Calibration and Reliability in Groundwater Modelling: From Uncertainty to Decision Making
(Proceedings of ModelCARE'2005, The Hague, The Netherlands, June 2005). IAHS Publ. 304, 2006.

Identification of hydrogeological models: application to tracer test analysis in a karst aquifer

CAROLINA MASSMANN[1], STEFFEN BIRK[1,3], RUDOLF LIEDL[1,4] & TOBIAS GEYER[2]

1 *Center for Applied Geoscience, University of Tübingen, Sigwartstrasse 10, D-72076 Tübingen, Germany*

2 *Geoscientific Centre at the University of Göttingen, Goldschmidtstrasse 3, D-37077 Göttingen, Germany*

3 Now at: *Institute for Earth Sciences, University of Graz, Heinrichstrasse 26, A-8010 Graz, Austria*

4 Now at: *Institute for Groundwater Management, Dresden University of Technology, D-01062 Dresden, Germany*
rudolf.liedl@mailbox.tu-dresden.de

Abstract Several one-dimensional models were fitted to a tracer breakthrough curve measured at a karst spring. The models were compared and ranked using an information theory approach that considers the principle of parsimony. The tracer breakthrough could be most adequately reproduced by including a fraction of mobile water and some nonequilibrium process. It was found that the number of observations clearly affect the model ranking.

Keywords AIC; information theory; model calibration; model selection; tracer test

INTRODUCTION

The selection of a model by calibration is subject to a quite common dilemma: should a "simple" model be used, which involves only a few parameters, thus facilitating calibration but also reducing its ability to reproduce experimental data as important features may have been missed? Or is it preferable to apply a "complex" model with a large number of parameters to better reproduce the available data but decrease the information content carried by each of the model parameters? Methods relying on information theory approaches, that involve statistical measures based on the maximum likelihood, constitute one option for addressing this question. These methods allow one to find a compromise between a good fit and a small number of parameters (e.g. Burnham & Anderson, 2002; Poeter & Anderson, 2004).

This paper presents an application of an information theory approach to different one-dimensional models fitted to a fluorescein (Uranin) breakthrough curve (Massmann, 2004) obtained from a karst aquifer in southwest Germany (Birk *et al.*, 2005). This approach is used to rank the models and the impact of different numbers of observations on the ranking is studied.

THEORETICAL BACKGROUND

Model equations

Some models considered in this paper include physical nonequilibrium processes, which account for partitioning of the solute into a mobile- and an immobile-fluid

region, with solute transport restricted to the mobile phase (Field & Pinsky, 2000). This is described by two dimensionless parameters: a solute partitioning coefficient β between the mobile and immobile phases, and a mass transfer coefficient ω between both phases.

Equation (1) describes solute transport in the mobile phase, and equation (2) represents processes in the immobile phase, which explains the absence of the advection and dispersion terms. It can be seen that the solute exchange rate between the immobile and mobile phases is proportional to their difference in concentration. The model equations read:

$$\beta R\frac{\partial C_m}{\partial t} = \theta_m D\frac{\partial^2 C_m}{\partial x^2} - \theta_m v\frac{\partial C_m}{\partial x} - \frac{\omega v}{L}\left(C_m - C_{im}\right) - \beta R\mu_m C_m \tag{1}$$

$$(1-\beta)R\frac{\partial C_{im}}{\partial t} = \frac{\omega v}{L}\left(C_m - C_{im}\right) - (1-\beta)R\mu_{im}C_{im} \tag{2}$$

where θ is the volumetric water content [–], C the solute concentration [M L^{-3}], L the length of the flow path [L], D the hydrodynamic dispersion coefficient [L^2 T^{-1}], and v the linear velocity [L T^{-1}]. μ represents a decay coefficient [T^{-1}]. The subscripts m and im stand for mobile and immobile, respectively. It is important to note that β is related to the fraction f of adsorption sites that equilibrate with mobile water [–] and the retardation factor R by $\beta R = \theta_m + (R - 1)f$. The dimensionless retardation factor R equals $1 + 2K_a/r$ with r the conduit radius [L] and K_a a partitioning coefficient [L] quantifying the sorbed tracer mass per unit surface area of solid phase and unit solute concentration in the water.

Akaike's Information Criterion

Akaike's Information Criterion, AIC, allows 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. When AIC is used for model selection, it must be computed for each model (all fitted to the same data set) and the model with the smallest AIC is regarded as being the "best" of the tested models. The AIC is calculated as $AIC = -2\ln(L(\hat{\theta} \mid x)) + 2K$ (Burnham & Anderson, 2002) where K equals the number of estimated model parameters plus one (because the variance of residuals has to be considered). $L(\hat{\theta} \mid x)$ represents the likelihood of the estimated parameter $\hat{\theta}$ given the data x. For normally distributed residuals AIC equals:

$$AIC = n\ln(\hat{\sigma}^2) - n\ln(2\pi) - n + 2K \tag{3}$$

where $\hat{\sigma}^2$ represents an estimate of the variance of residuals and n the number of observations. The first term in equation (3) represents the lack of fit of the model, which decreases when more parameters are included. The last term can be seen as a "penalty" for incorporating more parameters, since it then gets larger. The two middle terms are constants for a specific data set, and are not affected if parameters are added or removed from the models.

Akaike (1978) defined weights w_i to obtain a relative measure of the likelihood of a model, given a set of N models. These weights are expressed as:

$$w_i = \exp(-0.5\Delta_i) / \sum_{n=1}^{N} \exp(-0.5\Delta_n)$$

where $\Delta_i = AIC_i - AIC_{min}$ is termed the AIC difference with AIC_i and AIC_{min} denoting the AIC of model i and the smallest AIC of all models considered. The larger the AIC difference of the model, the less plausible it is that this fitted model is the best one.

RESULTS AND DISCUSSION

Thirteen models were fitted to the breakthrough curve using the program CXTFIT 2.1 (Toride *et al.*, 1999). Model 1 only included advection and dispersion. Models 2–13 also account for specific combinations of other processes covered by equations (1) and (2). Residuals were assumed to be normally distributed and the number of observations was 2104. The parameters fitted in each model can be seen in Table 1.

Table 1 AIC and related information for the fitted models. v = linear advective velocity; D = dispersion coefficient; R = retardation factor; β = solute partitioning coefficient between the mobile and immobile phases; ω = mass transfer coefficient between both phases; μ = decay coefficient. Subscripts *m*, *im* *m+im* stand for mobile, immobile phase and a combined reaction for both phases, respectively.

Model	Fitted parameters						*RSS*/*n*	*K*	$-2\ln(L(\hat{\theta} \mid x))$	AIC	AIC Δ_i	AIC w_i
12	*v* *D*		β	ω	μ_m	μ_{im}		0.3617	7	3831.0	3845.0	0.0	0.284
6	*v* *D*		β	ω		μ_{im}		0.3620	6	3833.0	3845.0	0.0	0.284
11	*v* *D*	*R*	β	ω		μ_{im}		0.3620	7	3833.1	3847.1	2.0	0.102
13	*v* *D*	*R*	β	ω	μ_m	μ_{im}		0.3617	8	3831.2	3847.2	2.2	0.097
8	*v* *D*		β	ω			μ_{m+im}	0.3624	6	3835.6	3847.6	2.6	0.078
9	*v* *D*	*R*	β	ω			μ_{m+im}	0.3622	7	3834.1	3848.1	3.1	0.060
10	*v* *D*	*R*	β	ω	μ_m			0.3622	7	3834.4	3848.4	3.4	0.051
7	*v* *D*		β	ω	μ_m			0.3627	6	3836.8	3848.8	3.8	0.043
4	*v* *D*		β	ω				2.8818	5	8197.8	8207.8	4363	0.000
3	*v* *D*				μ_m			11.2439	4	11062.2	11070.2	7225	0.000
5	*v* *D*	*R*			μ_m			11.2438	5	11062.2	11072.2	7227	0.000
1	*v* *D*							20.7321	3	12349.6	12355.6	8511	0.000
2	*v* *D*	*R*						20.7321	4	12349.6	12357.6	8513	0.000

The impact of mass transfer, retardation, and decay

Figures 1 and 2 show the fitted and the measured curves. The fitted models can be separated into two groups: models 1 to 5, with fitted curves that can be clearly distinguished from the measured one; and models 6 to 13, which all look very similar and fit the measured data better than the models in the first group. Figure 1 reveals that model 4 (physical nonequilibrium) is able to produce a better fit than models that do not consider partitioning of the solute into a mobile and an immobile phase (models 1, 2, 3 and 5), especially regarding the tailing.

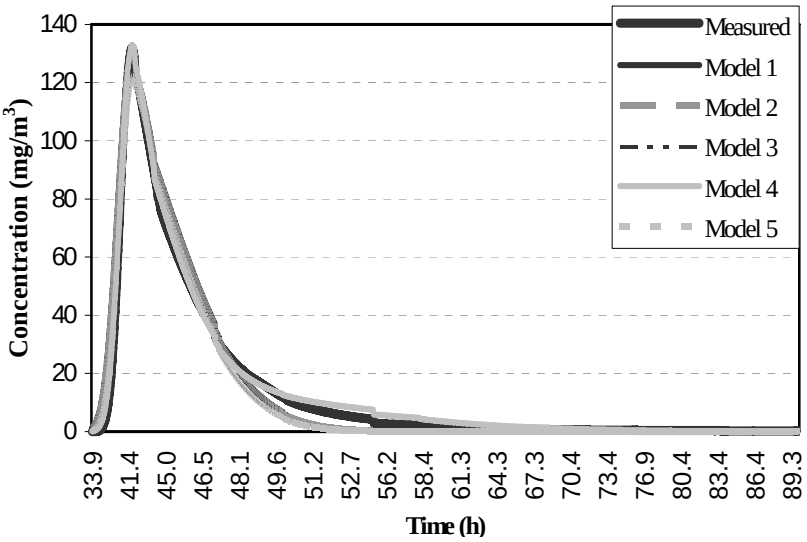

Fig. 1 Measured and fitted concentrations for models 1 to 5 (Table 1).

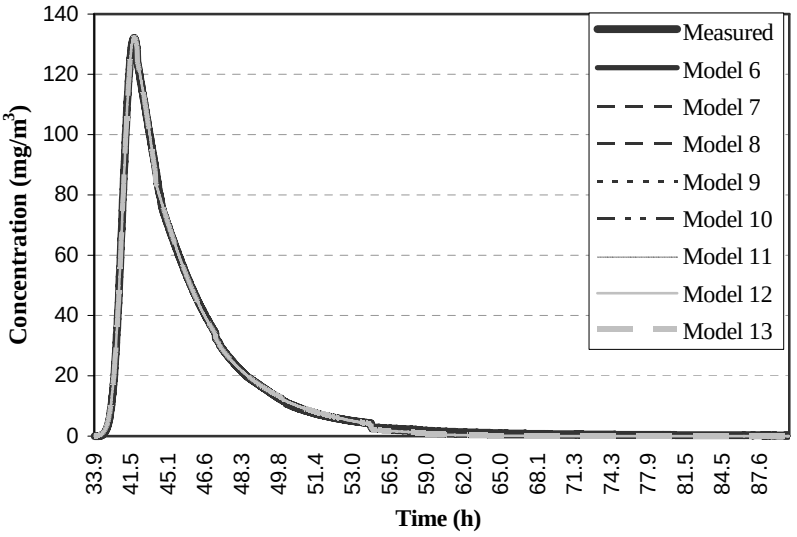

Fig. 2 Measured and fitted concentrations for models 6 to 13 (Table 1).

Models 1 to 5 have much higher AICs than the remaining models (Table 1). Models 12 and 6 have the lowest AIC, and can, therefore, be considered the best models. These two models account together for 56% of the Akaike weights w_i, i.e. there is a 56% probability that one of these models is the best of the 13 models. Models 1 to 5 have no support (Akaike weight = 0), implying that the remaining models that include physical nonequilibrium and decay reactions provide better approximations to the measured data.

The term describing the fitting is the most important AIC component. The term $2K$ has, however, an effect on the relative AIC values (and therefore on the model weights). The ranking of the models 11 and 13, for instance, is different if only the quality of fitting is considered; for example, the residual sum of squares divided by the number of observations, RSS/n, or the first term of equation (3).

Varying the number of observations

To analyse the effects of data sets with different numbers of observations on the model ranking, sixteen additional data sets were constructed by deleting observations at regular intervals. The models were then fitted to the reduced data sets. When all 2104 observations are considered, model 12 (six fitted parameters) is the best (Table 2). If the number of observations is reduced, model 6 (five fitted parameters) moves up to the first place. Model 4 (four fitted parameters) is the best when 22 or 18 observations are available.

Model 13 (the model with the highest number of parameters) takes place four in the set with 2104 observations, and successively a lower place as fewer observations are considered until reaching the last place when only 11 observations are available. More generally, models with six fitted parameters tend to be in higher positions of the ranking, the more observations are considered. With 421 observations or fewer, no model with six or more parameters occupies any of the first three places.

Focusing on each model separately, there is not much difference in the fitted parameters as long as the number of observations is larger than 53. For instance, the dispersivity (α) for model 9 is very similar when at least 141 observations are considered (Table 3). The differences increase for 106, 85 and 71 observations, but dispersivities are still close to those measured with more observations. With 53 or fewer observations, however, differences become considerably larger.

Table 2 Ranking of the models according to *AICc* when data sets with different number of observations are considered (*AICc* includes a correction term recommended in Burnham & Anderson (2002) when the ratio *n*/*K* is smaller than 40).

Number of observations	2104	1578	1403	1052	421	211	71	43	22	18	11
Best model	12	6	6	6	6	6	6	6	4	4	6
	6	12	12	12	8	7	8	8	7	7	8
	11	11	7	8	7	8	7	7	8	8	7
	13	13	8	7	9	12	12	12	6	6	4
	8	9	9	11	12	11	11	11	9	10	12
	9	8	13	9	11	9	9	10	10	9	11
	10	7	11	13	13	10	10	9	11	11	10
	7	10	10	10	10	13	13	13	12	12	9
	4	4	4	4	4	4	4	4	13	13	3
	3	3	3	3	3	3	3	3	2	2	5
	5	5	5	5	5	5	5	5	3	3	1
	1	1	1	1	1	1	1	1	1	1	2
Worst model	2	2	2	2	2	2	2	2	5	5	13

Table 3 Dispersivities (mobile phase) for model 9 fitted to data sets with different number of observations.

No. obs	α	No. obs	α	No. obs	α	No. obs	α
2104	3.88	421	3.87	85	3.85	27	5.31
1893	3.88	211	3.87	71	3.92	22	5.19
1578	3.88	141	3.89	53	3.70	18	5.27
1403	3.88	106	3.84	43	4.00	11	5.62
1052	3.88						

CONCLUSIONS

There is a difference in the model ranking when AIC is used instead of the residual sum of squares. This difference is important when the quality of the fitting is similar among the models.

There is no advantage in calculating AIC when the differences between the fitted curves are important enough, which is expected when there is a large number of available observations. It can also occur that there is no difference in the model ranking (when comparing the results obtained with AIC and least squares) if the fitted models have a similar number of parameters.

Data sets with less observations support simpler models (with fewer fitting parameters). On the contrary, models with more parameters are preferred when larger data sets are available.

Future work could involve non-Gaussian statistical distributions of the residuals. These distributions, of course, have to allow for positive as well as negative values. The ranking of the models may be altered if different distributions of residuals are assumed. However, the application of a statistical test is recommended beforehand in order to check whether the assumption of a specific residual distribution is adequate.

REFERENCES

Akaike, H. (1978) A Bayesian analysis of the minimum AIC procedure. *Ann. Inst. Statist. Math.* **30**, 9–14.

Birk, S., Geyer, T., Liedl, R. & Sauter, M. (2005) Process-based interpretation of tracer tests in carbonate aquifers. *Ground Water* **43**(3), 381–388.

Burnham, K. P. & Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer, New York, USA.

Field, M. S. & Pinsky, P. F. (2000) A two-region nonequilibrium model for solute transport in solution conduits in karstic aquifers. *J. Contam. Hydrol.* **44**, 329–351.

Massmann, C. (2004) Simplicity vs. complexity in tracer test modelling. MSc Thesis, University of Tuebingen, Center for Applied Geoscience, Germany.

Poeter, E. P. & Anderson, D. R. (2004) Multi-model ranking and inference in groundwater modeling. In: *Finite Element Models, MODFLOW, and More* (ed. by K. Kovar, Z. Hrkal & J. Bruthans) (Proc. Int. Conf. FEM_MODFLOW, September 2004, Karlovy Vary, Czech Republic), 85–89.

Toride, N., Leij, F. J. & van Genuchten, M. Th. (1999) The CXTFIT code for estimating transport parameters from laboratory or field tracer experiments – version 2.1. *US Salinity Laboratory Agricultural Research Service, US Department of Agriculture, Research report 137.*