Geophysical Research Abstracts
Vol. 14, EGU2012-12354-2, 2012
EGU General Assembly 2012

Analysing goodness of fit measures using a sensitivity based approach

C. Massmann and H. Holzmann

Institute of Water Management, Hydrology and Hydraulic Engineering University of Natural Resources and Life Sciences, Vienna, Austria

While the resources spend calibrating and evaluating models are often substantial, little thought is given to the selection of the evaluation criteria. The work presented here attempts to increase the knowledge about different criteria allowing a better justification of their use as well as aiding in the selection of complementary evaluation measures in multi-criteria approaches. This is done through variance decomposition based sensitivity analyses using the eFAST (extended Fourier amplitude sensitivity test) method. This method provides information about the proportion of the variance explained by each parameter individually (first order indices) and also allows calculating the impact of the interactions between parameters on the total variance. We present a comparison of the variances, first order indices and interactions for four criteria: the mean error (ME), mean absolute error (MAE), mean squared error (MSE) and the mean squared logarithmic error (MSLE). The calculations were done using a 3 and a 15-day moving window for a period of two years.

With respect to the variances it was found that MSE had the largest variance range, with extremes between one and two orders of magnitude higher than the other measures. The variance of ME is similar to the variance of MAE, and MSLE has the highest variance of all criteria during low flow periods. When looking at the whole period, ME has the highest first order indices and the smallest interactions, both characteristics facilitating model calibration. The first order indices are, for instance, around 18 % higher than for MSE and MSLE and the interactions are only 58 % of the interactions observed with MSE and MSLE.

On a daily basis, the sensitivities for the three day moving window look similar for all criteria, while showing larger differences for the 15 day moving window. The most striking features are the sharp boundaries observed for MSE, where it is clearly seen that only a few points determine the sensitivities for larger periods. MSLE has in contrast much smoother curves, with smaller indices than MSE for parameters related to fast flow components. Since it is observed for some periods that the parameters dominating are different if MSE or MSLE are considered, it is expected that a multi-criteria calibration taking both measures into account could make a better use of the available information in the time series. The sensitivities of ME and MAE are between the sensitivities of the other two criteria: sometimes more similar to one, on other times more similar to the other. The consideration of these measures could, however, still be beneficial due to their higher first order sensitivities.