Hydrol. Earth Syst. Sci. Discuss., 9, C4862–C4867, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C4862/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Complexity versus simplicity: an example of groundwater model ranking with the Akaike Information Criterion" *by* I. Engelhardt et al.

Anonymous Referee #1

Received and published: 16 October 2012

General Comments

This paper applies the Akaike Information Criterion (AIC) to rank alternative conceptual models of transient groundwater flow in graben-filling Quaternary sediments near Frankfurt. The alternative models differ in terms of how many hydraulic conductivity and storage parameters are independently estimated during model calibration using the inverse modeling software PEST. Calibration observations included 5081 hydraulic heads over a 20 year period.

AIC is one of several criteria that have been developed for ranking alternative models.

These criteria favor models with a better fit to the calibration observations, and at the same time penalize models with a larger number of parameters. They tend to highly rank models that have enough complexity to fit the observations well, yet do not over-parameterize the representation of model attributes such as system properties and boundary conditions. In addition to model ranking, these criteria have been used in the past decade or so for model averaging, in which they are the basis for calculating the likelihood or probability of each model under consideration.

This paper is a case study. It does not include development of new methods or concepts. It is a straightforward application of AIC to the alternative models of the Quaternary aquifer near Frankfurt. The paper first describes the field site, the flow model development, and the calibration approach. Then, the AIC criterion and calculation of AIC weights are presented and alternative models 1 through 7 are described, which increase in complexity with model number. Lastly, the results are presented, including sensitivity analyses, AIC model ranks, residuals, and parameter estimates. The results show that model number 4 is the optimal model according to AIC; of the seven models, model 4 is exactly in the middle in terms of the degree of complexity.

There are several recent papers in the literature that apply the AIC criterion to models of field groundwater systems (e.g. Folia et al. 2007, Kazumba et al. 2008, Parker et al. 2010, Singh et al. 2010, Ye et al. 2010). Most of these papers apply many different criteria for multi-model comparison, including BIC, AICc, KIC, and others, and evaluate which performs the best in terms of selecting a preferable model. Given the content of these previous papers, I do not believe that the paper under review is a significant original contribution to the hydrologic literature that will be of great interest. The previous papers explore the application of AIC more thoroughly in terms of comparing it to several other model ranking criteria.

Specific Comments

p. 9689. lines 17-20. The statement beginning "It identifies…" seems to state that calculating the AIC for a single model enables one to assess whether that model is satisfactory or, in contrast, needs more complexity introduced. The AIC for a single model cannot do this. The AIC is useful only when it can be calculated for a set of calibrated models with varying levels of complexity represented by varying numbers of model parameters.

p. 9689, lines 21-23. AIC has been applied in many more groundwater modeling studies than those listed here. The General Comment lists some of these studies, additional studies can be found by searching for AIC on the web sites for Water Resources Research, Ground Water, and the Journal of Hydrology, for example.

p. 9691, line 19: It is not clear what "Mio" means.

p. 9694, line 6: Regarding the statement that leakage was adjusted manually, do you mean that the hydraulic conductivity or conductance of the river bed and of Jacobi Pond bottom sediments were adjusted manually?

p. 9694, Section 2.2.4. Please state the total number of observations. It is given later (on p. 9699 line 5), but needs to be stated here. There is an alternative AIC measure, called AICc, that is recommended to be used when $(n/p)<40$ (Poeter and Anderson 2005). For your most complex model, $(n/p)=5081/30=169$. Stating the value of n will make it clear why you use AIC instead of AICc. Also, please briefly state how the observations were weighted.

p. 9695, lines 7-9: The concept of whether a model can be validated is a controversial subject among groundwater modelers. I suggest that the authors consider the article by Bredehoeft and Konikow (2012) before using the term validation to describe the exercise of assessing model fit to data excluded from the calibration.

p. 9695, equation (2): The commonly used expression for AIC applied to models calibrated by least squares regression omits the middle two terms of equation 2 (e.g., Poeter and Hill 2007, Singh et al. 2010). Burnham and Anderson (2002, p. 63), which

is cited for equation 2, also present the equation for AIC without the middle two terms for the least squares case. Please explain why you use the form of AIC in equation 2, and please cite a reference for this form of AIC.

p. 9695, equation (3). This expression needs to include the observation weights. Using the definition of weights in PEST, the squared term in the summation needs to be the product of the weight and the residual, not just the residual.

p. 9697, Section 3.1., General: Please provide the equation for the sensitivity coefficients plotted in Fig. 6. The measure plotted appears to be an overall measure of the sensitivity of all observations in a group to all parameters of the model. In addition, the measure appears to be normalized to the largest sensitivity in the model with the fewest parameters. For readers to evaluate the results in Fig. 6, the equation is needed.

p. 9697, line 11: What is the "colmation" layer? Does this refer to the river and pond bed sediments?

p. 9699, Section 3.3.1 and Table 3: Weighted residuals are dimensionless quantities, so the standard error of the weighted residuals is dimensionless rather than having units of meters as indicated in the manuscript.

p. 9699, line 26: The model validation data set is mentioned on lines 26-27, and is followed by the general statement about model fit on lines 27-29. That statement refers to Table 3, which summarizes the fit for all observations, those included in the calibration and those used just for validation. If you retain the idea of model validation in the paper, it would be best to show the fit to the calibration and validation observations separately.

p. 9700, lines 13-15: The Quaternary aquifer was described as consisting of unconsolidated sediments (section 2.1.1, Fig. 3). The explanation of possible highly permeable fractures in the aquifer is quite inconsistent with this aquifer geology.

Editorial Comments

Generally, the paper is written fairly well, and is well organized. Some parts would benefit from additional careful editing to improve the English usage. Some examples are:

p. 9691, line 18, change "rebuild" to "rebuilt"

p. 9693, line 16 "in the" not "in of the"

p. 9693, line 24 change to "The main inflow into the groundwater system resulted in…"

p. 9694, line 5 change "prevailed" to "were"

p. 9695, line 9 change "observations" to "observation"

p. 9696, lines 4-5 rephrase "since it gets larger then"

p. 9696, line 9 change "is denoting" to "denotes"

p. 9696, line 22 omit "amount"

p. 9697, line 18 change "amount" to "number"

p. 9697, line 20 change to "the highest …" Change "amount" to number"

p. 9697, line 22 Change "amount" to number"

p. 9698, line 2 change to "allowed evaluation of the conceptual…"

p. 9898, line 15 change "as similar worse" to "similarly poor"

p. 9899, line 5 omit "sum"

p. 9899, lines 12, change to "and also displayed the impact of…"

p. 9700, line 7, change to "Very limited information was available…"

REFERENCES CITED IN REVIEW (this list only includes references that are not cited in the manuscript itself)

Bredehoeft, J. D. and Konikow, L. F. (2012), Ground-Water Models: Validate or Invalidate. Ground Water, 50: 493–495. doi: 10.1111/j.1745-6584.2012.00951.x

Foglia, L., Mehl, S.W., Hill, M.C., Perona, P. and Burlando, P. (2007), 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

Kazumba, Shija, Gideon Oron, Yusuke Honjo, Kohji Kamiya, Lumped model for regional groundwater flow analysis, Journal of Hydrology, Volume 359, Issues 1–2, 15 September 2008, Pages 131-140, ISSN 0022-1694, 10.1016/j.jhydrol.2008.06.021.

Parker, A. H., West, L. J., Odling, N. E. and Bown, R. T. (2010), A Forward Modeling Approach for Interpreting Impeller Flow Logs. Ground Water, 48: 79–91. doi: 10.1111/j.1745-6584.2009.00600.x

Singh, A., Mishra, S. and Ruskauff, G. (2010), Model Averaging Techniques for Quantifying Conceptual Model Uncertainty. Ground Water, 48: 701–715. doi: 10.1111/j.1745-6584.2009.00642.x

Ye, M., Pohlmann, K. F., Chapman, J. B., Pohll, G. M. and Reeves, D. M. (2010), A Model-Averaging Method for Assessing Groundwater Conceptual Model Uncertainty. Ground Water, 48: 716–728. doi: 10.1111/j.1745-6584.2009.00633.x

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 9687, 2012.