Geophysical Research Abstracts,
Vol. 10, EGU2008-A-03053, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-03053
EGU General Assembly 2008

Complexity based robust hydrologic prediction

S. Pande (1), M. McKee (2), L. Bastidas (2), M. Kemblowski (2), W. Walker (3)

(1) Center for World Food Studies (SOW-VU), Vrije Universiteit, Amsterdam, Netherlands, (2) Utah Water Research Laboratory, Utah State University, Logan, Utah, USA, (3) Department of Biological and Irrigation Engineering, Utah State University, Logan, UT, USA (s.pande@sow.vu.nl / Fax: (31) 20 - 598 9325)

Policy making in water resource management requires robust assessment of the consequences of future states of the resource. When dependent on prediction models, it requires assessment of the uncertainties associated with predictions. Ensemble prediction systems have been extensively used to address such issues and seek to provide a collection of predictions, via a collection of parameters, with intent to envelope future observations. Other robust variants of such a paradigm consider a subset of predictions that perform better than the rest on some calibration set. However, such methods do not have well-established finite-sample properties and generally require large samples to additionally determine better performing predictions. We here propose a new paradigm of robust parameter selection and prediction based on Vapnik-Chervonenkis (VC) generalization theory. It is based on a concept of complexity (that is data independent), which essentially relates finite sample performance of a model to its performance when large (tending to infinity) sample of the same underlying process is available. More complex models tend to overfit the same amount of data and therefore provide less reliable future predictions. For a sample size, one can then judge infinite sample performance of different parameters (or models) and base predictions accordingly. What we then have is a robust prediction based on the principle of Occams' razor – to select the simplest model amongst the set of better performing models on given data. Not only finite sample properties of such a paradigm are well studied, it also doesn't depend on data to judge which subset of parameters are better performing. In this paper, we test this for the simplest of hydrologic models- a k-nearest neighbor time series model. We i) quantifying its complexity, ii) demonstrate how it can be used

for robust prediction, and iii) compare its performance with a probabilistic non-linear prediction method on Leaf river stream flow data set.