Experiments in Auto-Calibration of Large Scale Environmental Models using a Data Mining Approach

David Swayne
Markiyan Sloboda
Simon Hood
Vimal Sharma

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

Swayne, David; Sloboda, Markiyan; Hood, Simon; and Sharma, Vimal, "Experiments in Auto-Calibration of Large Scale Environmental Models using a Data Mining Approach" (2008). International Congress on Environmental Modelling and Software. 43. https://scholarsarchive.byu.edu/iemssconference/2008/all/43

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Experiments in Auto-Calibration of Large Scale Environmental Models using a Data Mining Approach

David A. Swayne¹, Markiyan Sloboda¹, Simon Hood¹, Vimal Sharma¹, William Booty², David Lam²

¹Computing Research Laboratory for the Environment
Department of Computing and Information Science, University of Guelph, Guelph, Ontario, Canada
²National Water Research Institute, Burlington Ontario, Canada
{dswayne, vsharma, msloboda}@uoguelph.ca, shood01@guelphhumber.ca {Bill.Booty, David.Lam}@ec.gc.ca

Abstract: Many artificial intelligence paradigms are good candidates for the very difficult problem of model autocalibration. Quasi-Newton methods, optimization techniques and genetic algorithms are typical candidates for current algorithm research for this problem. A less well-travelled road to discovery consists of techniques for knowledge acquisition and data mining for autocalibration. This problem is largely intractable except by “expert” guessing, and the results are usually unsatisfactory. Our approach has been, from the environmental side, to use well-known and understood models, rather than develop our own. Many of these models involve large amounts of computer time for a single instance. We have partially overcome the computer time issue through the use of virtual files and running the codes on a clustered supercomputer architecture. These two preliminary steps have left us with a very typical knowledge discovery problem: several thousand model runs from which we hope to abstract the following: “good” solutions, knowledge of the trends toward these good solutions, and possibly a transformation of the model into (for instance) a Bayesian network approximation of the original model. This would allow us to “throw away” the computational model, instead selecting an output set which is a good approximation to what the model would produce under ideal circumstances. This research was not well-timed for a paper presentation at iEMSs in Barcelona, but our project has produced some preliminary successes, from which we hope to provoke discussion at this workshop.

Keywords: autocalibration, parallel computing, data mining, clustering