Spatial Statistics 12 (2015) 96–108

Contents lists available at ScienceDirect

Spatial Statistics

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

Quantifying uncertainty for temperature maps derived from computer models

Lucia Paci [a,*], Alan E. Gelfand [b], Daniela Cocchi [a]

[a] Department of Statistical Science at the University of Bologna, Via delle Belle Arti, 41 - 40126 Bologna, Italy

[b] Department of Statistical Science at Duke University, Box 90251, Durham NC 27708-0251, USA

ARTICLE INFO

Article history:
Received 3 July 2014
Accepted 24 March 2015
Available online 3 April 2015

Keywords:
Hierarchical modeling
Measurement error
Numerical models
LogCAR process
MCMC

ABSTRACT

Computer models are often deterministic simulators used to predict several environmental phenomena. Such models do not associate any measure of uncertainty with their output since they are derived from deterministic specifications. However, many sources of uncertainty exist in constructing and employing numerical models.

We are motivated by temperature maps arising from the Rapid Update Cycle (RUC) model, a regional short-term weather forecast model for the continental United States (US) which provides forecast maps without associated uncertainty.

Despite a rapidly growing literature on uncertainty quantification, there is little regarding statistical methods for attaching uncertainty to model output without information about how deterministic predictions are created. Although numerical models produce deterministic surfaces, the output is not the 'true' value of the process and, given the true value and the model output, the associated error is not stochastic. However, under suitable stochastic modeling, this error can be interpreted as a random unknown. Then, we infer about this error using a Bayesian specification within a data fusion setting, fusing the computer model data with some external validation data collected independently over the same spatial domain. Illustratively, we apply our modeling approach to obtain an uncertainty map associated with RUC forecasts over the northeastern US.

* Corresponding author. Tel.: +39 051 2094657; fax: +39 051 232153.
E-mail addresses: lucia.paci2@unibo.it (L. Paci), alan@stat.duke.edu (A.E. Gelfand), daniela.cocchi@unibo.it (D. Cocchi).

http://dx.doi.org/10.1016/j.spasta.2015.03.005

1. Introduction

Many computer models are deterministic simulation models developed, for example, to predict environmental phenomena such as temperatures or air pollution levels. In a spatial setting, numerical model outputs are displayed in the form of maps, provided as averages over grid cells, usually at high spatial and temporal resolution. Such computer models do not associate any measure of uncertainty with their output since they are derived from deterministic specifications. However, many sources of uncertainty exist in constructing and employing numerical models. In fact, with computer models providing spatially referenced output, uncertainty maps can be a useful tool to guide environmental agencies in refining and improving computer models. Furthermore, when we use model output as predictor for an environmental variable, we might seek to evaluate how these uncertainties propagate from the model output to the forecasting of the response. Altogether, it seems there is need for quantifying uncertainty in this setting.

The motivating context for us are temperature maps that arise from the Rapid Update Cycle (RUC) numerical weather model, a regional short term weather forecast model for the continental United States. This model yields maps that are publicly available, provided by the National Oceanic and Atmospheric Administration (NOAA)'s National Climatic Data Center (NCDC) but with no explicit detail regarding their development and no associated uncertainty in their forecasts. Our contribution, articulated in detail below, is to propose a hierarchical stochastic model along with the introduction of a validation data set consisting of temperature measurements collected at monitoring stations operating in the same study region. The model fuses the two data sources to enable assessment of uncertainty associated with RUC maps.

In applications, the sources of potential uncertainty associated with numerical models include input uncertainty, function uncertainty, model discrepancy and observational error (Cumming and Goldstein, 2010). The Bayesian approach represents a natural way to account for all of these uncertainty sources and several methods have been developed to deal with the uncertainty analysis for complex computer models. Customarily, numerical models are implemented as computer codes, dependent upon a number of inputs which determine the nature of the output. These inputs represent unknown parameters and the uncertainty about them propagates through the numerical model, inducing uncertainty in the output.

A general statistical framework has been presented by Raftery et al. (1995) for mapping from a set of input parameters to a set of model outputs, the so-called Bayesian synthesis, eventually led to the Bayesian melding approach (Poole and Raftery, 2000) Also, statistical methods have been proposed to handle sensitivity analysis which is concerned with understanding how the model output is influenced by changes in the model inputs (e.g. Oakley and O'Hagan, 2004). For deterministic numerical models, i.e. models with no random components, their predictions are subject to error because any model is a simplification of reality. So, model output will not equal the 'true' value of the process of interest and this discrepancy is well-known as model inadequacy (Kennedy and O'Hagan, 2001).

Structural uncertainty, which is introduced by scientific choices of model design and development, can be also quantified by analyzing multi-model ensembles. In this case, the output consists of different versions of a numerical model, i.e. a model is run several times with different initial conditions (scenarios). Statistical approaches for quantifying uncertainty with ensembles have recently received considerable attention (see e.g. Gneiting et al., 2005; Raftery et al., 2005; Berrocal et al., 2007; Smith et al., 2009; Di Narzo and Cocchi, 2010; Kleiber et al., 2011; Sloughter et al., 2013).

There is little in the literature about statistical methods for attaching uncertainty to model output when we do not have information about how such deterministic predictions are created, i.e., we have no information about model inputs. Our contribution to uncertainty quantification builds upon the notion of uncertainty introduced by Ghosh et al. (2012) when numerical models are unavailable; rather, only deterministic outputs at some spatial resolution are provided. In other words, we do not know how the deterministic surfaces have been developed. For us, they come from a entirely unknown 'black box'. Ghosh et al. (2012) proposed a general Bayesian approach to associate uncertainties with deterministic interpolated surfaces which requires some external validation data collected independently over the same spatial domain as the deterministic map.