Atmospheric Environment 43 (2009) 1204–1217

Contents lists available at ScienceDirect

Atmospheric Environment

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

The impact of spatial correlation and incommensurability on model evaluation

Jenise L. Swall*, Kristen M. Foley

Atmospheric Modeling Division, National Exposure Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 30 July 2008
Received in revised form
16 October 2008
Accepted 19 October 2008

Keywords:
Block kriging
Air quality modeling
Spatial interpolation
Statistical simulation

ABSTRACT

Standard evaluations of air quality models rely heavily on a direct comparison of monitoring data matched with the model output for the grid cell containing the monitor's location. While such techniques may be adequate for some applications, conclusions are limited by such factors as the sparseness of the available observations (limiting the number of grid cells at which the model can be evaluated) and the incommensurability between volume-averages and pointwise observations. We examine several sets of simulations to illustrate the effect of incommensurability in a variety of cases distinguished by the type and extent of spatial correlation present. Block kriging, a statistical method which can be used to address the issue, is then demonstrated using the simulations. Lastly, we apply this method to actual data and discuss the practical importance of understanding the impact of spatial correlation structure and incommensurability.

1. Introduction

The performance of an air quality model is typically evaluated against actual measurements of the pollutant in question from monitoring networks. Such models treat the region as a large grid, giving output for each cell, so that the analyst must determine how to make the comparison. In the most common situation, the observed value from each monitor is matched with the value for the grid cell in which the monitor is located. The resulting paired data are used to evaluate the performance of the model, both visually, using such graphical displays as scatterplots and spatial plots, and numerically, using various performance metrics such as bias and root mean squared error. Examples of comprehensive air quality model evaluations which utilize this approach include Eder et al. (2006), Eder and Yu (2006), and Appel et al. (2007).

For example, consider observed and modeled maximum 8 h ozone values in the northeastern United States on June 14, 2001. Fig. 1(a) shows the observations recorded at 124 air monitoring stations in the region in parts per billion (ppb). Output from a Community Multiscale Air Quality model (Byun and Schere, 2006) simulation for the same day is pictured in Fig. 1(b). This model run utilizes grid cells with each side of length 12 km. (Henceforth, we refer to grid cells by their side length, e.g. "12 km grid cell", "36 km grid cell"). As with many such models, the value for a grid cell represents the volume-average for the layer of the atmosphere closest to the earth's surface over the extent of the grid cell. A scatterplot and summary statistics of the sort often used in traditional evaluation approaches are shown in Fig. 2. A spatial plot of the differences between the model-monitor pairs is given in Fig. 3.

Figs. 1–3 are sufficient to form a general picture of model performance. An examination of Figs. 2 and 3 reveals that, for this particular day, the model underpredicts maximum 8 h ozone at more monitoring sites than it overpredicts. Fig. 3 shows that while underprediction seems to be an issue along the Canadian border, we have a mix of overprediction and underprediction in other areas, particularly for coastal sites.

While these figures are helpful in understanding how the model output compares with the observations, none of them is sufficient for a detailed assessment of model performance. For instance, model metrics (such as those shown in Fig. 2) can only be calculated for grid cells in which we have monitors, and this means that the model metrics may overly reflect model performance in areas with large numbers of monitors. These might most often be urban areas or regions which have been pinpointed for further study due to a perceived greater likelihood of problems. None of these plots allows assessment of the model's performance at unmonitored locations. Also, the effect of measurement error or other sources of fine-scale variability cannot be adequately considered. Lastly, to better interpret Fig. 2, it would be helpful to understand to what extent the differences between model-simulated values and observed values may be due to the inherent differences between point measurements and volume-averages.

The problem of comparing grid averages and point measurements is usually termed "incommensurability" in the atmospheric science literature. Statisticians refer to this same issue as one of "change of support"; the issue and the underlying mathematics are discussed more thoroughly by Gelfand et al. (2001), while Gotway

* Corresponding author. Tel.: +1 919 541 7655; fax: +1 919 541 1379.
E-mail addresses: swall.jenise@epa.gov (J.L. Swall), foley.kristen@epa.gov (K.M. Foley).

1352-2310/$ – see front matter Published by Elsevier Ltd.
doi:10.1016/j.atmosenv.2008.10.057

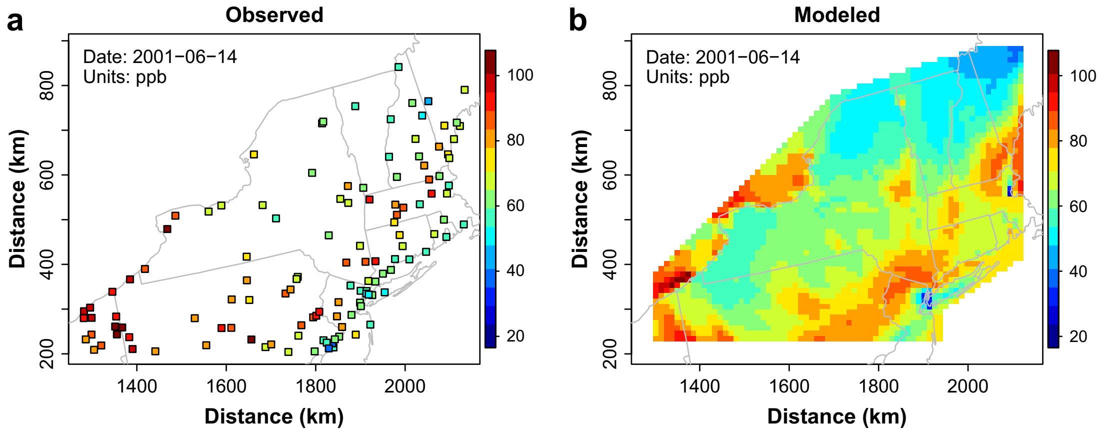

Fig. 1. Observed and modeled maximum 8 h ozone (2001-06-14).

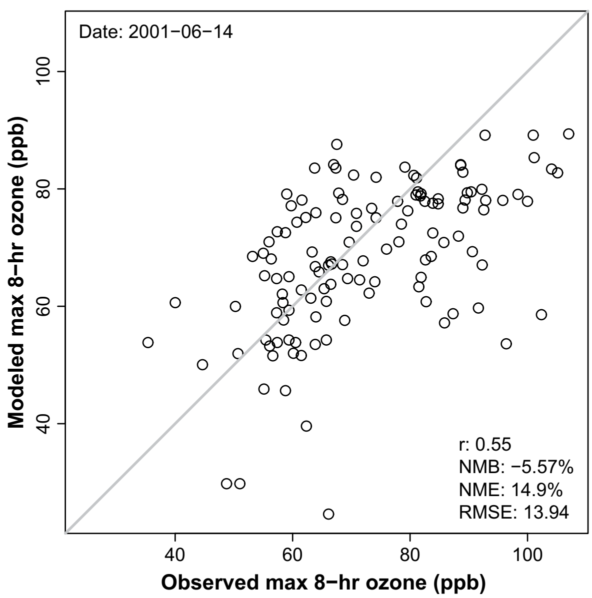

Fig. 2. Modeled vs. observed maximum 8 h ozone (2001-06-14).

and Young (2002) give an extensive review of statistical methods for combining data with different spatial supports proposed for a variety of scientific applications. Although in the context of air quality modeling most evaluations have not considered this issue in detail, some authors have considered the problem and developed statistical methods for addressing it. The papers by Fuentes et al. (2003), Fuentes and Raftery (2005), Swall and Davis (2006), and Davis and Swall (2006) present sophisticated statistical models for various applications, all of which address the incommensurability issue and are able to estimate pollutant levels for grid cells in which no observations lie. Even so, since there are no consistently available, regionally comprehensive sources of observational data at heights beyond that of the typical monitoring station, each of these techniques treat the model output as areal, rather than volume, averages, and work herein follows suit.

In this paper, we discuss these issues from an applied statistical perspective. We introduce simulated datasets to explore the impact of various spatial correlation structures on common model evaluation tools. Using these simulated cases, we illustrate the benefit of the "block kriging" technique, which uses the observations and the spatial correlation among them to estimate the levels of a pollutant at all of the grid cells, whether or not they contain monitors. This technique has the advantage of being relatively easy to implement in statistical software packages and of requiring relatively few assumptions, compared with some of the more complex approaches developed by the above authors. We compare and contrast this technique with traditional, point-based kriging techniques. Lastly, we apply these ideas to selected real-life cases and demonstrate their utility as part of a focused model assessment strategy.

2. Simulations

In this section, we make use of two sets of simulated spatial fields to demonstrate the potential impact of incommensurability

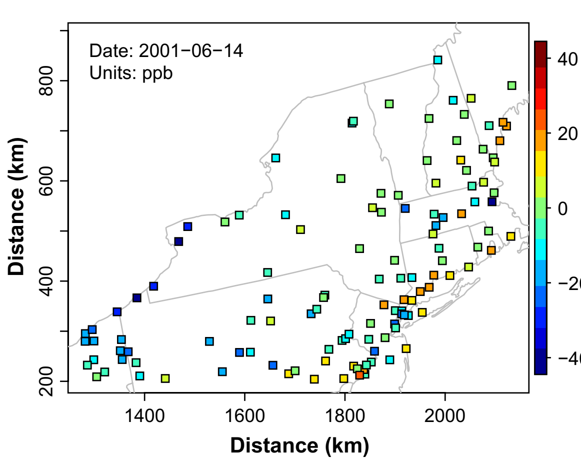

Fig. 3. Differences (modeled – observed) in maximum 8 h ozone (2001-06-14).

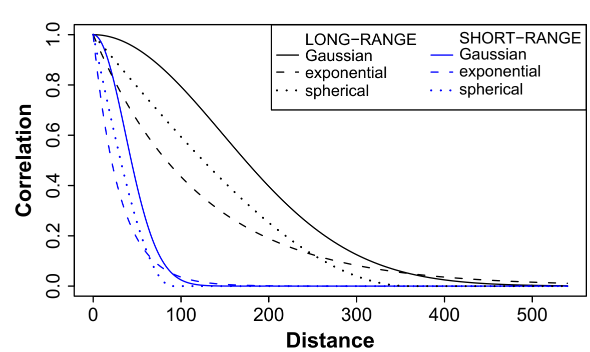

Fig. 4. Correlograms for long-range and short-range simulations.