Geophysical Research Abstracts
Vol. 13, EGU2011-6128, 2011
EGU General Assembly 2011


# Parameter dependent convergence bounds for a class of conceptual hydrologic models

Luis Bastidas (1), Saket Pande (2), Sandjai Bhulai (3), and Mac McKee (1)

(1) Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan, Utah, United States, (2) Department of Water Management, Delft University of Technology, Delft, Netherlands, (3) Department of Mathematics, VU University, Amsterdam, Netherlands

We provide analytical bound on convergence rates for a class of hydrologic models and consequently derive a complexity measure based on Vapnik-Chervonenkis (VC) generalization theory. The class of hydrologic models is spatially explicit interconnected set of linear reservoir models with the aim to represent globally nonlinear hydrologic behavior by locally linear models. Here, by convergence rate we mean convergence of the empirical risk to the expected risk. The empirical risk is a measure of deviation of the modeled output from the observed output for a given data set and the expected risk is the expectation of the empirical risk. The derived measure of complexity measures model's propensity to overfit data (larger the complexity, higher is the risk of overfitting data) .

We explore how data finiteness can affect model selection for this class of hydrologic models and provide theoretical results on how model performance on finite sample converges to its expected performance as data size approaches infinity. These bounds can then be used for model selection, since the bounds provide a tradeoff between model complexity and model performance on available finite information. This is akin to a regularized solution to an inverse problem.

We find that convergence bounds for considered hydrologic models depend on the magnitude of its parameters, which are recession parameters of constituting linear reservoirs. We analytically show that complexity of hydrologic models not only varies with the magnitude of its parameters but also depends on the network structure of hydrologic models (in terms of the spatial heterogeneity of parameters and the nature of hydrologic connectivity). One important result is that complexity of a linear reservoir model increases with the magnitude of its recession parameter, i.e. faster reservoir models (or models with poor memory) have higher propensity to overfit than slower ones (or models with better memory of past events). We complement and reinforce our results with numerical quantification of model complexity based on a global optimization algorithm.