Geophysical Research Abstracts
Vol. 19, EGU2017-13390-1, 2017
EGU General Assembly 2017

Bayesian model evidence as a model evaluation metric

Anneli Guthke (1,2), Marvin Höge (1,2), and Wolfgang Nowak (1)

(1) Institute for Modelling Hydraulic and Environmental Systems (LS^3)/SimTech, University of Stuttgart, Stuttgart, Germany (anneli.guthke@iws.uni-stuttgart.de), (2) Center for Applied Geoscience, University of Tübingen, Germany

When building environmental systems models, we are typically confronted with the questions of how to choose an appropriate model (i.e. which processes to include or neglect) and how to measure its quality. Various metrics have been proposed that shall guide the modeller towards a most robust and realistic representation of the system under study. Criteria for evaluation often address aspects of accuracy (absence of bias) or of precision (absence of unnecessary variance) and need to be combined in a meaningful way in order to address the inherent bias-variance dilemma.

We suggest using Bayesian model evidence (BME) as a model evaluation metric that implicitly performs a tradeoff between bias and variance. BME is typically associated with model weights in the context of Bayesian model averaging (BMA). However, it can also be seen as a model evaluation metric in a single-model context or in model comparison. It combines a measure for goodness of fit with a penalty for unjustifiable complexity. Unjustifiable refers to the fact that the appropriate level of model complexity is limited by the amount of information available for calibration. Derived in a Bayesian context, BME naturally accounts for measurement errors in the calibration data as well as for input and parameter uncertainty. BME is therefore perfectly suitable to assess model quality under uncertainty.

We will explain in detail and with schematic illustrations what BME measures, i.e. how complexity is defined in the Bayesian setting and how this complexity is balanced with goodness of fit. We will further discuss how BME compares to other model evaluation metrics that address accuracy and precision such as the predictive logscore or other model selection criteria such as the AIC, BIC or KIC.

Although computationally more expensive than other metrics or criteria, BME represents an appealing alternative because it provides a global measure of model quality. Even if not applicable to each and every case, we aim at stimulating discussion about how to judge the quality of hydrological models in the presence of uncertainty in general by dissecting the mechanism behind BME.