Computational Statistics and Data Analysis 71 (2014) 491–505

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Mixtures of experts for understanding model discrepancy in dynamic computer models

David J. Nott [a,*], Lucy Marshall [b], Mark Fielding [c], Shie-Yui Liong [d]

[a] Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore
[b] Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717-3120, United States
[c] School of Mathematics and Applied Statistics, University of Wollongong, 2522, Australia
[d] Tropical Marine Science Institute, National University of Singapore, Singapore, 119227, Singapore

ARTICLE INFO

Article history:
Received 1 June 2012
Received in revised form 8 February 2013
Accepted 29 April 2013
Available online 13 May 2013

Keywords:
Bayesian inference
Computer model
Emulator
Mixtures of experts

ABSTRACT

There are many areas of science and engineering where research and decision making are performed using computer models. These computer models are usually deterministic and may take minutes, hours or days to produce an output for a single value of the model inputs. Fitting mixtures of experts of computer models where the expert components use different values of the computer model parameters is considered. The efficient calibration of such models using emulators, which are fast statistical surrogates for the computer model, is discussed. It is argued that mixtures of experts are often insightful for describing model discrepancy and ways in which the computer model can be improved. This is not a strength of standard approaches to the statistical analysis of computer models where a certain "best input" assumption is usually made and model discrepancy is often described through a stationary Gaussian process prior on the discrepancy function. Application of the framework is presented for a dynamic hydrological rainfall–runoff model in which the mixture approach is helpful for highlighting model deficiencies.

1. Introduction

Computer models play an important role in research in many areas of modern science and engineering, including climatology, meteorology, oceanography, materials science and hydrology. Often the computer models used in these fields are deterministic and computationally intensive, requiring minutes, hours or days to run for a single value of the model inputs. This makes assessments of uncertainty in these models and the systems they represent a difficult problem. Statisticians have played an active role in developing a methodology for uncertainty assessment in computer models for some time, with increasing attention being paid to the area recently. Key problems include calibration of model parameters, prediction, sensitivity analysis and model validation. A recent non-technical overview of the area is O'Hagan (2006).

In this article we will mostly be concerned with model assessment, and in particular with attempting to understand deficiencies in the computer model and how it might be improved. We propose describing model discrepancy using an expansion of the statistical models conventionally used in the calibration of computer models. The expansion involves considering mixtures of experts models (Jacobs et al., 1991; Jordan and Jacobs, 1994) in which different values of the computer model parameters are allowed to hold in different regimes related to some covariates. This relaxes the "best input" assumption which is commonly used and criticized in the analysis of computer modeling experiments. We are mostly concerned in this paper with dynamic models with time series outputs and our later example is concerned with a hydrological rainfall–runoff

* Corresponding author. Tel.: +65 6516 2744; fax: +65 6872 3919.
E-mail address: standj@nus.edu.sg (D.J. Nott).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.04.020

model for prediction of streamflow. In that context, for example, it might be assumed that different model parameters are appropriate for describing the output in "high flow" and "low flow" situations, and examination of the computer model parameter estimates in the two states and the posterior probabilities for the two states over time can be informative about model deficiencies. We believe that asking modelers to think about model inadequacy in terms of regime specific parameters for the current model and covariates that might be informative for prediction of such regimes is a powerful and practical one for eliciting judgements about model discrepancy.

There is a large literature on computer model validation in the broadest sense. Santner et al. (2003) give an overview. Bayarri et al. (2007) give a Bayesian framework for model validation but they are mostly concerned with addressing the question of whether predictions from the model are adequate for an intended purpose and not directly with diagnosing model deficiencies. Wilkinson et al. (2011) assume a structural form for the discrepancy at the level of the dynamics with a linear regression term for bias and a white noise Gaussian process to quantify the residual variability. In hydrology, the idea of diagnosing model inadequacies by allowing model parameters to be time varying goes back many years (for example Beck and Young, 1976). Smith et al. (2008) have considered time varying parameter models with particle filtering approaches to computation. They undertake model assessment by considering what distribution for time varying parameters would be needed to reproduce the observed data, and consider whether the estimated distributions are physically plausible. Reichert and Mieleitner (2009) also consider time varying parameter models, but formulate a systematic approach to identifying possible causes of model bias by considering the degree of bias reduction that can be achieved by relaxing time invariance for different parameters. Time varying parameters are also used in hydrology as a way of dealing with input uncertainties (Kuczera et al., 2006), although issues of input uncertainty and model inadequacy are sometimes dealt with less formally such as in the generalized likelihood uncertainty (GLUE) method (Beven and Binley, 1992; Beven et al., 2008; Liu et al., 2009b). The time varying parameter approaches to model assessment are very similar to our approach in this article but there are a number of differences. First, the clustering of different output components induced by our mixture model can be very helpful for interpretation. Second, our approach is modular in the sense of only requiring runs of the computer model for fixed parameter values — these fixed parameter runs are then mixed together in the mixture model. In contrast, continuously time varying parameter approaches generally allow parameters to change in the one step transitions that define the dynamic computer model. Modifying computer model code to handle this kind of time varying parameter may in some cases be a non-trivial exercise. However, associated with this is a corresponding advantage — allowing time varying parameters in dynamic one step transitions generally allows us to respect physical constraints such as mass balance, whereas our approach with mixing may not (although the individual fixed parameter model runs do respect such constraints). However, we believe that for model assessment purposes this may not be a big disadvantage for our method and the modularity we have mentioned is an important compensating advantage.

Most existing approaches to the analysis of computer models handle uncertainty due to model inadequacy by a discrepancy function that is usually given a stationary Gaussian process prior, possibly with an added mean term involving covariates (see, for instance Kennedy and O'Hagan, 2001). Many authors have noted the desirability of improving on this and perhaps the most ambitious framework yet suggested for understanding model discrepancy is the reified modeling method of Goldstein and Rougier (2009). They criticize the commonly used "best input" approach to analysis of computer models, where evaluation of the simulation model at a single best input is sufficient for understanding the system being modeled, and encourage thinking about model discrepancy by imagining a better model (the reified model) that one might construct and about the relationship between the reified model and the current one. Implementation of the reified modeling approach in practice is enormously demanding as noted by the discussants of Goldstein and Rougier (2009) although in their response to the discussion Goldstein and Rougier (2009) suggest some simplifications. Recently Goldstein et al. (2008) suggest an interesting alternative using a multi-model ensemble although this does rely on such an ensemble being available. The approach of Goldstein et al. (2008) builds on previous work on Bayes linear methodology for the analysis of computer modeling experiments (Craig et al., 1996, 2001; Goldstein and Rougier, 2006). Our mixture framework can be thought of as relaxing the assumptions of a "best input" and of a simple additive model discrepancy term for the purposes of model assessment, and we believe that mixtures and the use of hierarchies is an easily implemented way of eliciting some expert knowledge about model deficiencies. As we discuss later, though, this framework has some limitations of its own. In particular, in some applications not all significant model deficiencies will be able to be reduced by the use of a time dependent parameter in the current model. Nevertheless we have found the framework effective and insightful in a number of applications.

Mixtures of experts have been used in hydrological modeling by Marshall et al. (2006) (see also Marshall et al., 2007a,b). However, the general framework described here contains a number of advances. First, in Marshall et al. (2006) it is assumed that model discrepancy can be accounted for completely by mixing over models — they note the possibility of using correlated errors in the mixture components but this was not implemented. Here we consider mixtures of experts of time series models for the discrepancy, so that model inadequacy is described both by the use of mixture component specific parameters in the computer model and of component dependent autoregressive residuals. Second, we also consider here efficient approaches to fitting mixtures of experts of computer models using emulators. The construction of dynamic emulators that can work with long time series outputs is a complex task and we use a recently developed method due to Liu and West (2009) for this purpose. The development of emulator based methods for fitting our mixture models is particularly significant since without an emulator evaluating the mixture model likelihood once requires running the computer model several times for different parameter values. Also, in general there is a price to be paid in terms of speed of convergence of the MCMC scheme for estimation of the mixing mechanism so that longer run times may be required compared to a single