Geophysical Research Abstracts,
Vol. 10, EGU2008-A-05387, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-05387
EGU General Assembly 2008

Runoff prediction with partial mutual information as tool for identification and construction of input variables for artificial neural networks

M. Hanel, **P. Maca**
Faculty of Environmental Sciences, Czech University of Agriculture, Prague

Artificial neural networks are very flexible and often used tools for runoff modeling. They can combine any reasonable set of input variables and learn the way how to map it to output variable(s). Among other parameters (in particular parameters that influence speed and quality of learning process) there are two crucial things that have to be chosen: the architecture of the network and the set of input variables. The increasing number of neurons leads to greater flexibility of the network but can cause tendency of overfitting the data, similarly increasing number of inputs brings more information into the model but together with noise. Finally, more complex structure of neural network costs more computational time. The key issue is then to choose the smallest architecture that is able to represent the process and the smallest set of input variables that includes as much necessary information for reliable prediction as possible. In the presented study the main attention was paid to input variable selection. Because neural networks are often chosen because of their ability to represent nonlinear relationships between variables the measure of nonlinear dependence called partial mutual information (PMI) was applied. PMI is based on entropy and express the reduction of uncertainty in one variable due to the knowledge of other variable. Compared to basic mutual information PMI counts for dependence between the variable of interest and already chosen variables as well and thus prevents from including variables that bring no extra information into input data set. Significance of input variable is tested against hypothesis of no dependence by standard bootstrap technique. The method was applied to prediction of runoff in experimental catchment Modrava III in the Czech republic for 1 and 12 hours leading times. Runoff was predicted using

hourly data on precipitation and past runoff. The aim of the study was to identify set of input variables (time lags of runoff and precipitation) that significantly contains information about predicted runoff. It was found that the variables of significance vary a lot between different runoff events. Therefore the mean dependence between PMI and time lag was explored for both precipitation and runoff and both leading times. Resulting relation between time lag and PMI was parametrized and used for construction of new input variables. In this way the information about past is summarized in few new variables instead of including all of these variables into input data set. The results shows that comparing the neural networks with same number of inputs, the network that uses input constructed with PMI produces more stable output. As a part of this study freely available software tools for calculation of PMI (PMI Estimator) and application of neural networks (ANN) were developed.