Calibration and Reliability in Groundwater Modelling (Proceedings of the ModelCARE 99 Conference held at Zürich, Switzerland, September 1999). IAHS Publ. no. 265, 2000.

Computationally efficient methods for the quantification of uncertainties in groundwater modelling

HARALD KUNSTMANN*

Fraunhofer Institute for Atmospheric Environmental Research, Kreuzeckbahnstraße 19, D-82467 Garmisch-Partenkirchen, Germany

e-mail: kunstmann@ifu.fhg.de

WOLFGANG KINZELBACH

Institute of Hydromechanics and Water Resources Management, Swiss Federal Institute of Technology, HIL G 37.3, CH-8093 Zurich-Hoenggerberg, Switzerland

Abstract When conservative decisions have to be made in water management questions, the quantification of uncertainties is necessary. Monte Carlo techniques are suited for this analysis but usually require a huge computational effort. An alternative and computationally efficient approach is the First Order Second Moment (FOSM) method which directly propagates the uncertainty originating from parameter uncertainty into the result. We apply the FOSM method to both the groundwater flow and solute transport equations. It is shown how calibration on the basis of measured heads and concentrations yield the "Principle of Interdependent Uncertainty" that correlates the uncertainties of feasible transmissivities and recharge rates. The method is used to compute the uncertainty of steady state heads and of steady state solute concentrations. The method is illustrated by application to the Palla Road aquifer in semiarid Botswana, for which the quantification of the uncertainty range of groundwater recharge is of prime interest. The uncertainty bounds obtained by the FOSM method correspond well with the results obtained by the Monte Carlo method. It is shown that at the planned abstraction rate the probability of exceeding the natural replenishment of the Palla Road aquifer by overpumping is 30%.

THE FIRST ORDER SECOND MOMENT (FOSM) METHOD: UNCONDITIONAL MOMENTS

The numerical solution of the two-dimensional groundwater steady state flow equation for confined aquifers:

$$\vec{\nabla} \cdot (T\vec{\nabla} h) + q = 0 \tag{1}$$

is obtained by finite difference schemes for the unknown piezometric heads h. The solution for h is represented as vector **h** whose entries are the values at the grid points. The aquifer parameters **p**, i.e. T: transmissivity $[L^2\ T^{-1}]$, q: recharge/pumping rates $[L^3\ T^{-1}\ L^{-2}]$ must be given at all gridpoints. In the following it is assumed that both the transmissivity T (or hydraulic conductivity $k_f = T/m$, with m denoting the thickness of the aquifer), and the groundwater recharge q are lognormally distributed, i.e. $Y = \ln T$, and $z = \ln q$ are normally distributed stochastic variables. The aquifer parameters **p** (i.e.

* Formerly at: Institute for Hydromechanics and Water Resources Management, ETH Zürich.

log transmissivity, Y, or log recharge, Z) are assumed to be uncertain. They are described as having a mean value $\hat{\mathbf{p}}$ and a covariance $\mathbf{Cov(p)}$, which is, e.g. of Gaussian type. If the model setup is based on a zonation, in log transmissivity for instance, the entries of $\mathbf{Y}$ are given by the zonal values of Y: $\mathbf{Y} \in \Re^{N_Y}$ with N_Y equal to the number of zones in log transmissivity Y. The same holds analogously for $\mathbf{Z}$. If the zonal values are uncorrelated, $\mathbf{Cov}_{YY}$ and $\mathbf{Cov}_{ZZ}$ are of diagonal shape. The entries of the diagonal elements are the squared standard deviations of the uncertainties of the zonal values.

The first moment of the heads is approximated in first order accuracy by calculating the expectation value of the solution of equation (1) (Dettinger, 1981). The propagation of parameter uncertainties into the head uncertainties, given by the covariance matrix $\mathbf{Cov(h)} = \mathbf{Cov}_{hh} = E[(\mathbf{h}-\hat{\mathbf{h}})\cdot(\mathbf{h}-\hat{\mathbf{h}})$, can be approximated by:

$$\begin{aligned}\mathbf{Cov}_{hh} \stackrel{1}{=} \; & \mathbf{D}_{hY}\cdot\mathbf{Cov}_{YY}\cdot\mathbf{D}_{hY}{}^{T} + \mathbf{D}_{hZ}\cdot\mathbf{Cov}_{ZZ}\cdot\mathbf{D}_{hZ}{}^{T} + \\ & \mathbf{D}_{hZ}\cdot\mathbf{Cov}_{ZY}\cdot\mathbf{D}_{hY}{}^{T} + \mathbf{D}_{hY}\cdot\mathbf{Cov}_{YZ}\cdot\mathbf{D}_{hZ}{}^{T}\end{aligned} \tag{2}$$

The superscript T denotes the transpose of the matrix. The propagation of the parameter–head covariance matrices is given by:

$$\mathbf{Cov}_{hY} = E[(\mathbf{h}-\hat{\mathbf{h}})\cdot(\mathbf{Y}-\hat{\mathbf{Y}})] \stackrel{1}{=} \mathbf{D}_{hY}\mathbf{Cov}_{YY} \tag{3}$$

$$\mathbf{Cov}_{hZ} = E[(\mathbf{h}-\hat{\mathbf{h}})\cdot(\mathbf{Z}-\hat{\mathbf{Z}})] \stackrel{1}{=} \mathbf{D}_{hZ}\mathbf{Cov}_{ZZ} \tag{4}$$

(Dettinger & Wilson 1981; Townley & Wilson, 1985). The superscript above the equal sign indicates the order of the approximation. $\mathbf{D}_{hY} \equiv (\mathbf{D}_{\vec{Y}^T}\mathbf{h})\big|_{\vec{Y}=\hat{\vec{Y}}}$ and $\mathbf{D}_{hZ} \equiv (\mathbf{D}_{\vec{Z}^T}\mathbf{h})\big|_{\vec{Z}=\hat{\vec{Z}}}$ are the first derivatives of $\mathbf{h}$ with respect to the transpose of the parameter vectors $(\mathbf{Y},\mathbf{Z})$, evaluated at their mean values. They are the Jacobian matrices expressing the sensitivity of h_i with respect to Y_j and Z_j.

The FOSM method is now applied to the two-dimensional (2-D) steady state solute transport equation:

$$n_m(\vec{\nabla}\cdot(\vec{v}c) - \vec{\nabla}\cdot((D_{mol}+D_{disp})\cdot\vec{\nabla}c)) - qc_{in} = 0 \tag{5}$$

(e.g. Bear, 1972), with c: solute concentration [M L^{-3}], $\mathbf{v}$: 2-D pore velocity vector [L T^{-1}], n_m: mobile porosity [–], D_{mol}: molecular diffusion coefficient [L^2 T^{-1}], D_{disp}: dispersion tensor [L^2 T^{-1}], q: recharge per unit horizontal area [L^3 T^{-1} L^{-2}], c_{in}: pollutant concentration in inflows to aquifer [M L^{-3}], m: thickness of aquifer [L]. Assuming a lognormally distributed transmissivity, the velocity v is given by Darcy's law:

$$\mathbf{v} = -\frac{1}{n_m}\frac{e^Y}{m}\vec{\nabla}h \tag{6}$$

The first order approximation of mean concentrations $\hat{\mathbf{c}}$ (the first moments) are again obtained by the solution of the PDE using the mean values of the aquifer parameters $\hat{\mathbf{Y}}, \hat{\mathbf{Z}}$ and the piezometric heads $\hat{\mathbf{h}}$. The uncertainty propagation of the concentration in steady state can be derived in analogy to the flow equation and is given by:

$$
\begin{aligned}
\mathbf{Cov}_{cc} = {} & \mathbf{D}_{ch} \cdot \mathbf{Cov}_{hh} \cdot \mathbf{D}_{ch}{}^{T} + \mathbf{D}_{cY} \cdot \mathbf{Cov}_{YY} \cdot \mathbf{D}_{cY}{}^{T} + \mathbf{D}_{cZ} \cdot \mathbf{Cov}_{ZZ} \cdot \mathbf{D}_{cZ}{}^{T} \\
& + \mathbf{D}_{ch} \cdot \mathbf{Cov}_{hY} \cdot \mathbf{D}_{cY}{}^{T} + \mathbf{D}_{cY} \cdot \mathbf{Cov}_{Yh} \cdot \mathbf{D}_{ch}{}^{T} \\
& + \mathbf{D}_{ch} \cdot \mathbf{Cov}_{hZ} \cdot \mathbf{D}_{cZ}{}^{T} + \mathbf{D}_{cZ} \cdot \mathbf{Cov}_{Zh} \cdot \mathbf{D}_{ch}{}^{T} \\
& + \mathbf{D}_{cY} \cdot \mathbf{Cov}_{YZ} \cdot \mathbf{D}_{cZ}{}^{T} + \mathbf{D}_{cZ} \cdot \mathbf{Cov}_{ZY} \cdot \mathbf{D}_{cY}{}^{T}
\end{aligned} \tag{7}
$$

$\mathbf{D}_{ch}$, $\mathbf{D}_{cY}$, and $\mathbf{D}_{cZ}$ indicate the sensitivity matrices of the solute concentration with respect to the piezometric heads, log transmissivity Y, and log recharge Z. The computation of $\mathbf{Cov}_{cc}$ requires the evaluation of $\mathbf{Cov}_{hh}$, $\mathbf{Cov}_{hY}$, and $\mathbf{Cov}_{hZ}$, which are determined by the covariance propagation of the flow equation. The uncertainty of the concentrations σ_c is given by the diagonal elements of $\mathbf{Cov}_{cc}$.

THE FOSM METHOD: CONDITIONAL MOMENTS

The same head distribution can essentially be obtained for high recharge and high transmissivity, as well as for low recharge and low transmissivity. There is a set of combinations of Y and Z, which lead to a practically identical model output at the measurement locations. The feasible transmissivities and recharge rates are related to each other. This "Principle of Interdependent Uncertainty" can be derived from the equations of the unconditional FOSM method using the idea that if specific entries of $\mathbf{Cov}_{hh}$ or $\mathbf{Cov}_{cc}$ are known through measurement, the uncertainty propagation equations turn into relationships between the parameter covariances $\mathbf{Cov}_{ZZ}$ and $\mathbf{Cov}_{YY}$.

The measurement of a piezometric head at grid point location k gives knowledge of that head within its uncertainty bounds coming from measurement- or model errors. It therefore reduces its uncertainty, σ_{h_k}, to the model/measurement error $\sigma_h\big|_{Model-/Measurement\ Error}$. All covariance matrix elements composed of factors σ_{h_k}, such as $\mathbf{Cov}_{hh}$, $\mathbf{Cov}_{hY}$ and $\mathbf{Cov}_{hZ}$, then read:

$$\mathbf{Cov}_{hh}\big|_{k,l} = \rho_{h_k h_l}\sigma_{h_k}\sigma_{h_l} \qquad \text{if } k \text{ or l are measurement locations} \tag{8}$$

$$\mathbf{Cov}_{hY}\big|_{k,i} = \rho_{h_k Y_i}\sigma_{h_k}\sigma_{Y_i} = 0, \quad \forall i = 1,\ldots,N_Y, \text{ if } k \text{ is measurement location} \tag{9}$$

$$\mathbf{Cov}_{hZ}\big|_{k,j} = \rho_{h_k Z_l}\sigma_{h_k}\sigma_{Z_j} = 0, \quad \forall j = 1,\ldots,N_Z, \text{ if } k \text{ is measurement location} \tag{10}$$

because it is assumed that there is no correlation between model/measurement errors of heads and the parameters Y and Z. These equations determine the relationship between uncertainties in Y and Z when heads are used for conditioning.

If the uncertainty range of log recharge Z is of interest $\mathbf{Cov}_{ZZ}$ is found to be:

$$
\begin{aligned}
\mathbf{Cov}_{ZZ} = {} & (\tilde{\mathbf{D}}_{hZ}{}^{T} \cdot \tilde{\mathbf{D}}_{hZ})^{-1} \cdot \tilde{\mathbf{D}}_{hZ}{}^{T} \cdot (\mathbf{Cov}_{hh}\big|_{Model-/Measurement\ Error} - \tilde{\mathbf{D}}_{hY} \cdot \mathbf{Cov}_{YY} \cdot \tilde{\mathbf{D}}_{hY}{}^{T} \\
& - \tilde{\mathbf{D}}_{hZ} \cdot \mathbf{Cov}_{ZY} \cdot \tilde{\mathbf{D}}_{hY}{}^{T} - \tilde{\mathbf{D}}_{hY} \cdot \mathbf{Cov}_{YZ} \cdot \tilde{\mathbf{D}}_{hZ}{}^{T}) \cdot \tilde{\mathbf{D}}_{hZ} \cdot (\tilde{\mathbf{D}}_{hZ}{}^{T} \cdot \tilde{\mathbf{D}}_{hZ})^{-1}
\end{aligned} \tag{11}
$$

with:

$$\mathbf{Cov}_{ZY} = -(\widetilde{\mathbf{D}}_{hZ}{}^{T} \cdot \widetilde{\mathbf{D}}_{hZ})^{-1} \cdot \widetilde{\mathbf{D}}_{hZ}{}^{T} \widetilde{\mathbf{D}}_{hY} \cdot \mathbf{Cov}_{YY} \tag{12}$$

and $\widetilde{\mathbf{D}}_{hY} = \mathbf{P} \cdot \mathbf{D}_{hY}$, $\widetilde{\mathbf{D}}_{hZ} = \mathbf{P} \cdot \mathbf{D}_{hZ}$. P is the projection matrix whose entries indicate the locations of the measurement points. For each value of log transmissivity Y that lies within its uncertainty range (given by $\mathbf{Cov}_{YY}$) there will be a value of log recharge Z within its uncertainty range (given by $\mathbf{Cov}_{ZZ}$), so that the model output at the measurement points will remain the same. Equation (11) quantifies this interdependence of the uncertainty ranges of log transmissivity Y and log recharge Z. Similarly to the derivation of $\mathbf{Cov}_{ZZ}$, the uncertainty range of log transmissivity Y, $\mathbf{Cov}_{YY}$, at given uncertainty range of $\mathbf{Cov}_{ZZ}$, can be derived.

The uncertainty range of log recharge Z, if concentration measurements and uncertainty in log-transmissivity Y are given, can be evaluated to be:

$$\begin{aligned}
\mathbf{Cov}_{ZZ} = {} & ((\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ})^{T} \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ}))^{-1} \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ})^{T} \\
& \cdot \Big(\mathbf{Cov}_{cc}\big|_{\text{Model-/Measurement Error}} - \widetilde{\mathbf{D}}_{cY} \cdot \mathbf{Cov}_{YY} \cdot \widetilde{\mathbf{D}}_{cY}{}^{T} \\
& - \widetilde{\mathbf{D}}_{cY} \cdot \mathbf{Cov}_{YZ} \cdot \widetilde{\mathbf{D}}_{cZ}{}^{T} - \widetilde{\mathbf{D}}_{cZ} \cdot \mathbf{Cov}_{ZY} \cdot \widetilde{\mathbf{D}}_{cY}{}^{T} \\
& - \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hY} \mathbf{Cov}_{YY} \cdot \mathbf{D}_{hY}{}^{T} \cdot \widetilde{\mathbf{D}}_{ch}{}^{T} - \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ} \mathbf{Cov}_{ZY} \cdot \mathbf{D}_{hY}{}^{T} \cdot \widetilde{\mathbf{D}}_{ch}{}^{T} \\
& - \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hY} \cdot \mathbf{Cov}_{YZ} \cdot \mathbf{D}_{hZ}{}^{T} \widetilde{\mathbf{D}}_{ch}{}^{T} \\
& - \widetilde{\mathbf{D}}_{cY} \cdot \mathbf{Cov}_{Yh} \cdot \widetilde{\mathbf{D}}_{ch}{}^{T} - \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{Cov}_{hY} \cdot \widetilde{\mathbf{D}}_{cY}{}^{T} \\
& - \widetilde{\mathbf{D}}_{cZ} \cdot \mathbf{Cov}_{ZY} \cdot \mathbf{D}_{hY}{}^{T} \cdot \widetilde{\mathbf{D}}_{ch}{}^{T} - \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hY} \cdot \mathbf{Cov}_{YZ} \cdot \widetilde{\mathbf{D}}_{cZ}{}^{T}\Big) \\
& \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ}) \cdot ((\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ})^{T} \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ}))^{-1}
\end{aligned} \tag{13}$$

with:

$$\begin{aligned}
\mathbf{Cov}_{ZY} = {} & -((\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ})^{T} \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ}))^{-1} \\
& \cdot (\widetilde{\mathbf{D}}_{cZ} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hZ})^{T} (\widetilde{\mathbf{D}}_{cY} + \widetilde{\mathbf{D}}_{ch} \cdot \mathbf{D}_{hY}) \cdot \mathbf{Cov}_{YY}
\end{aligned} \tag{14}$$

The incorporation of head or concentration information into the Monte Carlo approach is much more tedious and costly; the conditional Monte Carlo simulation requires the application of the inverse method for each realization. Conditional Monte Carlo simulation generates realizations of log transmissivity (or log recharge respectively) according to its covariance structure, and fits the parameters of log recharge (or log transmissivity respectively) such that the weighted least square deviation χ^2 becomes minimal for each realization.

COMPARISON OF FOSM AND MONTE CARLO METHOD

As a check of the conditional FOSM method we compare the diagonals of $\mathbf{Cov}_{ZZ}$ as obtained by the "Principle of Interdependent Uncertainty" on one hand, and by Monte Carlo statistics on the other hand, for an aquifer in semiarid Botswana, the Palla Road aquifer. We used the results of Siegfried & Kinzelbach (1997) who calculated a deterministic model for the Palla Road aquifer by adjusting the hydraulic conductivities. The recharge pattern was divided into six different zones. The mean natural replenishment of the Palla Road aquifer was estimated to be around 4100 m^3 day^{-1}. Three hydraulic conductivity (k_f) zones were chosen. The uncertainties given in Table 1 for the hydraulic conductivity were assumed.

Table 1 Mean and standard deviation of lnk_f (Palla Road aquifer).

Zone	Mean (lnk_f)	Sigma (lnk_f)	Mean (lnk_f)	Sigma (lnk_f)
1	–2.40	0.40	0.09	0.04
2	1.10	0.40	2.99	1.25
3	2.62	0.20	13.79	2.79

When comparing the resulting diagonals of $\mathbf{Cov}_{ZZ}$, as calculated via equation (11) and conditioned by 10 head measurements, with a corresponding Monte Carlo analysis, the following values listed in Table 2 are obtained.

Table 2 Mean $\hat{Z}'$ and standard deviation σ_Z with $Z' = \ln(Q)$, $Q = (q\,x\,y)$ and recharge q in m day^{-1}, no measurement/model errors assumed, Monte Carlo versus FOSM method.

Zone	Monte Carlo: Mean (Z')	Sigma (Z')	FOSM: Mean (Z')	Sigma (Z')
1	–0.58	0.78	–0.47	1.21
2	2.77	0.30	2.80	0.35
3	3.13	0.32	3.21	0.59
4	3.03	0.26	3.11	0.22
5	1.73	0.29	1.81	0.32
6	2.76	0.81	3.00	0.98

The satisfying correspondence the between the Monte Carlo and FOSM method results shows the consistency of the proposed FOSM formulae. A CPU-time advantage of the FOSM method against the Monte Carlo method, of up to a factor of 100 is achieved in the example presented.

Now 19 TDS (total dissolved solids) concentration measurements are applied to obtain the uncertainty bounds of groundwater recharge. No comparison to Monte Carlo results will be given since our code UFLOW (Kunstmann, 1998) does not yet include an inverse stochastic modelling tool for solute transport. Here we assume uncertainties for log hydraulic conductivity of $\sigma_{\ln kf} = 1$ for zones 1 and 2, and $\sigma_{\ln kf} = 0.5$ for zone 3. We take the deviations between mean model output and measurement as square roots of the diagonal elements of $\mathbf{Cov}_{cc}\big|_{Model-/Measurement\ Error}$. The uncertainty bounds of log recharge are estimated via equation (13) and the uncertainty bound of the entire six recharge zones is then calculated. The uncertainty of this value is found to be around 2400 m^3 day^{-1}. The probability density function of the resulting lognormally-distributed recharge is illustrated in Fig. 1.

In 1996 the abstraction from the Palla Road wellfield was around 3100 m^3 day^{-1} which is also the long-term average value considered for future operation. The probability of exceeding the natural replenishment by overpumping is derived to be 30%.

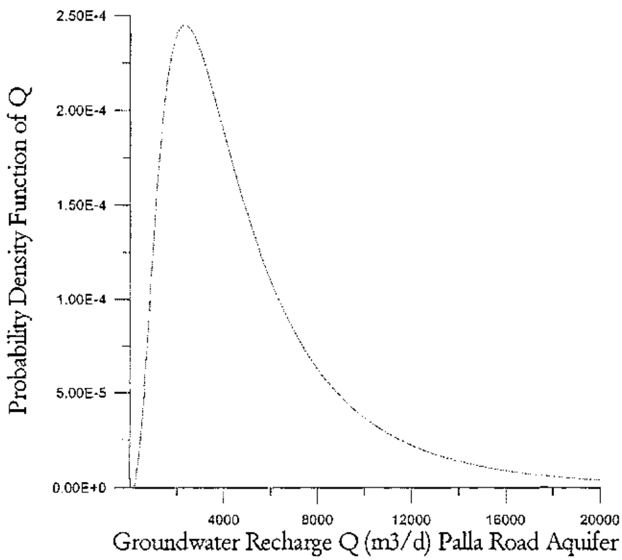

Fig. 1 Probability density function of groundwater recharge for the Palla Road aquifer.

SUMMARY AND CONCLUSIONS

Based on a first order Taylor series expansion, the First Order Second Moment (FOSM) method was extended and applied to the groundwater flow and solute transport equations in both unconditional and conditional analysis. The direct calculation of sensitivity matrices, and the effective method of multiplication of the matrices occurring, make its application to field studies now feasible. The FOSM method is a fast and reliable method for estimating the uncertainty of groundwater models. The two main limitations are the size of the covariance matrices appearing and the restriction of the size of the variability of parameters as it is inherently a linear method. Closed formulae obtained by the method clearly explain how conditioning information diminishes the uncertainty of the model output. The uncertainty arising from the non-uniqueness of the inverse problem is quantified. This "Principle of Interdependent Uncertainty" shows how the uncertainties of transmissivity and groundwater recharge are related to each other in the presence of head or concentration measurements. In addition, the influence of model/ measurement errors is quantified.

In general, all model computations should be subject to a stochastic analysis of uncertainty. Far too often deterministic results for design purposes are not critically used. Only quantified uncertainty gives the kind of safety required for robust management decisions. Also, the value of data can be quantified by conditional analysis. If the reduction of uncertainty by measurements is insufficient to provide enough safety, further data must be obtained instead of trying to do more elaborate modelling. The FOSM uncertainty analysis is possible as a kind of post-processing recommended for all modelling efforts.

REFERENCES

Bear, J. (1972) *Dynamics of Fluids in Porous Media*. Dover Publications, New York.

Dettinger, M. D. & Wilson, J. L. (1981) First order analysis of uncertainty in numerical models of groundwater flow. Part 1. Mathematical development. *Wat. Resour. Res.* **17**(1), 149–161.

Kunstmann, H. (1998) Computationally efficient methods for the quantification of uncertainties in groundwater modeling. PhD Thesis, Eidgenössische Technische Hochschule Zürich, Institut für Hydromechanik und Wasserwirtschaft, Zürich, Switzerland. Diss. ETH no. 12772.

Siegfried, T. & Kinzelbach, W. (1997) *Modeling of Groundwater Withdrawal and Its Consequences to the Aquifer—A Case Study in Semi-arid Botswana.* Institute of Hydromechanics and Water Resources Management, Swiss Federal Institute of Technology Zürich, Switzerland.

Townley, L. R. & Wilson, J. L. (1985) Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow. *Wat. Resour. Res.* **21**(12), 1851–1860.