Geophysical Research Abstracts
Vol. 17, EGU2015-1368, 2015
EGU General Assembly 2015

Distributed multi-criteria model evaluation and spatial association analysis

Laura Scherer and Stephan Pfister
Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland (scherer@ifu.baug.ethz.ch)

Model performance, if evaluated, is often communicated by a single indicator and at an aggregated level; however, it does not embrace the trade-offs between different indicators and the inherent spatial heterogeneity of model efficiency.

In this study, we simulated the water balance of the Mississippi watershed using the Soil and Water Assessment Tool (SWAT). The model was calibrated against monthly river discharge at 131 measurement stations. Its time series were bisected to allow for subsequent validation at the same gauges. Furthermore, the model was validated against evapotranspiration which was available as a continuous raster based on remote sensing. The model performance was evaluated for each of the 451 sub-watersheds using four different criteria: 1) Nash-Sutcliffe efficiency (NSE), 2) percent bias (PBIAS), 3) root mean square error (RMSE) normalized to standard deviation (RSR), as well as 4) a combined indicator of the squared correlation coefficient and the linear regression slope (bR2).

Conditions that might lead to a poor model performance include aridity, a very flat and steep relief, snowfall and dams, as indicated by previous research. In an attempt to explain spatial differences in model efficiency, the goodness of the model was spatially compared to these four phenomena by means of a bivariate spatial association measure which combines Pearson's correlation coefficient and Moran's index for spatial autocorrelation.

In order to assess the model performance of the Mississippi watershed as a whole, three different averages of the sub-watershed results were computed by 1) applying equal weights, 2) weighting by the mean observed river discharge, 3) weighting by the upstream catchment area and the square root of the time series length.

Ratings of model performance differed significantly in space and according to efficiency criterion. The model performed much better in the humid Eastern region than in the arid Western region which was confirmed by the high spatial association with the aridity index (ratio of mean annual precipitation to mean annual potential evapotranspiration). This association was still significant when controlling for slopes which manifested the second highest spatial association. In line with these findings, overall model efficiency of the entire Mississippi watershed appeared better when weighted with mean observed river discharge. Furthermore, the model received the highest rating with regards to PBIAS and was judged worst when considering NSE as the most comprehensive indicator.

No universal performance indicator exists that considers all aspects of a hydrograph. Therefore, sound model evaluation must take into account multiple criteria. Since model efficiency varies in space which is masked by aggregated ratings spatially explicit model goodness should be communicated as standard praxis - at least as a measure of spatial variability of indicators. Furthermore, transparent documentation of the evaluation procedure also with regards to weighting of aggregated model performance is crucial but often lacking in published research. Finally, the high spatial association between model performance and aridity highlights the need to improve modelling schemes for arid conditions as priority over other aspects that might weaken model goodness.