Geophysical Research Abstracts,
Vol. 11, EGU2009-9711-2, 2009
EGU General Assembly 2009

Model Input Data Selection by the Gamma Test

D. Han (1), W. Yan (2), and A. Moghaddamnia (3)

(1) University of Bristol, Civil Engineering, Bristol, United Kingdom, d.han@bristol.ac.uk, (2) GE Research, USA, yan@crd.ge.com, (3) Department of Range and Watershed Management, University of Zabol, Iran, ali.moghaddamnia@gmail.com

Hydrological processes are usually complex and involve nonlinear dynamic systems. A hydrological modeller needs to use trial and error to build mathematical models (such as ANN) for different input combinations. This is very time consuming since the modeller needs to calibrate and test different model structures with all the likely input combinations. In addition, there is no guidance about how many data points should be used in the calibration and what accuracy the best model is able to achieve. In this study, the Gamma Test developed by computer scientists in Cardiff University is explored for its ability in reducing model development workload and providing input data guidance before models are developed (i.e., its result is independent of the models to be developed). Basically, the Gamma Test is able to provide the best mean square error that can possibly be achieved using any nonlinear smooth models. In this study, different combinations of input data were explored to assess their influence on the evaporation estimation modelling. There were meaningful combinations of inputs (n is the number of total potential inputs); from which, the best one can be determined by observing the Gamma value, which indicates a measure of the best estimation attainable using any modelling methods for unseen smooth functions of continuous variables. The case study to validate the Gamma Test is based on the evaporation data in the Sistan region of Iran with 11 years of wind, temperature, saturation vapour pressure deficit, relative humidity and pan evaporation. The nonlinear dynamic model tested is the generalized regression neural network (GRNN), a special type of neural network. The training and testing data are partitioned by random selection from the original data set. Multiple linear regression models are used as a benchmark to check how nonlinear the system is. It has been found that the overall performance of the Gamma Test is quite encouraging and GT demonstrates its huge potential for an efficient ANN model development (as proved by the holdout validation process). The Gamma values are able to provide a good indication about the achievable accuracy for the ANN models and this has a huge advantage over the traditional model selection approaches such as cross validation or AIC which only inform the modeller the relative performance among the candidate models (e.g., if all candidate models are bad, they will pick up the best of the bad lot). But the Gamma Test not only tells the best input combination, but also how good the model's performance is relative to the best achievable result so that it is possible for a modeller to decide if any further effort is worthwhile to further improve the existing model. The study demonstrates that more explorations are needed to gain wider experience about this data selection tool and how it could be used in assessing the validation data.

Keywords: Model Input Selection, Gamma Test, Artificial Neural Networks, Evaporation