Geophysical Research Abstracts
Vol. 12, EGU2010-12375, 2010
EGU General Assembly 2010

A flexible Bayesian assessment for the expected impact of data on prediction confidence for optimal sampling designs

Philipp Leube (1), Andreas Geiges (2), and Wolfgang Nowak (3)

(1) Department of Hydrology Engineering, University of Stuttgart, Pfaffenwaldring 7a, 70176 Stuttgart, Germany, philipp.leube@iws.uni-stuttgart.de, (2) Department of Hydrology Engineering, University of Stuttgart, Pfaffenwaldring 7a, 70176 Stuttgart, Germany, andreas.geiges@iws.uni-stuttgart.de, (3) Department of Hydrology Engineering, University of Stuttgart, Pfaffenwaldring 7a, 70176 Stuttgart, Germany, wolfgang.nowak@iws.uni-stuttgart.de

Incorporating hydrogeological data, such as head and tracer data, into stochastic models of subsurface flow and transport helps to reduce prediction uncertainty. Considering limited financial resources available for the data acquisition campaign, information needs towards the prediction goal should be satisfied in a efficient and task-specific manner. For finding the best one among a set of design candidates, an objective function is commonly evaluated, which measures the expected impact of data on prediction confidence, prior to their collection. An appropriate approach to this task should be stochastically rigorous, master non-linear dependencies between data, parameters and model predictions, and allow for a wide variety of different data types. Existing methods fail to fulfill all these requirements simultaneously.

For this reason, we introduce a new method, denoted as CLUE (*Cross-bred Likelihood Uncertainty Estimator*), that derives the essential distributions and measures of data utility within a generalized, flexible and accurate framework. The method makes use of Bayesian GLUE (*Generalized Likelihood Uncertainty Estimator*) and extends it to an optimal design method by marginalizing over the yet unknown data values. Operating in a purely Bayesian Monte-Carlo framework, CLUE is a strictly formal information processing scheme free of linearizations. It provides full flexibility associated with the type of measurements (linear, non-linear, direct, indirect) and accounts for almost arbitrary sources of uncertainty (e.g. heterogeneity, geostatistical assumptions, boundary conditions, model concepts) via stochastic simulation and Bayesian model averaging. This helps to minimize the strength and impact of possible subjective prior assumptions, that would be hard to defend prior to data collection. Our study focuses on evaluating two different uncertainty measures: (i) expected conditional variance and (ii) expected relative entropy of a given prediction goal.

The applicability and advantages are shown in a synthetic example. Therefor, we consider a contaminant source, posing a threat on a drinking water well in an aquifer. Furthermore, we assume uncertainty in geostatistical parameters, boundary conditions and hydraulic gradient. The two mentioned measures evaluate the sensitivity of (1) general prediction confidence and (2) exceedance probability of a legal regulatory threshold value on sampling locations.