the mass extraction rate can be interpreted as exploiting the natural attenuation capacity of the aquifer to a maximum extent.

[25] Depending on whether total pumping rate or mass extraction should be minimized, the mathematical formulation of the objective function is

$$\min_{Q_w,i_w,j_w} Q = \sum_{w=1}^{W} Q_W(i_w,j_w) \tag{12a}$$

or

$$\min_{Q_w,i_w,j_w} Q\overline{C}_w^* = \min\left[\sum_{w=1}^{W} Q_W(i_w,j_w)\cdot \max\left(1,\frac{C_w(i_w,j_w)}{C_{MCL}}\right)\right] \tag{12b}$$

such that

$$Q_{\min} \le Q_w \le Q_{\max} \tag{13a}$$

$$i_{\min} \le i_w \le i_{\max} \tag{13b}$$

$$j_{\min} \le j_w \le j_{\max} \tag{13c}$$

$$C_c(i,j_{cp}) \le C_{MCL} \tag{14a}$$

$$\frac{\Delta h(i,j)}{h_{ref}(i,j)} \le h' \tag{14b}$$

where well-specific pumping rates are Q_w, with row coordinates i_w and column coordinates j_w. Wells may be positioned within a rectangular well placement area, predefined by the intervals $[i_{\min}, i_{\max}] = [15\text{ m}, 85\text{ m}]$ and $[j_{\min}, j_{\max}] = [15\text{ m}, 85\text{ m}]$. By keeping a 15 m distance, the influence of the fixed model boundaries will be kept low. Also, only pumping outside of the source zone ($i = 10$ m) in the upgradient area of the control plane ($i = 90$ m) shall be examined. Please note that objective function (12b) simplifies to the minimization of the pumping rate when bulk contaminant concentrations in the extracted groundwater lie below standards C_{MCL} and no treatment is necessary.

[26] Equations (14a) and (14b) represent constraints on maximum allowable contaminant concentrations at the control plane (i.e., represented by column $j = j_{\mathrm{cp}} = 90$ m), and on maximum drawdown at grid nodes i and j. Both are implemented as penalties in the objective functions, and we obtain modified objective functions from equation (12)

$$\min_{Q_w,i_w,j_w} J_1 = Q\cdot \Im_C \cdot \Im_D \tag{15a}$$

$$\min_{Q_w,i_w,j_w} J_2 = Q\overline{C}_w^*\cdot \Im_C \cdot \Im_D \tag{15b}$$

[27] Function $\Im_C$ penalizes all invalid pumping well schemes that cannot ensure C_{MCL} at the control plane. It is implemented as

$$\Im_C = 10^{\left(\rho_1 C_c'\right)^{\rho_2}} \tag{16a}$$

given that

$$C_c' = \sum_i \max\left[0, C_c(i,j_{cp}) - C_{MCL}\right], \tag{16b}$$

where coefficient ρ_1 ($\mathrm{m^3/mg}$, $\mathrm{l/\mu g}$) is used to scale the sum of target concentration violations at the control plane C_{c}'. In our case, it is set at 10 $\mathrm{l/\mu g}$, in order to achieve severe penalization of any invalid well scheme. Exponent ρ_2 (<1) is used to prevent an increase of $\Im_C$ beyond computationally tractable values (here $\rho_2 = 0.6$). Parameter i indexes all row cells (1, . . ., 100) at the control plane. Please note that this formulation of $\Im_C$ is one of many options for calculating and contrasting the degree of "invalidity" by a penalty term. As long as small concentration violations are already severely penalized, the exact formulation of $\Im_C$ will not exert a major influence on the course of the optimization. This is because of the fact that CMA-ES relies on rank-based rather than fitness-proportional selection.

[28] If constraints on drawdown are to be incorporated, a formulation with a multiplicative penalty term is suggested:

$$\Im_D = 1 + 100\cdot \max\left(0, \frac{\Delta h(i,j)}{h_{ref}(i,j)} - h'\right) \tag{17a}$$

otherwise

$$\Im_D = 1 \tag{17b}$$

Parameter h_{ref} denotes any specific reference head distribution at nodes i and j. Most straightforward would be to refer to the head distribution under undisturbed flow conditions. Then Δh is the absolute change in head if wells are operating, and h' is the maximum allowable relative change. As subsequently shown, instead of absolute head values, h_{ref} can also represent a drawdown matrix calculated for another well scheme. In such a case, the purpose is to ensure that the drawdown caused by the well scheme under

consideration does not reach a certain percentage h' of the drawdown computed for a reference case. In equation (17a), factor 100 effects a severe penalization, so that 1% of relative head change beyond threshold h' already doubles the value of the objective function without the penalty. Please note that this constraint was only considered for the five-well case.

4.2. Application and Assessment of Optimization

[29] Because of the stochastic nature of heuristic optimization techniques such as CMA-ES, it cannot be guaranteed that these algorithms will find the optimum in a single optimization attempt (i.e., optimization run OR). As a consequence, several ORs are usually mandatory and the CMA-ES is repeatedly applied to increase the probability of success. As for evolutionary algorithms in general, the total number of function evaluations, which are equivalent to the number of model runs, are determined by the number of ORs and the function evaluations I per OR. As long as no other termination criterion is considered (e.g., measure of stagnation or threshold of objective function value), I has to be a priori defined as a multiple of the individuals λ per generation. For practical applications, the essential questions are how many optimization runs n_{OR} and how many function evaluations I to choose. In principle, the probability of finding the global optimum by using a stochastic search algorithm will never reach 100% and therefore performing as many ORs as computationally possible is worthwhile.

[30] In order to identify an ideal value I_{deal} of I for each problem considered in our study, we follow an approach presented in an earlier publication [*Bayer and Finkel*, 2004]. The suggested procedure is as follows.

[31] 1. Define a computationally feasible number of repeated optimization attempts n_{OR} (e.g., 25) with a problem-specific conservative estimation of I per attempt.

[32] 2. Randomly initialize first population $\mathbf{x}_{k,1}$ for each OR (to exclude bias).

[33] 3. Apply CMA-ES in n_{OR} attempts (with respect to one objective function).

[34] 4. Determine *value J** of the objective function that has to be achieved by formal optimization, for example, the global optimum that is identified before by an appropriate method (e.g., full enumeration, direct search), or any other case-specific threshold of the objective function value.

[35] 5. Define a relative tolerance range (e.g., 10%) centered around J^*, which should reflect the required accuracy (well schemes that ensure objective function values within the range are called optimal solutions).

[36] 6. Compare the best individuals found after each consecutive generation $t = 1, \ldots, t_{max}$ between all ORs, and, after each generation, counting of the number of ORs which yielded solutions within the tolerance range.

[37] 7. Calculate the generation-specific success rate Φ_t over all attempts, which is the cumulative frequency of successful ORs versus the total number of attempts n_{OR}.

[38] 8. Divide the number of function evaluations ($= \lambda t$) required from generation 1 to t by success rate after generation t, which gives a generation-specific number of total model runs $MR_t = \lambda t/\Phi_t$ required on average to find one optimal well scheme. The number of generations where MR_t is minimal denotes the theoretically ideal number of generations t_{ideal}, and $MR_{t,ideal}/t_{ideal}$ the ideal number of n_{OR}. Since $MR_{t,ideal}/t_{ideal}$ is generally not an integer, the practical ideal values of MR_t and t are computed by

$$MR_{t,ideal} = \lambda \min_t \left\lceil \frac{MR_t}{\lambda} \right\rceil, \quad t = 1, .., t_{max} \tag{18}$$

where $\lceil\ \rceil$ denotes the ceil function; it returns the smallest integer that is not less than its argument. The approach yields an estimation of the ideal number of model runs $I_{deal} = \lambda t_{ideal}$ per OR. For the subsequent presentation of results, both the success rate, Φ, and the total number of model runs, MR, are shown as a function of generations, rather than in relation to the corresponding number of model runs (i.e., function evaluations) per OR. This allows for a more transparent presentation.

[39] The presented statistical investigation of stochastic solvers such as CMA-ES is crucial to understanding their reliability, the reproducibility of results and to give generally valid performance estimates. By utilizing this procedure, we adhere to the recommendations of *Mayer et al.* [2002] regarding the application of optimization algorithms for design problems involving flow and transport simulations. They advocate multiple optimization runs when using nondeterministic solvers to derive a measure of error in the objective function value returned, and to estimate the reliability of the algorithm. Only through repeated attempts and various application problems (e.g., with variable hydrogeology) is it possible to derive suggestions for an ideal application of the optimization method to other problems and, in this way, ensure transferability. After the subsequent discussion of results, we will therefore subsume recommendations for CMA-ES application.

[40] Please note that the aforementioned procedure does not guarantee the identification of an optimum, but calculates the minimum number of model runs required on average. It is therefore best suited as a measure to compare optimization algorithms' performance. A global optimum can only be ascertained by (1) full enumeration of solution alternatives or (2) by comparison of local optima that are obtained for convex subspaces of the entire decision space (i.e., for subproblems with global and without local optima). The latter was chosen for the subsequent analysis. A direct search method [*Nelder and Mead*, 1965] was used to examine the characteristics of objective functions (fitness landscapes) in detail. This Nelder-Mead method is based on evaluating a function at the vertices of a simplex, then iteratively shrinking the simplex as better points are found until some desired bound is obtained. It is thus suited to rapidly identifying the solutions of subproblems, such as minimum pumping rates of fixed wells.

5. Results: One-Well Case

5.1. Identification of Optima

[41] First, the results for the homogeneous aquifer scenario will be discussed. When the objective is to position one pumping well to minimize pumping (objective function J_1) or mass extraction rate (J_2), it can be expected that, because of the symmetric form of the plume, the best position will be somewhere at the centerline at row $i = 49$ m. In an anticipatory examination, the Nelder-Mead method was successively used to identify the exact optimum for fixed

Table 1. Optimal Well Settings for One-Well Case

	Degradation Rate k_c	Q, m^3/d	C_w, mg/m^3	Row i, m	Column j, m
		Homogeneous Aquifer			
min(J_1)	0.5 k_0	0.349	634.8	49	22
	0.75 k_0	0.217	586.5	49	24
	0.9 k_0	0.114	530.0	49	25
min(J_2)	0.5 k_0	0.541	69.1	49	85
	0.75 k_0	0.350	34.3	49	85
	0.9 k_0	0.203	21.7	49	85
		Heterogeneous Example (Realization A)			
min(J_1)	0.5 k_0	0.159	250.9	38	26
	0.75 k_0	0.061	468.9	42	25
	0.9 k_0	0.025	453.1	42	25
min(J_2)	0.5 k_0	0.629	39.3	58	81
	0.75 k_0	0.224	47.5	57	80
	0.9 k_0	0.106	30.8	63	55

wells at steps of 1 m between columns j = 15–85 m, that is, in the well placement area. For this direct search method, we used a maximum of 20 iterations, initializing each consecutive optimization by the optimum pumping rate found for the previous well position. Interestingly, pumping rates are minimized for both objectives (i.e., also with respect to J_2). This means that a decrease in mass extraction rate by increasing the pumping rate through dilution (i.e., a disproportional decline of the concentration in the extracted groundwater) was not possible. Optimal well positions differ depending on objective function type and change with the chosen degradation rate. The best well positions were found at j = 22 m (k_c = 0.9 k_0), j = 24 m (k_c = 0.75 k_0) and j = 25 m (k_c = 0.5 k_0) when only the pumping rate was minimized. The lower the natural attenuation capacity, the wider the plume expands, and thus the further the well is positioned downgradient. The differences, however, are small. Compared with the optimized well positions, the required pumping rates change significantly. They increase from Q = 0.11 m^3/d (k_c = 0.9 k_0) to Q = 0.35 m^3/d (k_c = 0.5 k_0) (Table 1).

[42] In contrast to the optimal well positions when minimizing pumping rates, mass extraction minimization always led to the positioning of wells at maximum distance from the source (i.e., at the eastern boundary of the well placement area (j = 85 m), independent of assumed degradation rate). Obviously, the reason for this trend is that distant wells maximize the effect of natural attenuation, which is reflected by the optimal solutions listed in Table 1. That is, compared with the results for objective function J_1, pumping rates nearly doubled, while concentrations in the extracted groundwater decreased to about 10% or less depending on the value of the degradation rate.

[43] In addition, one heterogeneous aquifer scenario, arbitrarily selected from the ensemble of 100 stochastically generated heterogeneous transmissivity distributions, was analyzed in detail. Figure 3 depicts the plume with undisturbed flow conditions assuming a moderate degradation rate of k_c = 0.75 k_0. Similar to the treatment of the homogenous aquifer, the Nelder-Mead method was applied to investigate the spatial changes of objective function values relating to minimization of J_1 and J_2 for fixed well positions. All well positions in the 1 m regular base grid within the well placement area were examined (71 × 71 = 5041 positions). The interpolated contour maps are shown in Figures 4a and 4b. Figure 4 illustrates how the different objectives (Figure 4a: minimization of pumping rate, Figure 4b: minimization of mass extraction rate) lead to totally different schemes. Figure 4 depicts irregular variations for the values of min(J_1) and min(J_2) over one to more than two orders of magnitude, reflecting the significant influence of variations in the hydraulic conductivity distribution. The existence of multiple local optima indicates the difficulty of the problem at hand.

[44] Similar to the global optima found for the homogeneous aquifer, the best well position when minimizing Q is in vicinity of the source, and the best position when

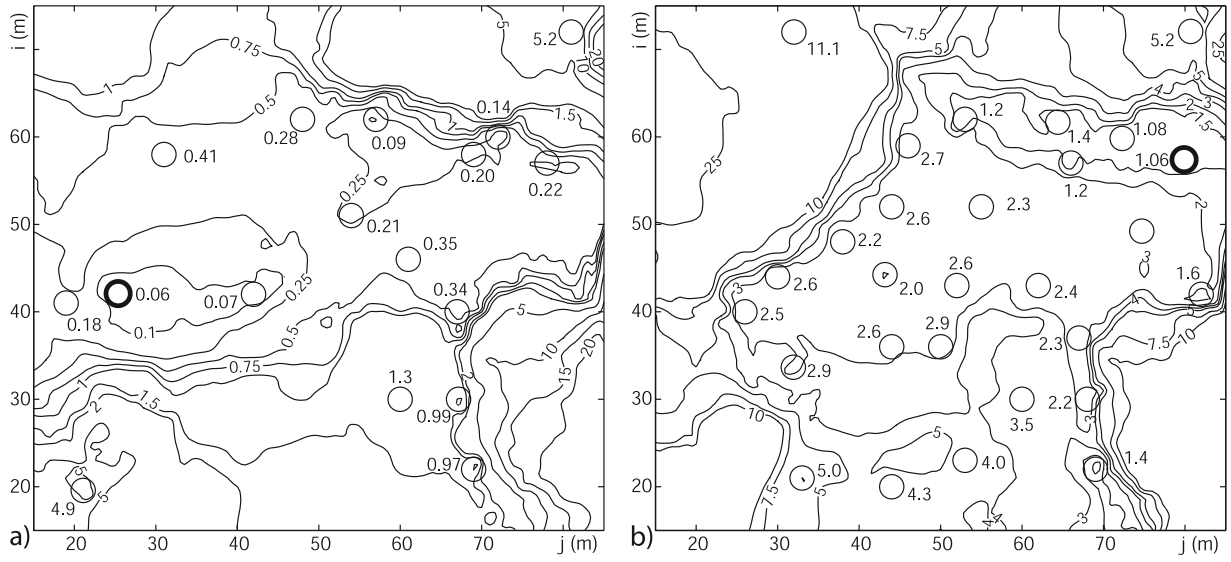

Figure 4. Contour plots, of equal objective function values within the well placement area, for the heterogeneous aquifer example (realization A, k_c = 0.75 k_0) provided for minimizing (a) pumping rates (m^3/d) and (b) normalized mass extraction rates (mg/d)/C_{MCL}. Thin circles denote local optima, and thick circles represent a global optimum.

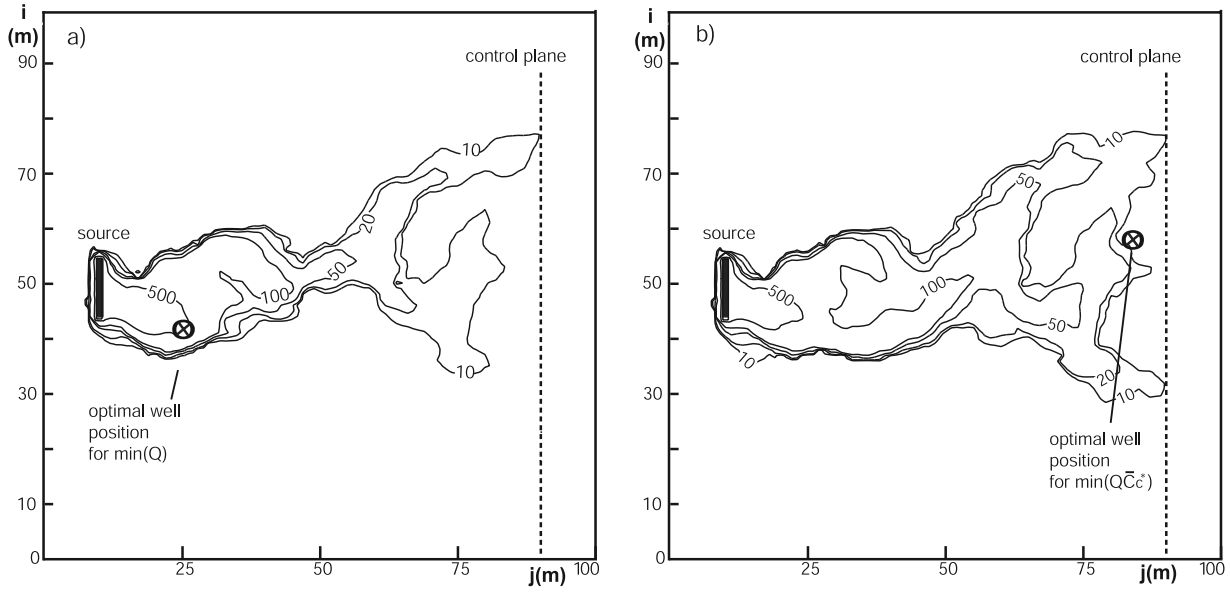

Figure 5. Concentration contours (μg/L) if optimized one-well solution is operating for the heterogeneous aquifer example (realization A, $k_c = 0.75\ k_0$): (a) contours for solution minimizing pumping rate and (b) contours for solution minimizing mass extraction rate (compare Figure 2 for undisturbed flow).

minimizing $Q\overline{C}^*_w$ is close to the eastern boundary of the well placement area. Concentration contour maps for the optimized well configurations are shown in Figures 5a and 5b ($k_c = 0.75\ k_0$). An interesting feature is that the well positions do not turn out to be in the center of the plume, nor in the zone of least plume width ($j = 49$ m). In contrast to optimized systems, comparatively higher groundwater flow rates in these zones may lead to the need for elevated pumping rates, which are required to achieve the compliance criterion at the control plane.

5.2. CMA-ES Application to Homogeneous Aquifer Scenario

[45] In the first step, we tested the (3_W,7)-CMA-ES on the homogeneous aquifer scenario (cf. Figure 2), for the entire well placement area (within *i*/*j* coordinates 15–85 m). In three different scenarios, first-order degradation rate constants k_c were set 50%, 75% and 90% of k_0. For each rate constant, we conducted a total of 25 optimization attempts with approximately 400 (exactly: 406, 58 generations for seven individuals each) model runs each. Assuming limited knowledge of the problem before the optimization procedure, an interval of three orders of magnitude was set for the range of pumping rates to be examined. Oriented on the findings of the primordial direct search, the upper limit Q_{max} was defined as two orders of magnitude higher than the minimum found, and the lower limit Q_{min} was set to one order less than this minimum. By doing so, it was assured that optimal pumping rates were within the decision space. Furthermore, in order to attain a comparable initialization, this approach was also repeated in the subsequent application of the CMA-ES for the heterogeneous aquifer example.

[46] The individual characteristics (i.e., rates of convergence) of each separate CMA-ES run yield nonuniform shapes of cumulative frequency histograms and *MR* curves. Hence overall trends can be identified, but even after 25 ORs, it is not possible to recognize clear differences in the CMA-ES performance on the different problem formulations. In particular, varying k_c did not show any notable effect, and therefore, as an example, only the results for $k_c = 0.75\ k_0$ are depicted in Figure 6. The chosen value of k_c determines the objective function values of each well configuration, but apparently does not significantly control the complexity of the fitness landscape. The success rate for finding the optimal solutions, as determined by the Nelder-

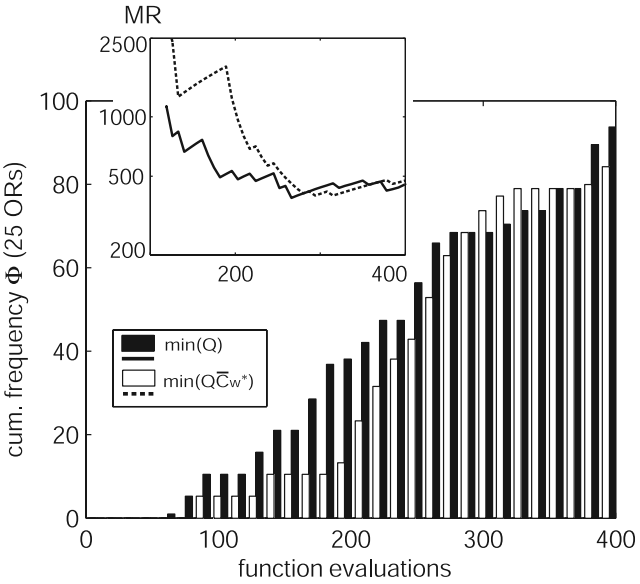

Figure 6. (3_W,7)-CMA-ES application: development of cumulative frequency (i.e., success rate Φ) over 25 optimization attempts to find the optimum (tolerance of 10%) when one well is optimized for a homogeneous aquifer, assuming $k_c = 0.75\ k_0$. The inset depicts calculated average number *MR* of model runs to find an optimum in dependence of function evaluations per attempt.

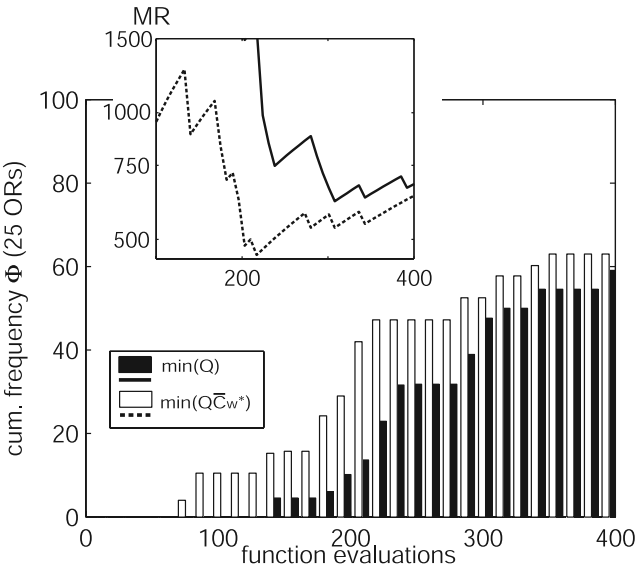

Figure 7. (3_W,7)-CMA-ES application: development of success rate Φ to find the optimum (tolerance of 10%) when one well is optimized for the heterogeneous aquifer example (realization A, k_c = 0.75 k_0). The inset depicts calculated average number *MR* of model runs to find an optimum in dependence of function evaluations per attempt.

Mead method at a tolerance of 10%, reaches more than 75% after all function evaluations for each specification of k_c are complete. However, the success rates Φ for optimization of Q (J_1) differ to the success rates of minimizing mass extraction $Q\overline{C}_w^*$ (J_2). If the pumping rate alone is the objective, then nearly all optimization attempts located the optimum (90% ± 5% depending on k_c). Focusing on mass extraction yields about 10% lower success rates. After ca. 100 to 150 function evaluations per attempt, there is a continuous increase of the overall success rate Φ, with a delayed increase when mass extraction is to be minimized. The reason for these differences can be seen in the particular formulations of the objective functions. Both require a model run to check the contaminant concentrations at the control plane. For minimization of mass extraction, J_2 additionally computes concentrations in the extracted groundwater that depend on Q and well location. Incorporating this additional parameter potentially complicates the objective function. For example, consider a well that is located outside the center of a plume. If there is no failure, raising the pumping rate means a linear rise of J_1, but a nonlinear increase of J_2 depending on the change of concentrations in the extracted groundwater. As the capture zone widens with Q, reaching highly contaminated aquifer zones (for example the center of the plume) could cause a disproportional increase of extracted mass. Depending on the spatial concentration gradients, the further extending capture zone then covers less contaminated groundwater and could even decrease the mass extraction rate.

[47] The values of the average number of function evaluations required to find the optimum, *MR*, are within similar ranges for both objectives. Distinct influences of the results of individual ORs yield oscillations in the tradeoffs between *MR* and function evaluations per attempt. Consequently, the number of ORs is not large enough to determine I_{deal} exactly and, instead, it seems reasonable to deduce recommendable ranges. The ideal number of model runs here are within I_{deal} = 250 to 400 (or slightly higher).

[48] The results for the homogeneous aquifer indicate a high reliability of the CMA-ES and a required (approximately) *MR* = 400 model runs to find the optimum. Well placement assuming a homogeneous aquifer is chosen as the most simple problem, as no nonconvexities of the fitness function are affected by spatially varying hydraulic conductivities. We have not inspected the 3-D fitness landscapes further, which could potentially have local optima. However, as long as the conductivity distribution is homogeneous, problems involving well configuration are principally not too demanding for classical or heuristic search techniques. Figure 6 also shows that, no matter how simple the problem is, the heuristic algorithm cannot guarantee a success rate of 100%. We chose this case as a reference for optimization of heterogeneous aquifer scenarios. However, it is obvious that the problem could be solved by focusing on the centerline of the plume, probably with a lower number of model runs required when utilizing a direct search algorithm.

5.3. CMA-ES Application to Heterogeneous Aquifer Scenarios

[49] The (3_W,7)-CMA-ES was first applied to the selected single heterogeneous aquifer scenario (subsequently denoted as realization A, Figure 3) for k_c = 0.75 k_0. As in the example of a homogeneous aquifer, 25 ORs of 406 model runs each were conducted and the tolerance was set to 10% of the optimum previously detected with the Nelder-Mead method. Figure 7 depicts the rise of success rates versus the number of function evaluations per OR. The calculated minimum total number of function evaluations *MR* is approximately 500 for minimizing mass extraction and around 650 for pumping rate minimization. The ideal range of model runs per OR is between 200 and 400 (or higher).

[50] The spatial conductivity variations aggravate the complexity of the objective function and thus reduce the success rate in identifying an optimal solution. In contrast to the findings for the homogeneous aquifer scenario, the average mass extraction minima for this particular heterogeneous case were found earlier and gave success rates superior to those obtained by minimizing Q. A detailed inspection of the generation-wise, evolutionary search during each individual CMA-ES run revealed that the major reason for this discrepancy lies in the individual characteristics of the objective functions. When minimizing for Q, several ORs got (temporally) stuck in the region around well location 62/57, which is a local optimum in the contour map depicted in Figure 4a. The reason why this local optimum is more attractive than others could not be clarified with certainty. Apparently this region competes severely with the global optimum as it allows very close optimal solutions. Similarly, when minimizing mass extraction, the well position 62/53 acts as a local optimum ''trap,'' although the deteriorative effect on the evolutionary search is less pronounced.

[51] In the next step, optimization was performed for the bulk ensemble of 100 realizations. For each heterogeneity distribution and the respective three rate constants, five optimization attempts were executed for each objective. The boundaries of Q_w were set over a range of four orders

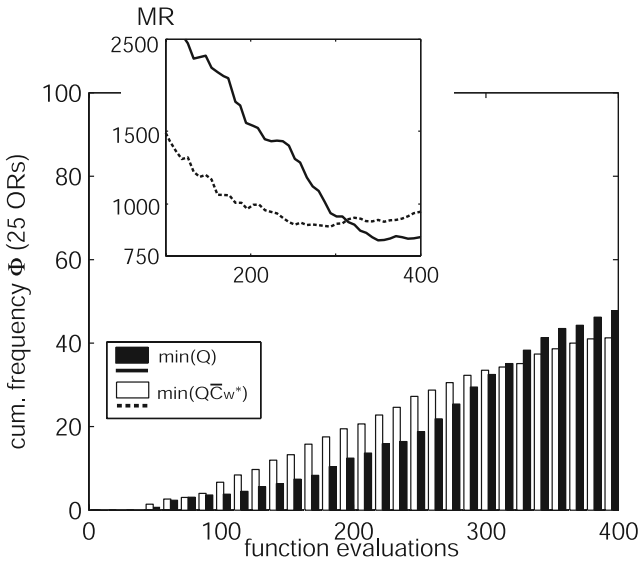

Figure 8. (3_W,7)-CMA-ES application: development of success rate Φ to find the optimum (tolerance of 10%) when one well is optimized for 100 heterogeneous aquifer realizations, assuming $k_c = 0.75\ k_0$ (five attempts for each problem). The inset depicts calculated average number *MR* of model runs to find an optimum in dependence of function evaluations per attempt.

of magnitude, taking the best first guess—the minimum pumping rates for the homogeneous case—as log mean. In view of the expected minor effect of the degradation rate constant on the CMA-ES performance, we will highlight the results for $k_c = 0.75k_0$. Again, the direct search method was employed to identify the case-specific global optimum in advance.

[52] The results show a general trend that was already observed for the homogeneous case. The success rates of detecting an optimum are higher when only the pumping rate Q, and not mass extraction, is the objective. In detail, for 77 realizations, five ORs detected a Q-optimized well configuration with equivalent accuracy as determined by the Nelder-Mead method (differences in objective function value of less than 1%). If we raise the tolerance to 10% around the optimal value, this number rises to 96. On one hand, this means a high reliability of the evolutionary algorithm. On the other hand, there were four cases in which even five ORs were not enough to detect the global optimum. Although the small number of ORs is not sufficient to obtain statistically representative performance evaluations for each particular case, the results indicate case-specific success rates, reflecting the individual nature of each conductivity pattern and thus a realization-specific complexity of the objective function. The same is observed when considering mass extraction as an objective: at a tolerance of 1% an acceptable solution was discovered in 67 cases, at 10% in 88 cases.

[53] The overall development of success rates Φ at increasing function evaluations per attempt (after totally $n_{OR} = 500$) is shown in Figure 8. Averaging of Φ over 100 realizations yields smoother plots for *MR* compared with the curves for the single heterogeneous realization A (cf. Figure 7). Also, the success rates Φ for the bulk ensemble are lower, which might by affected by the broader ranges set for decision variable Q. Specifically, after 400 evaluations, they reach Φ = 48% for selecting Q as optimization criterion, and 42% when minimizing $\overline{QC}^*_w$. Again, setting mass extraction minimization as a criterion means an earlier increase of Φ. For this objective, the recommendable number of model runs per attempt is within $I_{deal} = 250-400$, and in the range of $I_{deal} = 300-400$ for objective J_1. The tradeoffs in Figure 8 indicate that no clear benefits can be achieved beyond 400 model runs. For both objectives, we calculate similar total numbers of minimum function evaluations required to find an optimum, which are *MR* = 820 for J_1, and 880 for J_2.

[54] The different performance for both objectives may be attributed to the same factor as suggested for the homogeneous case: introducing the contaminant concentration of the extracted groundwater into the objective function mostly means additional complexity. However, the results of individual realizations may deviate from overall findings (cp. results for realization A). As there is no straightforward method at hand which is suited to generally quantifying complexity, we inspected realization-specific objective function contour maps (similar to Figure 4) on the basis of the results from the direct search. A general observation is that these contour maps tend to exhibit more local optima when mass extraction is minimized. Also, we often found extended nonconvex regions of a higher number of tolerable close optimal well positions, rather than one prominent global optima, when Q is minimized. This is reflected in the increase of success rates, which occurs earlier as the number of function evaluations per OR rise (Figure 7).

5.4. Examination of Optimal Solutions

[55] The optimization of the hydraulic conductivity realizations was not only utilized to derive recommendations for configuring the (3_W,7)-CMA-ES, but also to identify characteristic features of the solutions obtained. Subsequently, we will examine the optimized well schemes and, in particular, the positions of wells.

[56] The previous application showed that five ORs have substantial reliability in finding a close optimal solution to the one-well problem. Consequently, to further investigate the influence of variations of the degradation rate constant k_c, we avoided a computationally demanding direct search of optima. Instead, only an evolutionary search was performed for $k_c = 0.5\ k_0$ and $k_c = 0.9\ k_0$, equivalent to the application for $k_c = 0.75\ k_0$. Although the solutions of some of the realizations are probably not the global optima, this procedure should demonstrate the general effects of changing the value of k_c.

[57] The optimal well positions found for minimizing Q are shown in Figure 9a. There turns out to be a preference area in the downgradient vicinity of the source. For all degradation rates assumed, a small rectangular area could be delineated, which encompasses more than 75% of the optimal positions ($i = 41-58$ m, $j = 15-30$ m). More than 75% of the optimal well positions are located less than 18 m away from the center of the source, which is in the upgradient quarter of the well placement area (Figure 10). As a result, if natural attenuation occurs independent from aquifer heterogeneity, extracting a fraction of the highly contaminated groundwater released by the source, rather than using distant pumping wells, is favorable. Further

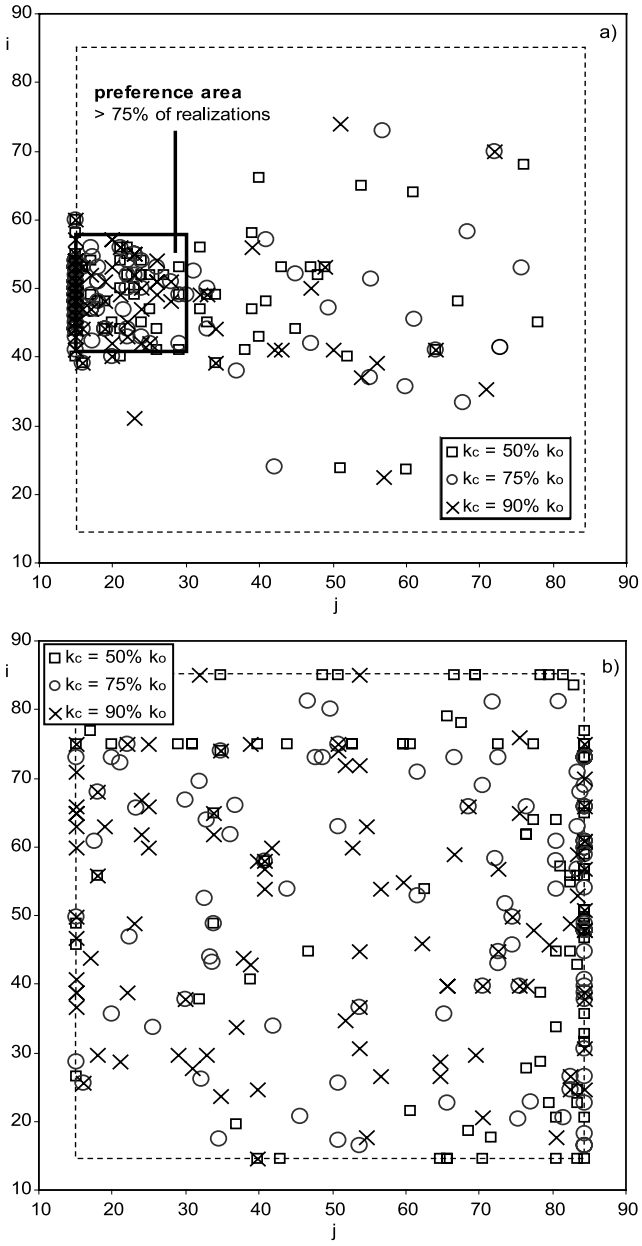

Figure 9. Well positions within well placement area for 100 heterogeneous aquifer realizations: (a) minimized pumping rates and (b) minimized mass extraction rates.

downgradient, more extensive depression cones are obviously needed to extract an ample mass of the contaminant from the underground. This is to ensure the compliance objectives at the control plane. As shown for the exemplary realization A, this conclusion is even valid for situations in which the plume width considerably varies and is locally very small (see Figures 4 and 5). In the case of substrate limited degradation, such as described by Monod instead of first-order kinetics, these findings can be expected to be similar, or even more pronounced. Extracting water upgradient in the plume decreases the downgradient contaminant concentrations, and thus the demand on substrate, weakening its role as a limiting factor for the intrinsic degradation capacity of the aquifer. The degree of degradation has a minor influence on these findings (at least for the ranges of k_c considered here). However, in most cases, different values of k_c lead to different optimal well positions. An overall trend is that higher degradation rates move the best well positions closer to the source (Figure 10).

[58] The set of realizations which produced well positions outside of the preference area was further examined to answer the question: how much higher is the pumping rate for wells which are limited to this area, compared to the realization-specific, optimal solution for the entire well placement area? In line with the previous procedure, the direct search method was used for the 20 realizations of $k_c = 0.75\ k_0$, and CMA-ES for the realizations of the other k_c scenarios. The results from the direct search revealed that the solutions are distinctively worse for only 7 out of these 20 realizations (i.e., required pumping rates more than double the rate for wells outside this area). For $k_c = 0.9\ k_0$, only 7 (of 16) and for $k_c = 0.5\ k_0$, 5 (of 24) realizations were identified. A further examination showed that between 48 ($k_c = 0.5\ k_0$) and 53 ($k_c = 0.9\ k_0$) of the best well positions have a minimal distance to the source (j = 15 m). These findings are of crucial importance in practice, as any efforts to reduce the commonly existing uncertainty in the description of hydraulic aquifer properties can be limited to a small subarea of the plume downgradient of the contaminant source.

[59] Figure 9b depicts the well positions for minimizing mass extraction. As indicated for the homogeneous and heterogeneous examples above, there is a trend toward well positions which are distant from the source, in areas where the contaminant concentrations are low. This is more evident in scenarios with smaller degradation rate constants k_c, since a certain reduction in contaminant concentrations occurs at greater distances from the source. In contrast to the focused well positions when minimizing Q (Figure 9a), ideal locations of wells appear to be significantly dependent on the form of plume. An interesting feature is that several optimal wells are not located in the zone of the original plume (defined by the concentration contour equal to the

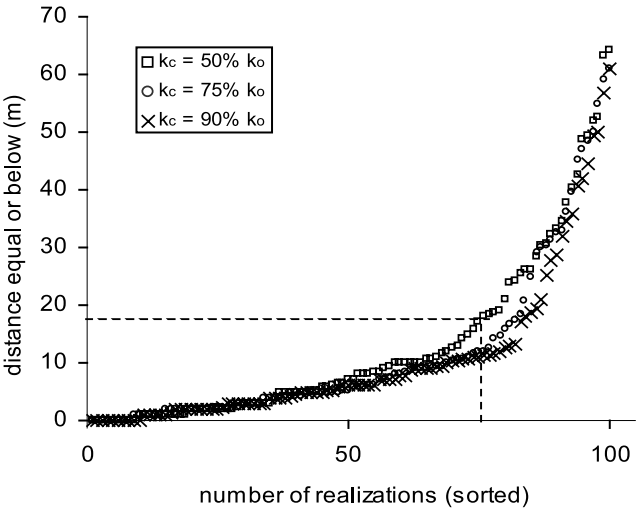

Figure 10. Distance of Q-optimized well positions for 100 aquifer realizations at different intrinsic contaminant degradation rate constants, k_c. As indicated by the dashed line, at least 75% of all positions are at a distance of less than 18 m.

desired standard C_{MCL} = 10 at the control plane). Specifically, the concentrations $\overline{C}_w$ in the extracted groundwater of the optimized systems in 48 (k_c = 0.9 k_0), 34 (k_c = 0.75 k_0) and 10 (k_c = 0.5 k_0) realizations do not reach more than 10 μg/L. This reflects the fact that minimizing mass extraction can lead to solutions wherein the contaminant plume hydraulically deviates from the natural path to increase the traveltime. This way, the fraction of contaminant mass degraded in situ rises. These solutions fulfill the desired objectives to restrict the downgradient propagation of the plume, but the longitudinal extension is hindered at the expense of transversal size. This would be a reasonable approach, if neighboring properties were threatened by industrial sites, and pumping wells could lead to a local extension of the plume. The acceptability of such strategies is largely dependent on local regulations, but in principle any hydraulic measure means a change of the flow situation and, consequently, a change in plume propagation. For example, head build up in front of passive remediation systems generally causes a lateral extension of contaminant plumes to be captured. *Haggerty and Gorelick* [1994] present a critical discussion of such "smearing" effects, particularly regarding their influence on operation time. However, in their study, degradation is not considered. In contrast, *Schäfer* [2001] models the effect of operating wells in the presence of natural attenuation. The study identifies changes in the spatial degradation rates of Xylene when pumping wells are installed, especially as the contact between dissolved oxidants and contaminants is improved. A detailed investigation of changes in relevant biochemical processes, however, is beyond the scope of this paper.

6. Results: Five-Well Case

[60] In this section, we present the results from the installation of five wells for the selected heterogeneous aquifer realization, which was already considered for the one-well case (realization A). Additionally, four further realizations (B, C, D and E) are arbitrarily selected as representatives of the other subsets of geostatistically equivalent realizations. Examination of more realizations is restricted by the associated computational burden for this 15-dimensional optimization problem. In accordance with the previous discussion of one-well results, the primary focus is placed on the (6_W,12)-CMA-ES performance. Hence we will first highlight the algorithm's progression when searching for increasingly better PTS designs, and then examine typical features of the optimized systems. Since previous investigations with different degradation rates yielded similar results for the performance of the optimization routine, only the scenario with k_c = 0.75 k_0 will be considered here.

[61] The optimization procedure was applied in 10 different, randomly initialized ORs for both objectives and all realizations. This means that, in total, 100 ORs, of 3504 objective function evaluations each (292 generations from 12 individuals), were performed. Maximum pumping rates Q_{max} were set five times higher, and the lower boundary Q_{min} was set one order of magnitude lower than the optimum from the one-well case. In order to exclude the previously detected one-well solution from the decision space, an additional constraint on acceptable drawdown was incorporated (penalty term given in equation (17)). Its purpose was to ensure that the maximum drawdown of the five-well case would not reach more than 90% of the drawdown from the previous one-well solution. Accordingly, the reference value $h_{ref}(i, j)$ in equation (17a) was oriented at the head distribution of the optimal one-well solution. However, instead of referring to an absolute head value, the drawdown, which was calculated for the one-well solution, was used. Specifically, $h_{ref}(i, j)$ was assigned the absolute head change for the one-well case compared to undisturbed flow conditions. Parameter Δh of equation (17a) represents the head change for the five-well case, with the threshold h' set accordingly at 0.9.

[62] A major hindrance, when judging the optimization algorithm's performance for the five-well case, is that optimal solutions are not known. This is unlike the previous application, where the Nelder-Mead method could be used to identify the problem-specific best single well solutions. For this much more demanding problem, an analogous approach is virtually impossible within a reasonable time as each additional decision parameter exponentially raises the number of model runs required for an optimization. As an alternative to referring to the global optimum, we suggest comparing the fitness of an optimized system to that identified for the previously discussed one-well case. The rationale behind this is that experience with PTS shows that formally optimized systems in many applications involve only a few number of wells, with little gain shown by adding further ones [e.g., *Guan and Aral*, 1999; *Zheng and Wang*, 1999; *Matott et al.*, 2006]. Accordingly, the results of the one-well case provide an initial approximation of the objective function value that can be expected for the five-well case. Because of the additional drawdown constraint, it is not assured that multiple wells will improve the fitness of the one-well optima.

[63] Figure 11 shows the convergences of median best fitness values over 10 ORs. Independently from the objective, the curves follow a sharp decline that is most pronounced within the first 1500 objective function evaluations, and then levels off. Each realization-specific tradeoff is unique and reflects the individual characteristic of each conductivity distribution and its influence on the respective objective function. After 3504 iterations, minimization of Q (Figure 11a) yields results with fitness values that are, generally, slightly higher than those found for the one-well problem. Apparently, further wells hardly improve a solution. In contrast, for minimization of mass extraction, the fitness of the best solutions is slightly (realization C) to significantly (E) lower.

[64] A very common feature is that one well of the optimized systems is assigned a disproportionately high extraction rate (compared to the extraction rates of the other four wells), and placed in close vicinity to the best one-well position. Alternatively, several wells of similarly high extraction rates are grouped in areas that have been identified as close optimal positions by the direct search method for single wells. Wells with much lower discharges (close to Q_{min}) are irregularly distributed over the area, and showed to have a negligible effect on the plume. Repeated attempts of the CMA-ES did not yield the exact same results, (i.e., the fittest well configuration was not reproduced). However, individual ORs placed the dominating wells in the same zones which are visualized in Figure 12a. Pumping rate

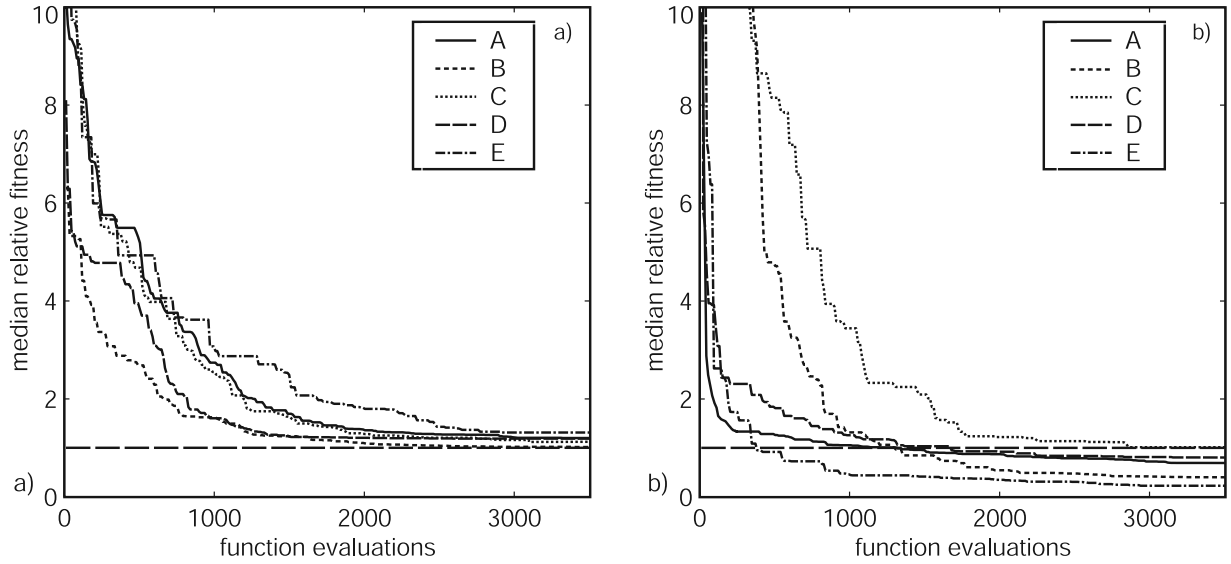

Figure 11. Convergence of median best fitness (objective function) values of 10 ORs for the five-well case ($k_c = 0.75\ k_0$, realizations A–E): (a) minimization of total pumping rate and (b) minimization of total mass extraction rate; fitness values are expressed as relative values, normalized to realization-specific global optima for the one-well case (horizontal dashed line at relative fitness of 1 marks equal fitness for both cases).

distribution among wells differed, even though the fitness values of the entire system were nearly the same. This reflects the higher degree of freedom for multiple well systems compared with the one well case.

[65] Figure 12b shows the concentration contour lines for realization A when the PTS that reduces mass extraction most is in operation, and well positions of close optimal solutions. Additionally, well positions of close optimal well schemes are displayed. Similar to the results when minimizing Q, all of the well schemes found by repeated ORs are singular, however accrue at preferable locations. Some wells of very low rates are also not properly adapted. Examples include the wells located at the southeastern corner of the well placement area, which only have a negligible effect on the plume. The paramount difference to the Q-optimized well schemes is the positioning of individual wells of the best five-well scheme, as compared to the one-well results. This is similar to the findings for the

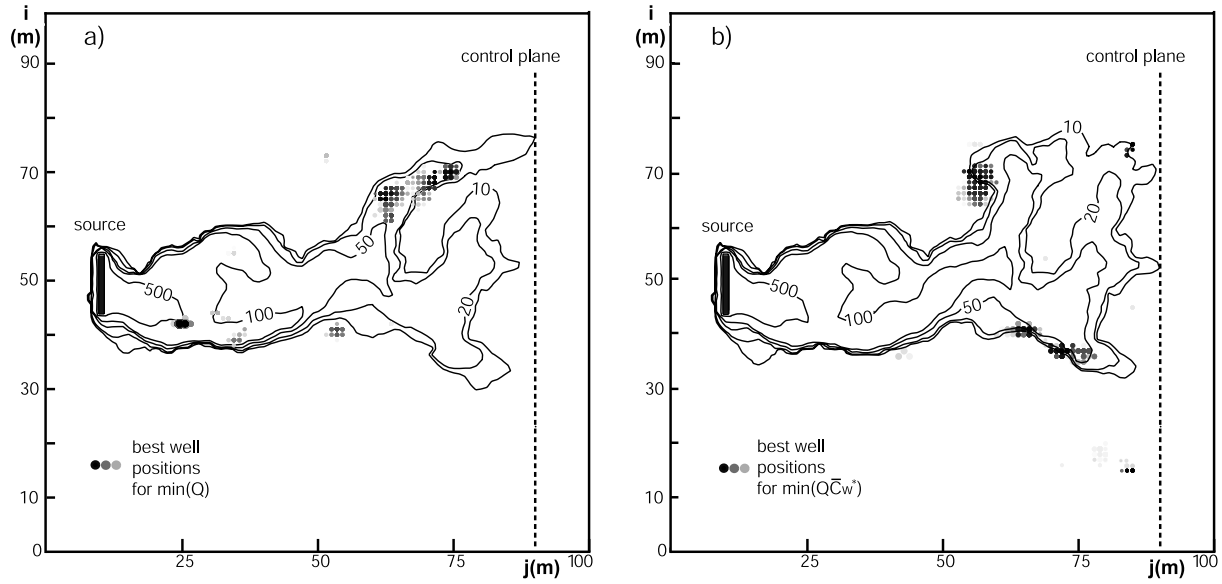

Figure 12. Concentration contours for heterogeneous aquifer example ($k_c = 0.75\ k_0$, realization A) if best five-well solution is operating: (a) contours for the solution minimizing pumping rate and (b) contours for the solution minimizing mass extraction rate. Circles show well positions of well schemes that are better than the median best of 10 ORs. Gray scale reflects fitness of specific well combination; that is, the darker the circle, the better the solution. Diameters of circles illustrate (relative) pumping rates.

other realizations: although the best one-well position can be part of an ideal five-well solution, there are cases where new preferred positions are determined. While minimizing Q favors further upstream and plume-centered well positions, the ideal well configurations for minimization of $Q\overline{C}_w^*$ tend to be laterally distributed to spread the plume. Obviously, the latter can be more efficiently achieved by installing multiple wells. Figure 12b illustrates that a major role is played by the specific form of the plume, which tends to divide when approaching the control plane. This downgradient area is also most attractive because of the lower contaminant concentrations. Accordingly, it is most efficient to manage the downgradient plume fingers separately.

7. Summary and Conclusions

[66] The presented study demonstrated the capacity of the CMA-ES with μ rank update to optimize PTS for concentration control of contaminated plumes. A central question related to how this optimization algorithm could be most efficiently applied to solve common design problems, such as pumping rate determination and well placement. This was discussed for two somewhat opposite design criteria: minimization of pumping rate and minimization of mass extraction. Both objectives are identified as main economic determinants of PTS if the technology is implemented for contaminant control rather than cleanup.

[67] In order to derive a largely valid assessment of the optimization approach, a huge number of dissimilar, singular problems have been set up. These basically vary in hydraulic and hydrochemical characteristics of the modeled aquifer. Such site-specific properties are expected to have a major influence on the variability of objective function values (i.e., the (multidimensional) fitness landscape, which can exhibit significant nonlinearities and nonconvexities). The huge computational effort, which occurs when thousands of model runs are required to iteratively calculate the objective function values, was maintained at a low level. This was at the expense of a more sophisticated modeling concept. One could consider temporally variable mass transfer processes or transient hydraulic conditions, as well as incorporate substrate limited degradation instead of first-order kinetics. Though simplifications are common for the demonstration of optimization methods in this field, the generic modeling results should be regarded as rough approximations of real world processes. Further, using more process-oriented models can lead to different optimized systems. The hypothetical application examples were set up as simplified surrogates. As such, they exposed the dominant role of hydraulic heterogeneity in relation to the complexity of the objective functions to be solved.

[68] In fact, even with the presented modeling concept, it was possible to reveal relevant characteristics of ideal well configurations. These observations are essential when an optimization problem has to be set up without an in depth knowledge of its complexity, of potential objective function optima, and of uncertainty in the site description. The optimal well positions of the one-well case aimed at minimizing the pumping rate ended up being concentrated in a preference area in the upgradient plume, which was in the vicinity of the source. Although the 100 different realizations of hydraulic conductivity distributions, and the variations in the degradation rates, are not sufficient for an exact statistical examination, this could be interpreted as a general tendency. Even if the optimal well position is not located in this preference area, we showed, in nearly all cases, that at least close optimal solutions exist. Thus it is recommendable to focus principally on this area if low decision space should be maintained in order to speed up the optimization procedure. In practice, the existence of a preference area can be utilized as an orientation for site investigation.

[69] When the extracted mass is a decision criterion, the optimal well schemes can differ significantly from those that are found for minimization of discharge only. If economic considerations suggest retaining low concentrations in the extracted groundwater, ideal well positions tend to be further downgradient from the source, in areas of lower contaminant concentration. Hence use would be made of the intrinsic contaminant degradation capacity of the aquifer. As could be expected, the distance from the source area increases as the intrinsic degradation capacity of the aquifer decreases. In summary, there are no preferable placement areas, rather, highly unique characteristics of the well layouts, which are mainly controlled by the spatial heterogeneity distribution.

[70] Multiple-well configurations do not inevitably consist of a combination of one-well solutions. This is demonstrated in the case of the five-well systems optimized regarding pumping rates and mass extraction. Optimal and close optimal positions detected for one-well cases appear only to be favorable for multiple well systems if the total pumping rate is minimized. Also, despite a broad well placement area and small-scale variations of hydraulic conductivity, a small number of wells seems to be sufficient. In principle, this is only valid for the type of problem considered here, and assumes that the well configurations found are (close to) optimal solutions. Of course, the existence of unconfined, less productive aquifers and the need for broad capture zones may increase the ideal number of wells.

[71] Again, it seems much harder to predict an ideal multiple-well system if contaminant concentration in the extracted groundwater is involved in the objective function. The positions of wells found in five-well cases rarely reflect good positions detected for single wells. We found that there is considerable potential to improve the extraction system by raising the number of wells. The five-well solutions, which were identified for the realizations in this study, benefit from local deviations of the plume, which can be relevant even if one well extracts at a much lower rate than the others.

[72] A central issue in this study was how to implement the CMA-ES in the most economical way. That is, how to obtain an optimal solution at a minimum number of model runs. We observed success rates of over 40% in most of the applications, even when the population size λ was held at a minimum. This also means that even higher success rates may be expected if the number of individuals per generation is increased. As the CMA-ES is a stochastic search method, it is reasonable to restart it and apply it in more than one optimization run OR. Therefore the settings used for λ (and offspring size μ) are basically recommendable.

[73] Although the application examples differed in complexity of objective functions and size of decision space, we

could obtain largely valid results for the ideal number of model runs I_{deal} per OR. It has been found that for the one-well case, the number of model runs should range between 250 and 400. In most applications within this range, no clear differences were observed in the total average number of model runs *MR* required to identify the optimum after multiple ORs. The exemplary five-well scenarios denote an appropriate setting of I_{deal} between 2500 and 3500, independent of the objective function to be solved. If it can be assumed that, first, as observed for the one-well case, these findings will not significantly change for other applications, and secondly, I_{deal} can be interpolated for other well cases, then we can conclude that the ideal number of model runs per CMA-ES application nonlinearly rises when the number of wells is raised. Concerning generations, however, a linear trend was identified. As a rough estimation, we calculated 50 generations per well, that is, 350 iterations for the one-well case and 3000 for the five-well case. This can be taken as a rule of thumb for further applications. The success rate, on the other hand, is more problem specific and could be controlled by the population size.

[74] A main determinant in any application is the desired accuracy of the solution (i.e., the closeness of the identified optimum to the global optimum). In many scenarios, the tolerance of 10%, which was applied, led to the acceptance of more than one well scheme. Thus, for one specific problem, well locations may vary slightly. It is not common to specify a higher precision level as the models used rarely accomplish higher accuracy. Regardless, the CMA-ES is principally capable of identifying the exact global optimum. However, as has been demonstrated by the application of a preliminary CMA-ES variant [*Bayer and Finkel*, 2004], stricter tolerance settings can considerably raise the total number of model runs required. In such cases, it is recommended that the number of generations per OR be increased, or decreased if the desired accuracy of results is less strictly set.

[75] In this paper, the modern version of CMA-ES, with weighted recombination and μ rank update, was utilized for the optimization of one common remediation technology, a pump-and-treat system. Together with the limited existing work on CMA-ES applications in our field, these studies reveal the high efficiency and unspecific applicability of this optimization algorithm. There is clear indication that such promising results for pumping well optimization can also be expected for further application problems in water research.

Notation

α_L longitudinal dispersivity (L).
α_T transversal dispersivity (L).
δ Dirac delta ($1/L^2$).
δ_{ij} Kronecker delta.
θ effective porosity.
λ population size (number of offspring), default $\lambda = 4 + \lfloor 3 \ln n \rfloor$.
ρ_1 scaling coefficient in penalty term (L^3/M).
ρ_2 exponent in penalty term (<1).
μ parent number (number of selected search points in the population), default $\mu = \lfloor \lambda/2 \rfloor$.
μ_{eff} constant $\mu_{eff} = 1/\sum_{i=1}^{\mu} w_i^2$ for normalization of steps in **p**, ("variance effective selection mass" in CMA-ES).
σ global step size of CMA-ES.
σ_{Y^2} variance of exponential variogram.
φ exponent of cost function.
ω recombination weight.
Φ_t specific success rate of detecting a solution J^* in generations 1, ..., t.
B orthogonal matrix with eigenvectors of **C** as columns.
C covariance matrix.
c_c learning rate for cumulation for the rank-one update of the covariance matrix, set default of $c_c = \frac{4}{n+4}$.
C_c solute concentration (M/L^3) of compound c.
C_c' sum of target concentration violations at control plane (M/L^3).
c_{cov} learning rate of **C**, set default value of $c_{cov} = \frac{2}{\mu_{eff}(n+\sqrt{2})^2} + \left(1 - \frac{1}{\mu_{eff}}\right) \min\left(1, \frac{2\mu_{eff}-1}{(n+2)^2+\mu_{eff}}\right)$.
C_{MCL} maximum contaminant concentration level (M/L^3).
C_w concentration (M/L^3) of contaminants (i.e., of one contaminant c) in water abstracted by well w, $\overline{C}_w$ is flow averaged concentration for more than one well.
c_σ learning rate for the cumulation for the step size control (backward time horizon of evolution path in CMA-ES), set default value of $c_\sigma = \frac{\mu_{eff}+2}{\mu_{eff}+3+n}$.
D diagonal matrix with square roots of the eigenvalues of **C** as diagonal elements.
D^* molecular diffusion coefficient (L^2/T).
$\mathbf{D}_c$ hydrodynamic dispersion tensor (L^2/T).
d_σ damping parameter for step size update of CMA-ES, set default value of $d_\sigma = 1 + 2\max\left(0, \sqrt{\frac{\mu_{eff}-1}{n+1}} - 1\right) + c_\sigma$.
h hydraulic head (L) at spatial model coordinates.
h' threshold for relative head change for drawdown constraint.
h_{ref} reference head distribution (L) for drawdown constraint.
I number of function evaluations per OR.
i row coordinate, i_w is well coordinate within interval $[i_{min}, i_{max}]$ (L).
I identity matrix.
I_{deal} ideal number of function evaluations for optimization algorithm (at $MR_{t,ideal}$).
I_Y correlation range of exponential variogram (L).
j column coordinate, j_w is well coordinate$[j_{min}, j_{max}]$, j_{CP} is column of control plane, (L).
J objective function, $J_1 = \min(Q)$, $J_2 = \min(Q\overline{C}_w^*)$.
J^* threshold value of objective function (e.g., global optimum).
k index of individuals in population ($k = 1, .., \lambda$).
k' index of parents sorted according to their fitness, best one first ($k' = 1, .., \mu$).
k_0 minimum degradation rate constant (1/T) that ensures C_{MCL} at the control plane.
k_c first-order degradation rate constant (1/T) of compound c.
K hydraulic conductivity tensor (L/T).

MR_t generation-specific number of total model runs required on the average to find one desired (optimal) solution (= $\lambda t/\Phi_t$); $MR_{t,ideal}$ is minimum of MR_t.
n problem dimension (number of decision variables).
n_{OR} number of optimization (CMA-ES) runs.
$\mathbf{p}$ sequence of successive normalized steps, ("conjugate evolution path" in CMA-ES).
Q well extraction rate (L^3/T) within range [Q_{min}, Q_{max}], Q' is rate per unit aquifer thickness (L^2/T).
t (index) of generation ($t = 1, .., t_{max}$).
t_{ideal} ideal number of generations for optimization algorithm (at $MR_{t,ideal}$).
t_m modeling time.
v groundwater flow velocity.
w well index (one-well case: $w = 1$; five-well case: $w = 1, \ldots, 5$).
$\mathbf{x}$ search point (individual).
$\langle \mathbf{x} \rangle_{\mu,t}$ center of mass of μ selected search points (recombined parent as origin for next generation).
$\mathbf{Z}$ covariance matrix of best individuals.
$\mathbf{z}$ random vector from multivariate normal distribution with zero mean and unity covariance matrix.

[76] **Acknowledgments.** Financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, contract BA 2850) for this project is gratefully acknowledged. We highly appreciate the suggestions of two anonymous reviewers, who helped to substantially improve this paper, particularly regarding the general assessment of CMA-ES performance. Special thanks are due to Claire Rose Aanes and Anna Cigolini for their assistance in preparing the manuscript.

References

Ahlfeld, D. P., and M. P. Sprong (1998), Presence of nonconvexity in groundwater concentration response function, *J. Water Resour. Plann. Manage.*, *124*, 8–14.

Aly, A. H., and R. C. Peralta (1999), Comparison of a genetic algorithm and mathematical programming to the design of groundwater cleanup systems, *Water Resour. Res.*, *35*, 2415–2425.

Bauer, S., C. Beyer, and O. Kolditz (2006), Assessing measurement uncertainty of first-order degradation rates in heterogeneous aquifers, *Water Resour. Res.*, *42*, W01420, doi:10.1029/2004WR003878.

Bayer, P., and M. Finkel (2003), Capture zone adaption using evolutionary algorithms, in *MODFLOW and More 2003: Understanding Through Modeling*, edited by E. Poeter et al., pp. 759–763, Int. Ground Water Modeling Cent., Golden, Colo.

Bayer, P., and M. Finkel (2004), Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones, *Water Resour. Res.*, *40*, W06506, doi:10.1029/2003WR002675.

Bayer, P., M. Morio, C. Bürger, B. Seif, M. Finkel, and G. Teutsch (2001), Funnel-and-gate vs. innovative pump-and-treat systems: A comparative economical assessment, in *Groundwater Quality: Natural and Enhanced Restoration of Groundwater Pollution*, *Publ. 275*, edited by S. F. Thornton and S. Oswald, pp. 235–244, Int. Assoc. of Hydrol. Sci., The Hague, Netherlands.

Bayer, P., S. Budesteanu, S. Bauer, and M. Finkel (2005), Innovative evolutionary algorithms to optimise innovative, in *Proceedings of ModelCare 2005 Fifth International Conference on Calibration and Reliability in Groundwater Modelling—From Uncertainty to Decision Making*, pp. 631–636, Int. Assoc. of Hydrol. Sci., The Hague, Netherlands.

Bear, J. (1972), *Dynamic of Fluids in Porous Media*, Elsevier, New York.

Chang, L.-C., and C.-T. Hsiao (2002), Dynamic optimal ground water remediation including fixed and operation costs, *Ground Water*, *40*, 481–490.

Chan Hilton, A. B., and T. B. Culver (2005), Groundwater remediation design under uncertainty using a robust genetic algorithm, *J. Water Resour. Plann. Manage.*, *131*, 25–34.

Cheng, A. H.-D., D. Halhal, A. Naji, and D. Ouazar (2000), Pumping optimization in saltwater-intruded coastal aquifers, *Water Resour. Res.*, *36*, 2155–2165.

Cieniawski, S. E., J. W. Eheart, and S. Ranjithan (1995), Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resour. Res.*, *31*, 399–409.

Clement, T. P., C. D. Johnson, Y. Sun, G. M. Klecka, and C. Bartlett (2000), Natural attenuation of chlorinated solvent compounds: Model development and field-scale application at the Dover site, *J. Contam. Hydrol.*, *42*, 113–140.

Deb, K., A. Anand, and D. Joshi (2002), A computationally efficient evolutionary algorithm for real-parameter optimisation, *Evol. Comput.*, *10*, 371–395.

Feyen, L., and S. M. Gorelick (2005), Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas, *Water Resour. Res.*, *41*, W03019, doi:10.1029/2003WR002901.

Freeze, R. A., and S. M. Gorelick (1999), Convergence of stochastic optimization and decision analysis in the engineering design of aquifer remediation, *Ground Water*, *37*, 934–954.

Gailey, R. M., and S. M. Gorelick (1993), Design of optimal, reliable plume capture schemes: Application to the Gloucester Landfill ground-water contamination problem, *Ground Water*, *31*, 107–114.

Guan, J. B., and M. M. Aral (1999), Optimal remediation with well locations and pumping rates selected as continuous decision variables, *J. Hydrol.*, *221*, 20–42.

Guan, J., and M. M. Aral (2004), Optimal design of groundwater remediation systems using fuzzy set theory, *Water Resour. Res.*, *40*, W01518, doi:10.1029/2003WR002121.

Haggerty, R., and S. M. Gorelick (1994), Design of multiple contaminant remediation: Sensitivity to rate-limited mass transfer, *Water Resour. Res.*, *30*, 435–446.

Hansen, N. (2006), The CMA evolution strategy: A comparing review, in *Towards a New Evolutionary Computation: Advances in Estimation of Distribution Algorithms*, edited by J. A. Lozano et al., pp. 75–102, Springer, New York.

Hansen, N., and A. Ostermeier (2001), Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.*, *9*, 159–195.

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

Huang, C., and A. S. Mayer (1997), Pump-and-treat optimization using well location and pumping rates as decision variables, *Water Resour. Res.*, *33*, 1001–1012.

Kaiser, R., O. Kolditz, and W. Zielke (2000), Automatic grid adaptation for multidimensional coupled processes in subsurface hydrosystems, in *Groundwater Updates, International Symposium 2000 on Groundwater (IAHR)*, edited by K. Sato and Y. Iwasa, pp. 339–344, Springer, New York.

Kern, S., S. D. Müller, N. Hansen, D. Büche, J. Ocenasek, and P. Koumoutsakos (2004), Learning probability distributions in continuous evolutionary algorithms—A comparative review, *Nat. Comput.*, *3*, 77–112.

Kolditz, O., and S. Bauer (2004), A process-orientated approach to compute multi-field problems in porous media, *Int. J. Hydroinformatics*, *6*, 225–244.

Liu, Y., and B. S. Minsker (2004), Full multiscale approach for optimal control of in-situ bioremediation, *J. Water Resour. Plann. Manage.*, *130*, 26–32.

Marryott, R. A. (1996), Optimal ground-water remediation design using multiple control technologies, *Ground Water*, *34*, 425–433.

Marryott, R. A., D. E. Dougherty, and R. L. Stollar (1993), Optimal groundwater management: 2. Application of simulated annealing to a field-scale contamination site, *Water Resour. Res.*, *29*, 847–860.

Matott, L. S., A. J. Rabideau, and J. R. Craig (2006), Pump-and-treat optimization using analytic element flow models, *Adv. Water Resour.*, *29*, 760–775, doi:10.1016/j.advwatres.2005.07.009.

Mayer, A. S., T. Kelley, and C. T. Miller (2002), Optimal design for problems involving flow and transport phenomena in subsurface systems, *Adv. Water Resour.*, *25*, 1233–1256.

McKinney, D. C., and M.-D. Lin (1996), Pump-and-treat groundwater remediation system optimization, *J. Water Resour. Plann. Manage.*, *122*, 128–136.

McPhee, J., and W. W.-G. Yeh (2004), Multiobjective optimization for sustainable groundwater management in semiarid regions, *J. Water Resour. Plann. Manage.*, *130*, 490–497.

Mulligan, A. E., and D. P. Ahlfeld (1999), Advective control of groundwater contaminant plumes: Model development and comparison to hydraulic control, *Water Resour. Res.*, *35*, 2285–2294.

Nelder, J. A., and R. Mead (1965), A simplex method for function minimization, *Comput. J.*, *7*, 308–313.

Orr, S., and A. M. Meystel (2005), Approaches to optimal aquifer management and intelligent control in a multiresolutional decision support system, *Hydrogeol. J.*, *13*, 223–246, doi:10.1007/s10040-004-0424-3.

Pebesma, E. J. (2004), Multivariable geostatistics in S: The GSTAT package, *Comput. Geosci.*, *30*, 683–691.

Ranjithan, S. R. (2005), Role of evolutionary computation in environmental and water resources systems analysis, *J. Water Resour. Plann. Manage.*, *131*, 1–2, doi:10.1061/(ASCE)0733-9496.

Reed, P. M., B. S. Minsker, and A. J. Valocchi (2000), Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation, *Water Resour. Res.*, *36*, 3731–3741.

Rogers, L. L., and F. U. Dowla (1994), Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, *Water Resour. Res.*, *30*, 457–481.

Schäfer, W. (2001), Predicting natural attenuation of xylene in groundwater using a numerical model, *J. Contam. Hydrol.*, *52*, 57–83.

Schwefel, H.-P. (1995), *Evolution and Optimum Seeking*, John Wiley, Hoboken, N. J.

Shieh, H.-J., and R. C. Peralta (2005), Optimal in situ bioremediation design by hybrid genetic algorithm-simulated annealing, *J. Water Resour. Plann. Manage.*, *131*, 67–78.

Smalley, J. B., B. S. Minsker, and D. E. Goldberg (2000), Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resour. Res*, *36*, 3043–3052.

Srivastava, P., J. M. Hamlett, P. D. Robillard, and R. L. Day (2002), Watershed optimization of best management practices using AnnAGNPS and a genetic algorithm, *Water Resour. Res.*, *38*(3), 1021, doi:10.1029/2001WR000365.

Stenback, G. A., S. K. Ong, S. W. Rogers, and B. H. Kjartanson (2004), Impact of transverse and longitudinal dispersion on first-order degradation rate constant estimation, *J. Contam. Hydrol.*, *73*, 3–14.

U.S. Environmental Protection Agency (1999), Hydraulic optimization demonstration for groundwater pump-and-treat systems, *Final Rep. EPA 542-R-99-011*, Washington, D. C.

Veith, T. L., M. L. Wolfe, and C. D. Heatwole (2003), Optimization procedure for cost effective BMP placement at a watershed scale, *J. Am. Water Resour. Assoc.*, *39*, 1331–1343.

Vink, K., and P. Schot (2002), Multiple-objective optimization of drinking water production strategies using a genetic algorithm, *Water Resour. Res.*, *38*(9), 1181, doi:10.1029/2000WR000034.

Wang, W., and D. P. Ahlfeld (1994), Optimal groundwater remediation with well location as a decision variable: Model development, *Water Resour. Res.*, *30*, 1605–1618.

Wolpert, D. H., and W.-G. Macready (1997), No free lunch theorems for optimization, *IEEE Trans. Evol. Comput.*, *1*, 67–82.

Yoon, J.-H., and C. A. Shoemaker (2001), Improved real-coded GA for groundwater bioremediation, *J. Comput. Civ. Eng.*, *15*, 224–231.

Zhang, Y., C. Zheng, B. Minsker, R. Greenwald, and R. Peralta (2004), *Application of Flow and Transport Optimization Codes to Groundwater Pump and Treat Systems*, vol. 1, *Final Tech. Rep. A252324*, 148 pp., Nav. Facil. Eng. Serv. Cent., Port Hueneme, Calif.

Zheng, C., and P. P. Wang (1999), An integrated global and local optimization approach for remediation system design, *Water Resour. Res.*, *35*, 137–146.

Zheng, C., and P. P. Wang (2002), A field demonstration of the simulation-optimization approach for remediation system design, *Ground Water*, *40*, 258–265.

P. Bayer and M. Finkel, Center for Applied Geosciences, University of Tübingen, Sigwartstrasse 10, Tübingen D-72076, Germany. (peter.bayer@uni-tuebingen.de)