Hydrol. Earth Syst. Sci. Discuss., 9, 9687–9714, 2012
www.hydrol-earth-syst-sci-discuss.net/9/9687/2012/
doi:10.5194/hessd-9-9687-2012

Hydrology and Earth System Sciences Discussions

Complexity versus simplicity: an example of groundwater model ranking with the Akaike Information Criterion

I. Engelhardt[1], J. G. De Aguinaga[2], H. Mikat[3], C. Schüth[1], O. Lenz[1], and R. Liedl[4]

[1]Technische Universität Darmstadt, Institute of Applied Geosciences, 64287 Darmstadt, Germany
[2]Bauhaus-Universität Weimar, Research Training Group 1462, 99423 Weimar, Germany
[3]Hessenwasser GmbH & Co. KG, Geschäftsbereich Wasserwirtschaft, 64521 Groß-Gerau, Germany
[4]Technische Universität Dresden, Institute for Groundwater Management, 01062 Dresden, Germany

Received: 8 July 2012 – Accepted: 13 August 2012 – Published: 21 August 2012

Correspondence to: I. Engelhardt (engelhardt@geo.tu-darmstadt.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

A groundwater model characterized by a lack of field data to estimate hydraulic model parameters and boundary conditions combined with many piezometric head observations was investigated concerning model uncertainty. Different conceptual models with a stepwise increase from 0 to 30 adjustable parameters were calibrated using PEST. Residuals, sensitivities, the Akaike Information Criterion (AIC), and the likelihood of each model were computed. As expected, residuals and standard errors decreased with an increasing amount of adjustable model parameters. However, the model with only 15 adjusted parameters was evaluated by AIC as the best option with a likelihood of 98 %, while the uncalibrated model obtained the worst AIC value. Computing of the AIC yielded the most important information to assess the model likelihood. Comparing only residuals of different conceptual models was less valuable and would result in an overparameterization of the conceptual model approach. Sensitivities of piezometric heads were highest for the model with five adjustable parameters reflecting also changes of extracted groundwater volumes. With increasing amount of adjustable parameters piezometric heads became less sensitive for the model calibration and changes of pumping rates were no longer displayed by the sensitivity coefficients. Therefore, when too many model parameters were adjusted, these parameters lost their impact on the model results. Additionally, using only sedimentological data to derive hydraulic parameters resulted in a large bias between measured and simulated groundwater level.

1 Introduction

Uncertainty is a key issue in hydrogeological modeling. Uncertainties are associated with parameter values, chosen scale, data quality, validity of boundaries, and initial conditions. Moreover, groundwater models are subject to several errors resulting from conceptual and stochastic uncertainty. Uncertainty in calibrated parameters can originate

HESSD
9, 9687–9714, 2012

from inaccuracies in field data, insensitivity with regard to changes in model parameters, and correlations within adjusted parameter sets (Singh et al., 2010). In many cases, measured field or laboratory data cannot be directly used to parameterize the model since they might reflect a different scale or boundary condition. Overparameterized models increase uncertainty since the information of the observations is distributed through all the parameters. To simulate a natural system with a numerical model, data have to be filtered, averaged and modified. A way to reduce this uncertainty is to select a parsimonious model, which provides good performance with as few calibrated parameters as possible.

There are several approaches to find this compromise between model fit and low number of calibration parameters (Hill and Tiedeman, 2007; Massmann et al., 2006). One of these approaches is the Akaike Information Criterion (AIC; Akaike, 1973). AIC is a probabilistic criterion based on the maximum likelihood and treats the problem of parsimonious model selection as an optimization problem across a set of proposed conceptual models (Burnham and Anderson, 2002). In addition, AIC allows to rank the models and also to improve the information about the model parameters. The AIC analysis determines the optimal model for a given data set. It identifies wherever the results of the selected model are already satisfactory or wherever an increased effort is needed by introducing more parameters into a model, so that AIC is able to select a more complicated model with a better fit to the observed data.

The application of the AIC is relatively new in groundwater modeling and still not standard, although it has been applied in several studies (e.g. Hill and Tiedeman, 2007; Foglia et al., 2007; Hill, 2006; Poeter and Anderson, 2005). In this study the Akaike Information Criterion was applied to a groundwater model developed for quantitative groundwater management issues in a region intensively used for industrial and drinking water production. Only very few data were available from pumping tests about the hydraulic properties of the aquifer at the water works and hydraulic parameters had to be estimated from sedimentological investigations. Several sedimentological investigations were obtained from borehole drillings conducted more than 100 yr ago and

HESSD
9, 9687–9714, 2012

were associated with high uncertainties deriving reliable hydraulic properties. On the other hand, long-term data, in form of high resolution groundwater level time series, were provided for the model calibration. This combination of a lack of model parameters and an extended calibration data set might result in a misleading model approach by using many adjustable parameters. In this study the uncertainty of different model approaches was assessed by gradually increasing the amount of adjustable model parameters to predict measured groundwater fluctuations. In addition to the residuals, the model assessment also considers the Akaike Information Criterion to define the optimal and reasonable model approach.

2 Materials and methods

2.1 Investigated field site

2.1.1 Geological setting

The study area is situated south of the city of Frankfurt and east of the Frankfurt International Airport in the German federal state Hesse. The site is located in the northern part of the Upper Rhine Graben (URG), which is part of the European Cenozoic Rift System (Ziegler and Dèzes, 2005). The URG, an approximately 300 km long and 40 km wide elongate lowland is flanked by uplift plateaus and terminated in the northern part by the WSW-ESE striking southern boundary fault of the Rhenish Massif, bounded to the west by the Mainz basin and to the east by the Hanau basin and the Odenwald Massif (Fig. 1a). The graben-filling sediments are of Eocene to Early Miocene and of Plio-/Pleistocene age (Berger et al., 2005). The subsidence of the graben resulted in up to 2000 m thick Tertiary deposits and more than 100 m thick fluvial Quaternary sediments (Anderle, 1968; Bartz, 1974). In the northernmost part of the URG between Mörfelden, Langen, Frankfurt, and the Lower Main area mainly fluvial sand and gravel with embedded clay lenses were deposited during the Pleistocene (Anderle, 1968).

HESSD
9, 9687–9714, 2012

The thicknesses of these deposits in the northern offset of the URG range between 10 and 40 m (Fig. 1b). Holocene eolian silty fine sand was deposited on top of this layer. The base of the Quaternary and Tertiary sand and gravel consists of Permian sandstone and conglomerates as well as Tertiary basalt.

2.1.2 Hydrogeology and hydrology

Average groundwater flow velocities within the Quaternary and Tertiary sand and gravel deposits are about 0.5 m d^{-1} and groundwater flows from the Sprendlinger Horst in the south-east towards the River Main. The depth to the groundwater table varies between 3 and 5 m near the River Main and gradually increases up to 15 m towards the south and East.

The long-term precipitation (1961–1993) averages around 675 mm a^{-1} as measured at the meteorological station in Frankfurt. About 15 % of the precipitation, thus 100 to 150 mm a^{-1}, can infiltrate into the groundwater (Berthold and Hergesell, 2005). The groundwater within this area is intensively used for drinking water and industrial purposes. Several water works are located within this region. In the water works Oberforsthaus, located directly in the study area, 18 production wells were operated. Groundwater extraction started already in 1894. About 100 yr later the water works was rebuild and then extraction rates increased within a few years from 560 000 $m^3 a^{-1}$ (1995) to 1.4 Mio $m^3 a^{-1}$ in 2000. Since 2005 the water works has been kept in stand-by operation. For sustainable groundwater management issues groundwater resources were recharged with treated water from the river Main to prevent an excessive groundwater table drop. Surface water was infiltrated by horizontal pipes and a small pond (named Jacobi Pond). During periods of high groundwater extraction rates treated surface water infiltration reached up to 35 to 40 % of the extracted groundwater volume and was reduced to about 25 % in periods with average extraction rates. The artificial groundwater recharge stopped in 2005 when the water works changed to stand-by operation.

HESSD
9, 9687–9714, 2012

2.2 Numerical model set-up

2.2.1 Discretization

The geological structure of the investigated Quaternary aquifer consists of a complex system of high and low permeable layers. Nine lithological units were identified from the borehole drillings. For translation of the complex geological information into a numerical model some simplifications were necessary. All geological information obtained from drillings and geological maps were summarized into three hydrostratigraphic layer (Fig. 2): (i) dominated by high permeable aquifer material (gravel and coarse sand), (ii) dominated by medium and low permeable aquifer material (medium and fine sand), (iii) dominated by high permeable material (gravel and coarse sand), and the impermeable aquifer base (iv) containing silt, clay, sandstone, limestone, or basalt. Then, 15 profiles were constructed containing these three hydrostratigraphic layer. Geological information between the profiles were interpolated to estimate the top and bottom of the three hydrostratigraphic layer (Fig. 2).

Based in these simplifications the spatial discretization contained 22 680 grid cells. The temporal discretization for the simulation period between 1990 to 2009 included 379 stress periods to capture the simulated period with monthly collected piezometric pressure heads.

2.2.2 Hydraulic properties

Only very few data were available about hydraulic conductivities and storage of the aquifer layers. Within a layer, several micro layers may be present and an averaging technique was applied to account for these heterogeneities. First, all data obtained from the geological description of the borehole data were used to assign an initial estimate on hydraulic conductivities and storage coefficients to each of the nine lithological units. For each of the three hydrostratigraphic layer an equivalent hydraulic conductivity

HESSD
9, 9687–9714, 2012

and storage coefficient was then calculated to account for the contribution of each lithological unit to the three hydrostratigraphic layers (Fig. 3).

As an example, the equivalent hydraulic conductivity (K_{eq}) of hydrostratigraphic layer 1 around well A was obtained by calculating the weighted arithmetic average of the lithological units with:

$$K_{eq} = \frac{d1A \cdot K_{gravel} + d2A \cdot K_{coarse\ sand} + d3A \cdot K_{fine\ sand} + d4A \cdot K_{gravel}}{d1A + d2A + d3A + d4A} \quad (1)$$

with

K_{eq} = equivalent hydraulic conductivity

dA = thickness of the layer at well A

$1, 2, \ldots$ = number of the layer

K = hydraulic conductivity according to the sedimentological description of the lithological unit

Equivalent hydraulic conductivities and storage values were interpolated over the model domain for each of the three hydrostratigrahic layers and subdivided into ten conductivity and storage zones, respectively (Fig. 4). Hydraulic conductivity and storage zones showed a different pattern and frequency in of the three layers or not developed at all. The interpolation of the equivalent hydraulic conductivity zones failed around geological structures such as faults. Therefore, a final manual adjustment of the hydraulic parameters to maintain relevant geological features was necessary.

2.2.3 Numerical model boundaries

The standard finite-difference model MODFLOW (Harbaugh et al., 2005) was used for the flow simulations. Groundwater levels measured in 47 observation wells in 1990 were interpolated and assigned as initial head distribution (Fig. 5).

The main inflow into the groundwater resulted from recharge that varied monthly over the investigated 20 yr. Further groundwater inflow was caused by surface water infiltration from the Jacobi Pond. Groundwater outflow mainly occurred by exfiltration

HESSD
9, 9687–9714, 2012

into the river Main (Fig. 5). The stage of the river Main was adjusted monthly during the investigated period by applying a linear interpolation between two hydrological stations close to the model domain: Frankfurt Osthafen (4 km upstream) and Raunheim (16 km downstream). The water level of the Jacobi Pond was assumed to remain constant during the investigated period since groundwater levels near the pond also prevailed fairly constant. Leakage between groundwater and surface water was adjusted manually. Along the south-west boundary, groundwater flowed out of the model domain towards the water works Goldstein, which started operation in 1995. This subsurface outflow was accounted for by a general head boundary. The piezometric head outside of the model domain was given by the monthly measured groundwater level at the pumping wells of the water works Goldstein. Within the model domain the water works Oberforsthaus operated about 18 pumping wells between 1990 and 2005. The monthly measured extraction rates were corrected by the injected artificial recharge, and resulting extraction volumes were assigned at the water works location.

2.2.4 Model calibration

The non-linear parameter estimator PEST (Doherty, 2010) was used for the automated model calibration through an inverse parameter estimation process based on the Gauss-Marquardt-Levenberg method. PEST minimizes discrepancies between model simulated outputs and the corresponding measurements by minimizing the weighted sum of squared differences between the respective values. PEST also computes the sensitivities with regard to selected parameters at all observation points. These sensitivities provide a measure of how much a simulated value changes in response to a perturbation of an adjustable parameter (Hill and Tiedeman, 2007).

Piezometric heads collected at 41 observation wells between 1990 and 2009 were used for the model calibration (Fig. 5). For a better overview, observation wells were categorized into six groups: (i) near Jacobi Pond, (ii) near the River Main, (iii) southern area, (iv) western area, (v) northern area, and, (vi), around the water works Oberforsthaus (Fig. 5) to account for the different factors influencing the hydraulic pattern of the investigated region. Hydraulic conductivities and storage coefficients were estimated using PEST. First guesses of these parameters were assigned as derived from sedimentological interpretation of the borehole data as described previously for the estimation of the equivalent hydraulic conductivities and storage coefficients (Figs. 3 and 4).

After calibration of the hydraulic parameters with the measured piezometric pressure heads collected at 41 observation wells a model validation was conducted. This validation used piezometric pressure heads measured at six further observation wells representing each observation group. These observations wells were not used for the parameter estimation during the inverse modeling procedure.

2.3 Model assessment using the Akaike Information Criterion (AIC)

The computation of the AIC allows the selection of a parsimonious model that uses the smallest number of parameters needed to provide an adequate approximation to the measured data. Thus, a compromise between a "good" fit and a small number of parameters can be found. For the model selection, AIC was computed for each model. All models were calibrated to the same data set, and the model with the smallest AIC is regarded as the optimal one of all proposed models (Massmann et al., 2006). For normally distributed residuals, AIC equals (Burnham and Anderson, 2002):

$$\mathrm{AIC} = n\ln(\hat{\sigma}^2) - n\ln(2\pi) - n + 2p \quad (2)$$

where p equals the number of estimated model parameters plus one, n the number of observations, and $\hat{\sigma}^2$ represents an estimate of the variance of residuals, which is given by:

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \quad (3)$$

where ε stands for the residuals: observed minus calculated values.

HESSD
9, 9687–9714, 2012

Groundwater model ranking with the Akaike Information Criterion

I. Engelhardt et al.

The first term in Eq. (2) represents the lack of the model fit, which decreases when more parameters are included. The two middle terms are constants for a specific data set, and are not affected if parameters are added or removed from the models. The last term can be seen as "penalty" term for incorporating more parameters, since it gets larger then.

Akaike (1973) defined weights w_j to obtain a relative measure of the likelihood of a model for a given set of N models. These weights are expressed as:

$$w_j = \exp(-0.5\Delta_j) \Big/ \sum_{j=1}^{N} \exp\left(-0.5\Delta_j\right) \quad (4)$$

where j is the counter of models and $\Delta_j = \mathrm{AIC}_j - \mathrm{AIC}_{\min}$ is denoting the AIC difference to the smallest AIC of all considered models. The larger the AIC difference of a model, the less plausible it is to be the best one.

First, the uncalibrated model using only sedimentological information was applied (Model 1), then the five most widespread horizontal hydraulic conductivities were estimated (Model 2). In Model 3, all horizontal hydraulic conductivities were considered and vertical hydraulic conductivities were tied by a factor of 0.1 ($K_\mathrm{v} = K_\mathrm{H}/10$). The next model (Model 4) computed additionally to the horizontal hydraulic conductivity the five most widespread storage coefficients. Model 5 estimated all horizontal conductivities and storage coefficients. In Model 4 and 5 vertical hydraulic conductivities were still tied. Then in Model 6 all horizontal and vertical conductivities were estimated independently and in addition the five most widespread storage coefficients. Finally, Model 7 independently estimated all horizontal and vertical hydraulic conductivities and all storage coefficients for all zones of the model domain giving a total amount of 30 adjustable parameters (Table 1).

HESSD
9, 9687–9714, 2012

3 Results

3.1 Sensitivity analysis

For each observation group time-dependent dimensionless sensitivity coefficients of the measured piezometric pressure heads are shown in Fig. 6. The pattern of the sensitivities between the groups was independent from the number of parameters used in the automated model calibration. Sensitivity was always highest for the northern area as the best optimization results could be obtained for this region. Lowest sensitivities were always computed for observation wells near the River Main and the Jacobi Pond (Fig. 6). These low sensitivities resulted from the impact of surface water-groundwater exchange on the groundwater level that was mostly driven by the stage of the surface water and leakage through the colmation layer. However, due to missing information about these processes both boundary conditions were not adjusted in the automated model calibration resulting in these low sensitivity coefficients of the influenced observation wells. Thus, hydraulic conductivities of the colmation layers of the Jacobi Pond and Main with fixed with $5 \times 10^{-6}\,\mathrm{m\,s^{-1}}$ and $1.2 \times 10^{-5}\,\mathrm{m\,s^{-1}}$, respectively.

Sensitivities of all observation groups followed changes of the groundwater level fluctuations and decreased when the groundwater extraction stopped in 2005.

Sensitivities were compared for the different models that differ by the amount of adjustable parameters from initially 5 to finally 30 parameters. The PEST optimization of five parameters revealed highest sensitivity coefficients (Fig. 5). Increasing the amount of adjustable parameters decreased the sensitivity of the piezometric heads. Therefore, considering a model set-up with large numbers of observation data, the amount of adjusted model parameters must be chosen with care to prevent an overparameterization and to maintain the influence of the measured data for the model calibration.

3.2 AIC model ranking

Computing the AIC allowed to evaluate the conceptual models with respect to their complexity and parameter uncertainty. Since Eq. (2) has to be minimized, the lowest AIC value indicates the best model. Model complexity was gradually increased from the uncalibrated stage to 30 adjustable model parameters (Table 2). This increase in complexity was linearly penalized; as expected, by considering more parameters the model fit steadily improved until reaching a constant level with no or little improvement (Fig. 7). By summing the model fit and the penalty the models can be ranked (s. "AIC line" in Fig. 7).

The scale of the y-axis is omitted in Fig. 7 since AIC is a relative measure and the absolute values are meaningless. Important are, however, the differences to the best model (AIC Δ_j; Table 2). The lowest AIC value, therefore the best option, is achieved by using Model 4 having 15 adjustable parameters. The highest (worst) AIC value was calculated for the uncalibrated model. However, Model 2 (5 adjustable parameters) and Model 7 (30 adjustable parameters) were assessed as similar worse due to a lack of model fit to the data (Model 2) or an unjustified complexity (Model 7).

Relative Akaike weights (AIC w_j), Eq. (4), were computed for all models to express in percent the likelihood of a model, where a likelihood of 100 % means that the corresponding model alone is regarded to represent the "best option", while a likelihood of 0 % corresponds to a model that has absolutely no support when compared to other models. In our case, the model selected as optimal (AIC $\Delta_j = 0$) is associated with a likelihood of about 98 %. All other models have practically no support according to the AIC and are either underparameterized (Models 1, 2 and 3) or clearly overparametrized (Models 5, 6 and 7).

HESSD
9, 9687–9714, 2012

3.3 Optimization results

3.3.1 Obtained residuals

The model calibration was based on monthly measured piezometric heads in 41 observation wells between 1990 and 2009. Measurements were not available every month at every observation well, giving a total sum of 5081 piezometric head data for the calibration. Computed and measured piezometric heads of the model with the smallest AIC (Model 4) are given for each observation group in Fig. 8.

Within the western area groundwater levels varied over 3 m. This fluctuation resulted from the impact of the water works Goldstein located south of this region. Around the water works Oberforsthaus groundwater levels varied over a range of 1.2 m. This lower groundwater level drop can be explained by the artificial recharge measures at this water works. Within the southern part groundwater levels varied up to 2.1 m and were also displaying the impact of the water works Goldstein. Within the northern area, near the River Main and the Jacobi Pond, groundwater levels remained almost constant with only minor fluctuations associated with changes in precipitation, soil cover and river discharge during the year.

Within most regions, measured groundwater levels were reasonably well reproduced by the flow model (Table 3, Fig. 8). The smallest standard error of the weighted residuals of 0.22 to 0.23 m was obtained near the Jacobi Pond (group 5; Table 3). Around the water works Oberforsthaus (group 1) and within the western part (group 4) computed standard errors of the weighted residuals increased to 0.47 to 0.51 m, which can still be assessed as sufficient with respect to the high uncertainties in boundary conditions and model parameter values. Calibration results obtained for observation wells located near the river Main (group 6) showed the highest calibration errors with up to 1.34 m that might result from the interpolation of the river stage within the model domain.

Six observation wells were used for the model validation giving 1445 observations or 22 % of initial available calibration data. Groundwater levels simulated by the optimal

HESSD
9, 9687–9714, 2012

model matched measured values reasonably well and demonstrated that model parameters were estimated within a reliable range (Table 3).

By increasing the amount of adjustable hydraulic conductivities, mean residuals decreased and the standard error of weighted residuals improved from 1.18 m (Model using only sedimentological information) to 0.74 m (Model 7).

3.3.2 Obtained parameter estimates

Only very few information was available from field investigations about hydraulic conductivity and storage. The ratio between vertical and horizontal hydraulic conductivities was assumed to be 1 : 10 within Models 2, 3, 4. Applying this assumption and additionally calibrating the most widespread storage coefficients (Model 4) was assessed by the AIC as the most certain model with a likelihood of 98 %. Hydraulic conductivities were estimated distinctively higher by PEST in most regions than derived from sedimentological information (Table 4). These differences may result from the impact of highly permeable fractures within the aquifer that were missed by the interpretation of the borehole data.

4 Concluding remarks

Limited information about model parameters and model boundaries linked with large amounts of observation data about groundwater fluctuations were investigated for model uncertainty. Such combination of model parameters and calibration data could lead to an overparameterized conceptual model. A sensitivity analysis clearly demonstrated that the sensitivity at all observation points decreased by increasing the number of adjustable parameters. This reduced the influence of collected field data to constrain the model calibration. Computing the AIC allowed to evaluate the benefit of adjusting high numbers of model parameters. The simplest as well as the complex models were rejected by the Akaike Information Criterion since they are likely to be under- or overparameterized. In our study, the best model is of "medium complexity". It calibrates five of ten storage coefficients and all ten horizontal conductivities while keeping the vertical conductivities tied by one order of magnitude lower. Remaining storage coefficients derived from sedimentological investigations could not be improved by further calibration. The results of the optimal model approximately resemble observed hydraulic piezometric heads while keeping estimated model parameters at a minimum. The evaluation of the AIC can improve model confidence as it avoids an under- or overparameterization of a conceptual model for a given data set.

Acknowledgements. We thank Elke Duhr and Stefan Pohl (both Hessenwasser GmbH & Co KG) and Sebastián Fernández for preparing and compiling the large amount of geological and hydraulic data.

References

Akaike, H.: Information theory and an extension of the maximum likelihood principle, in: Breakthroughs in Statistics, vol. 1, Foundations and Basic Theory, edited by: Kotz, S. and Johnson, N. L., Springer-Verlag, New York, USA, 610–624, 1973.

Anderle, H.-J.: Die Mächtigkeiten der sandig-kiesigen Sedimente des Quartärs im nördlichen Oberrhein-Graben und der östlichen Untermain-Ebene, Notizbl. Hess. L.-Amt Bodenforsch., 88, 185–196, Wiesbaden, 1968.

Anderle, H.-J. and Golwer, A.: Tektonik, in: Erläuterungen Geol. Kt. Hessen 1 : 25000, Bl. 5917, Kelsterbach, Wiesbaden, 50–64, 1980.

Bartz, J.: Die Mächtigkeit des Quartärs im Oberrheingraben, in: Approaches to Taphrogenesis, edited by: Illies, J. H. and Fuchs, K., Schweitzerbarth, Stuttgart, 78–87, 1974.

Berger, J.-P., Reichenbacher, B., Becker, D., Grimm, M., Grimm, K., Picot, L., Storni, A., Pirkenseer, C., Derer, C., and Schaefer, A.: Paleogeography of the Upper Rhine Graben (URG) and the Swiss Molasse Basin (SMB) from Eocene to Pliocene, Int. J. Earth Sci., 94, 697–710, doi:10.1007/s00531-005-0475-2, 2007.

Berthold, G. and Hergesell, M.: Flächendifferenzierte Untersuchungen zu möglichen Auswirkungen einer Klimaänderung auf die Grundwasserneubildung in Hessen, INKLIM 2012 – Integriertes Klimaschutzprogramm, Abschlussbericht, Hessisches Landesamt für Umwelt und Geologie, Wiesbaden, 2005.
Burnham, K. P. and Anderson, D. R.: Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd Edn., Springer-Verlag, New York, USA, 2002.
Doherty, J.: PEST – Model-independent Parameter Estimation, User's Manual, 5th Edn., Watermark Numerical Computing, Brisbane, Australia, 2010.
Foglia, L., Mehl, S. W., Hill, M. C., Perona, P., and Burlando, P.: Testing alternative ground water models using cross-validation and other methods, Ground Water, 45, 627–641, doi:10.1111/j.1745-6584.2007.00341.x, 2007.
Harbaugh, A. W.: MODFLOW-2005, The US Geological Survey Modular Ground-Water Model – the Ground-Water Flow Process, US Geological Survey Techniques and Methods 6-A16, Reston, VA, USA, 2005.
Hill, M. C.: The practical use of simplicity in developing ground water models, Ground Water, 44, 775–81, 2006.
Hill, M. C. and Tiedeman, C. R.: Effective Groundwater Model Calibration – with Analysis of Data, sensitivities, predictions, and uncertainty, John Wiley & Sons, Hoboken, USA, 2007.
Lahner, L. and Toloczyki, M.: Geowissenschaftliche Karte der Bundesrepublik Deutschland 1 : 2 000 000, Geologie, Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, 2004.
Massmann, C., Birk, S., Liedl, R., and Geyer, T.: Identification of Hydrogeological Models: Application to Tracer Test Analysis in a Karst Aquifer. in: Calibration and Reliability in Groundwater Modelling: From Uncertainty to Decision Making, Proceedings of ModelCARE'2005, June 2005, The Hague, The Netherlands, IAHS Publ., 304, 2006.
Poeter, E. P. and Anderson, D. R.: Multimodel ranking and inference in ground water modeling, Ground Water, 43, 597–605 doi:10.1111/j.1745-6584.2005.0061.x, 2005.
Singh, A., Mishra, S., and Ruskauff, R.: Model averaging techniques for quantifying conceptual model uncertainty, Ground Water, 48, 701–715, 2010.
Ziegler, P. A. and Dèzes, P.: Evolution of the lithosphere in the area of the Rhine Rift System, Int. J. Earth Sci., 94, 594–614, doi:10.1007/s00531-005-0474-3, 2007.

Table 1. Calibrated models analyzed with AIC.

Model	Number of adjusted parameters during automated model calibration	
	Conductivities	Storage coefficients
1	based on sedimentological data	
2	5	0
3	10	0
4	10	5
5	10	10
6	20	5
7	20	10

HESSD
9, 9687–9714, 2012

Table 2. Likelihood of the flow models according to the Akaike weights (AIC w_j).

Model	AIC Δ_j	AIC w_j
1	148.5	0.00
2	37.8	0.00
3	10.1	0.01
4	0.0	0.98
5	8.8	0.01
6	18.9	0.00
7	29.8	0.00

HESSD
9, 9687–9714, 2012

HESSD
9, 9687–9714, 2012

Table 3. Standard error of the weighted residuals of the six observation 504 groups and total sum of squared weighted residuals for each of the 505 seven conceptual models.

	Standard error of weighted residuals (m)						
Model	G1	G2	G3	G4	G5	G6	Total Residuals
1	0.7347	1.966	0.752	0.728	0.265	1.393	1.18
2	0.472	0.640	0.499	0.494	0.219	1.341	0.750
3	0.470	0.607	0.628	0.503	0.233	1.306	0.745
4	0.475	0.602	0.591	0.507	0.225	1.317	0.745
5	0.473	0.599	0.636	0.509	0.234	1.307	0.744
6	0.473	0.600	0.628	0.508	0.234	1.305	0.744
7	0.472	0.613	0.574	0.5084	0.217	1.306	0.743

Group 1: Around water works *Oberforsthaus*.
Group 2: Southern area.
Group 3: Northern area.
Group 4: Western area.
Group 5: Near Jacobi Pond.
Group 6: Near River Main.
Total residuals: obtained for 5081 piezometric pressure head data.

Table 4. Comparison of the initial guesses of the hydraulic conductivity 517 based on sedimentological information and values estimated by 518 PEST for the optimal model (Model 4).

	Hydraulic conductivity ($m s^{-1}$)			
	Horizontal		Vertical	
Zone	Estimated from sedimentological information	Estimated by PEST	Estimated from sedimentological information	Estimated by PEST
1	5.6×10^{-3}	1.7×10^{-1}	5.6×10^{-4}	1.7×10^{-2}
2	3.8×10^{-3}	4.8×10^{-1}	3.8×10^{-4}	4.8×10^{-2}
3	5.3×10^{-3}	1.5×10^{-1}	5.3×10^{-4}	1.5×10^{-2}
4	6.8×10^{-3}	3.5×10^{-2}	6.8×10^{-4}	3.5×10^{-3}
5	8.3×10^{-3}	5.7×10^{-3}	8.3×10^{-4}	5.7×10^{-4}
6	9.8×10^{-3}	1.8×10^{-2}	9.8×10^{-4}	1.8×10^{-3}
7	1.1×10^{-2}	2.0×10^{-2}	1.1×10^{-3}	2.0×10^{-3}
8	1.3×10^{-2}	6.8×10^{-2}	1.3×10^{-3}	6.8×10^{-3}
9	1.4×10^{-2}	6.6×10^{-2}	1.4×10^{-3}	6.6×10^{-3}
10	1.0×10^{-7}	4.3×10^{-7}	1.0×10^{-8}	4.3×10^{-8}

HESSD
9, 9687–9714, 2012

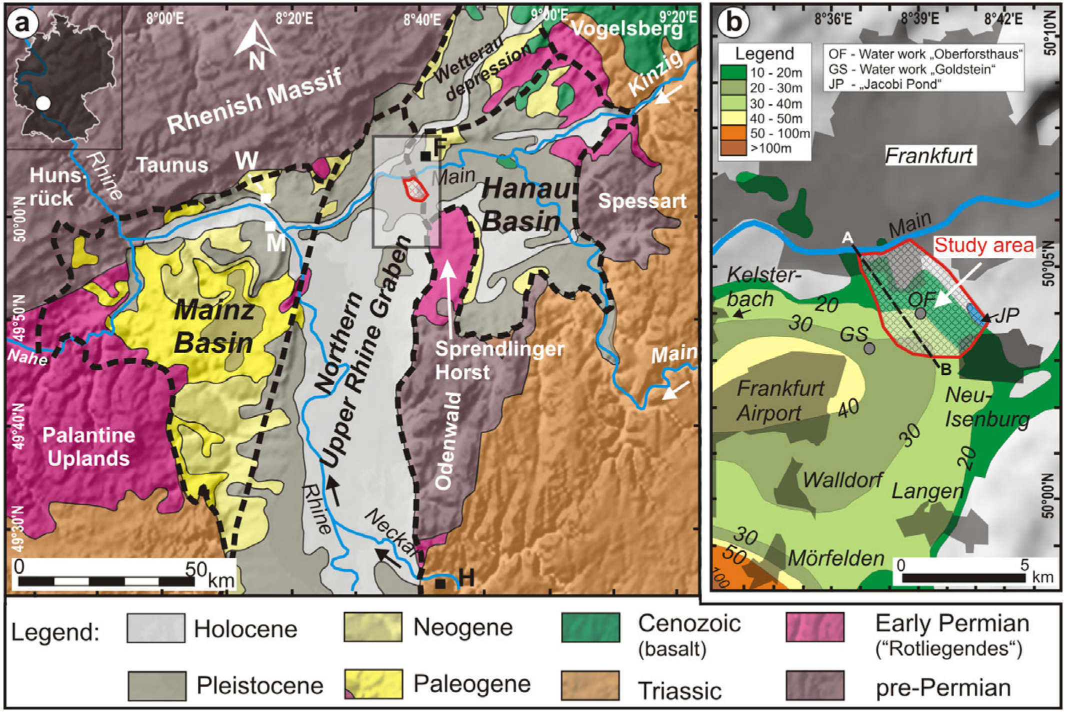

Fig. 1. (a) Simplified geological map showing the northern part of the Upper Rhine Graben, the adjacent Mainz and Hanau basins (modified after Lahner and Toloczyki, 2004; W: Wiesbaden, M: Mainz, F: Frankfurt, H: Heidelberg). **(b)** Thickness of the Quaternary sand and gravel deposits south of Frankfurt (after Anderle, 1968; Bartz, 1974; Anderle and Golwer, 1980). Location of the model domain, the water works, and of transect A-B.

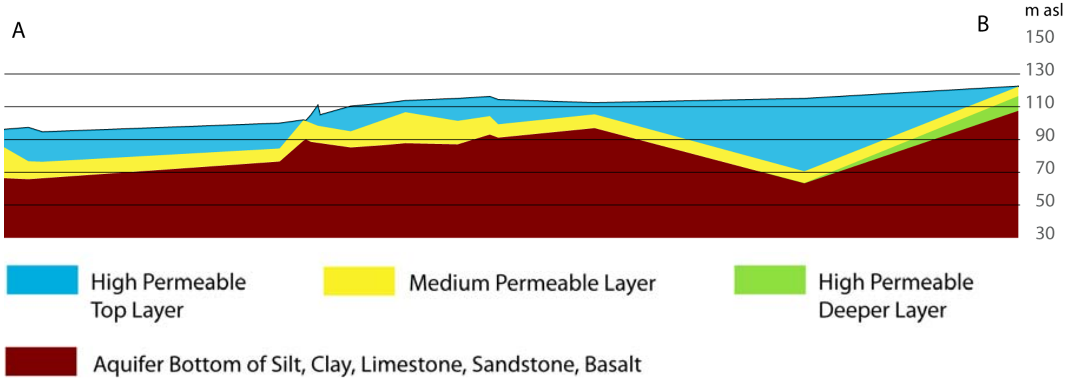

Fig. 2. Averaged hydrostratigraphic layer from nine lithologic units along transect A-B.

HESSD
9, 9687–9714, 2012

Groundwater model ranking with the Akaike Information Criterion

I. Engelhardt et al.

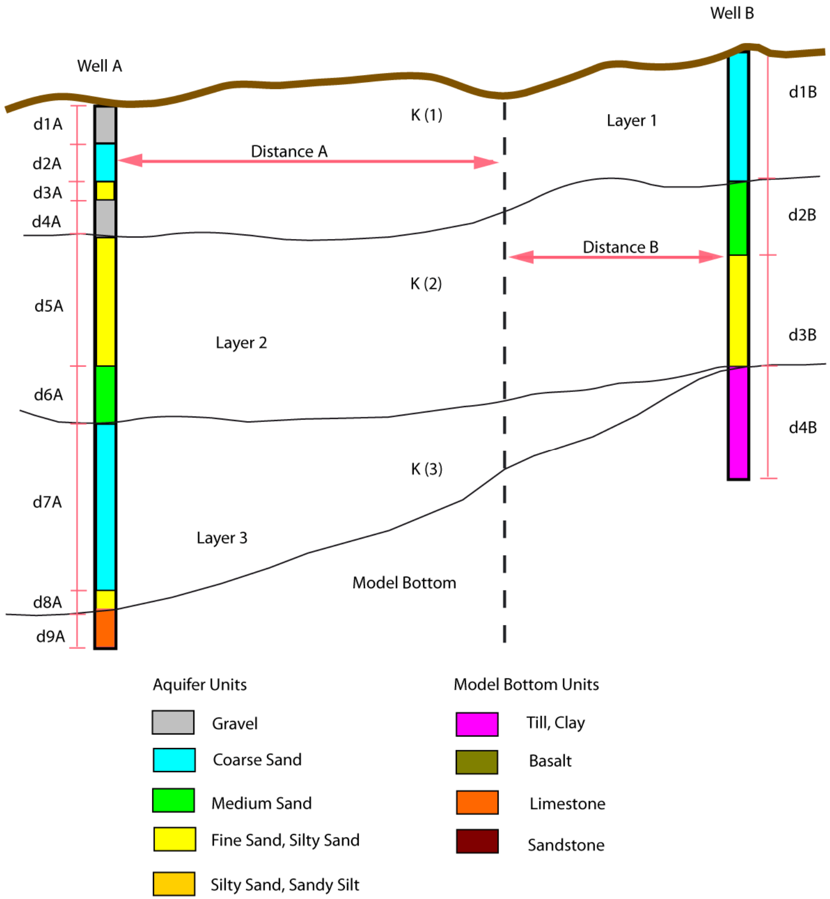

Fig. 3. Averaging technique to derive the equivalent hydraulic conductivities around two wells within the three hydrostratigraphic layer that contain nine lithologic units.

HESSD
9, 9687–9714, 2012

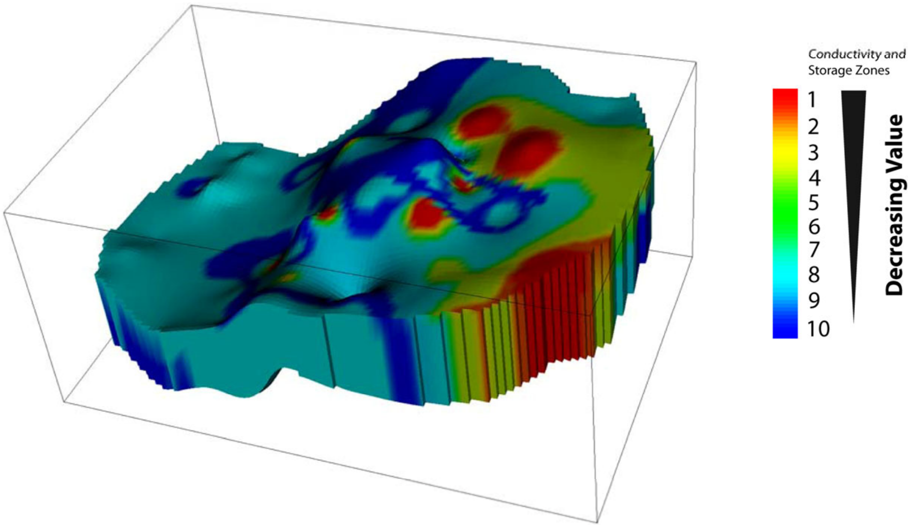

Fig. 4. Spatial distribution of the ten equivalent hydraulic conductivities of Model 1 (uncalibrated model based on sedimentological information) within the three hydrostratigraphic layer.

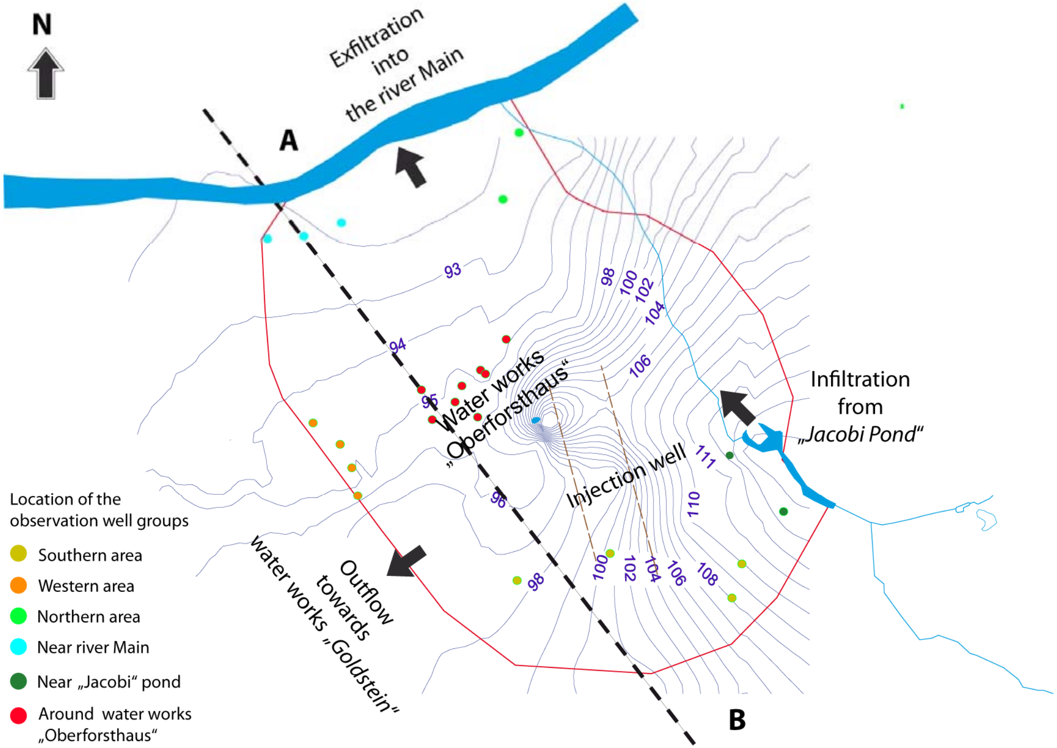

Fig. 5. Boundary conditions, initial head distribution of the numerical flow model and location of the observation well groups.

HESSD
9, 9687–9714, 2012

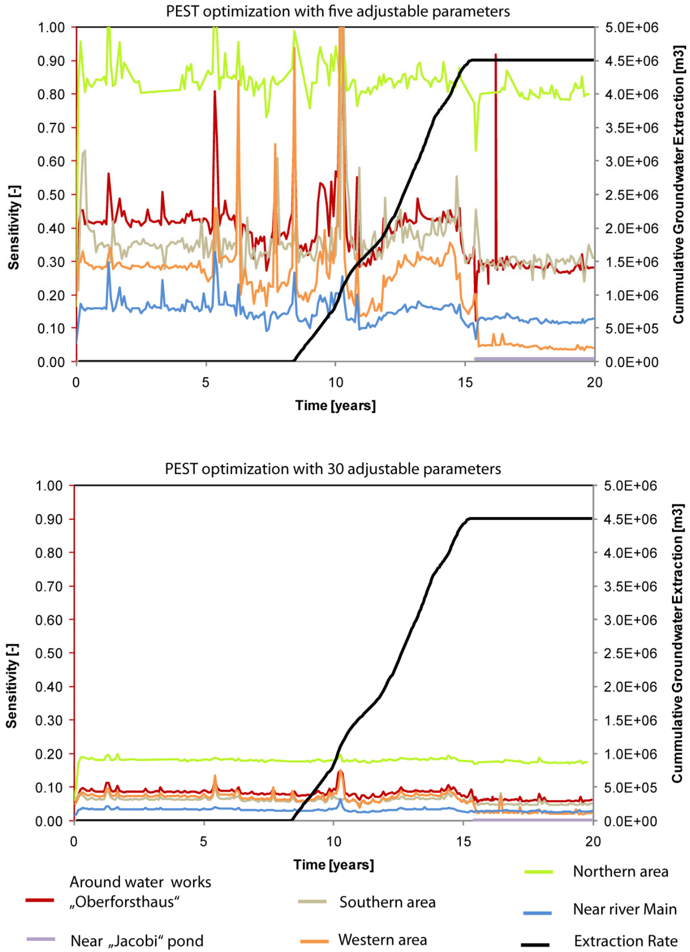

Fig. 6. Sensitivity of the six observation groups with respect to the adjustable amount of parameters and the cumulative groundwater extraction at the water works Oberforsthaus.

HESSD
9, 9687–9714, 2012

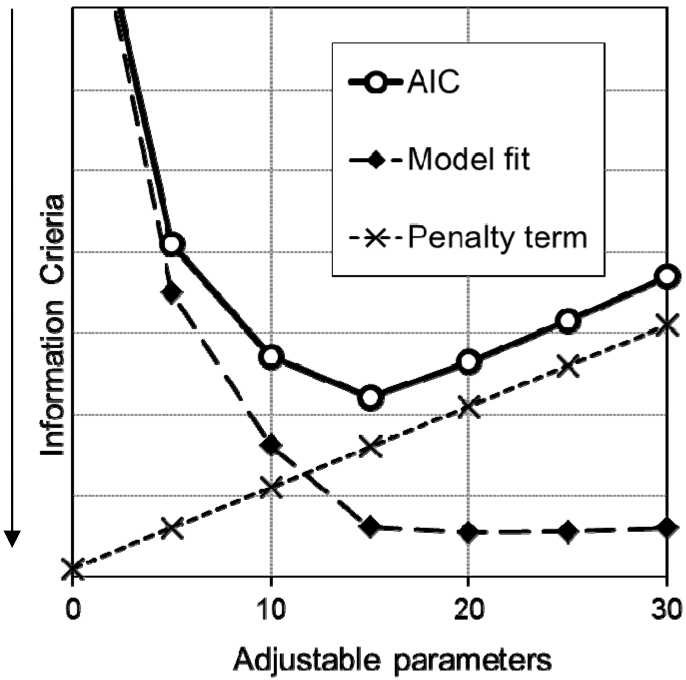

Fig. 7. AIC assessment (circle) of the calibrated models with respect to complexity (cross) and model fit (diamond).

HESSD
9, 9687–9714, 2012

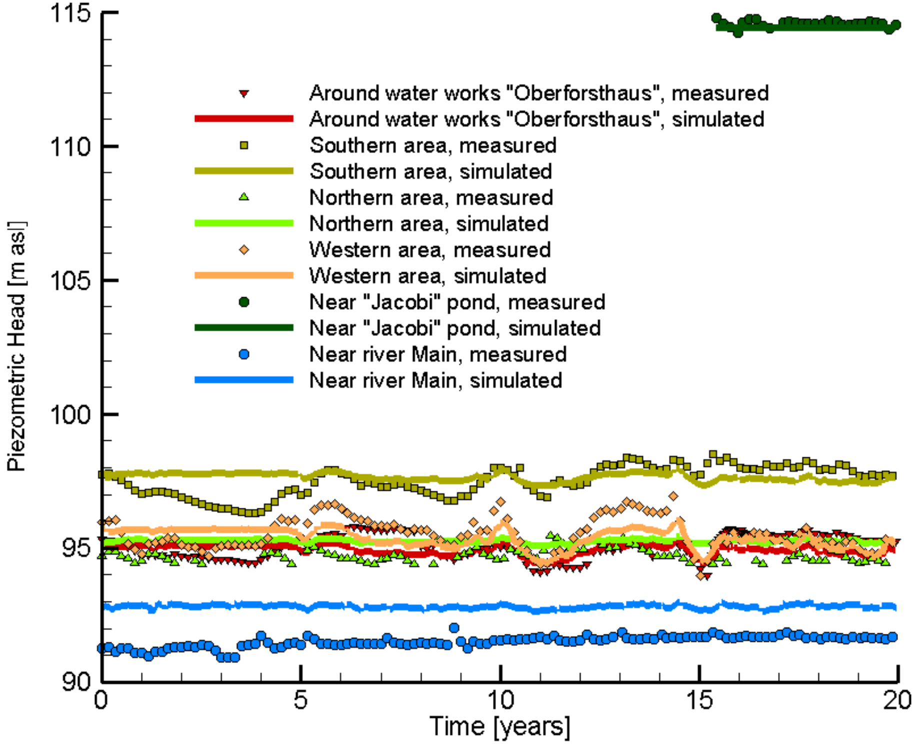

Fig. 8. Simulated piezometric heads of Model 4 (optimal model) versus measured piezometric heads between 1990 and 2009. Observation wells were summarized in six groups. One observation well of each group is illustrated within the figure.