Spatio-temporal model for short-term predictions of air pollution data

Francesca Bruno[1], Lucia Paci[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 28

[1] Department of Statistical Sciences "Paolo Fortunati" - University of Bologna, Via delle Belle Arti, 41 - Bologna
francesca.bruno@unibo.it
lucia.paci2@unibo.it

Abstract

Recently, the interest of the many Environmental Agencies is on short-term air pollution predictions referred at high spatial resolution. This permits citizens and public health decision-makers to be informed with visual and easy access to air quality assessment. We propose a hierarchical spatio-temporal model to enable use of different sources of information in order to provide short-term air pollution forecasting.

Keywords: Bayesian hierarchical model; data fusion; ozone forecasting; MCMC.

1 Introduction

Recently, several Environmental Agencies are interested to provide the public and experts with visual and easy access to air quality information. Short-term air pollution predictions are usually needed as spatial patterns. Numerical models devoted to estimate air pollution are available in short-time frames at high spatial resolution. However, these forecasts are biased and not equipped by any uncertainty measure. Conversely, measurements at monitoring stations give the 'true' air pollution level. The joint consideration of the two sources of information can improve air pollution forecasting [5, 4].

In this work, ozone data from the monitoring network (41 stations) and numerical model output are fused to obtain accurate short-term predictions of ozone level in the Emilia Romagna region. Two numerical models are available:

the first one is the multi-scale Chimere chemistry-transport which estimates the ozone levels (and other pollutants) for gridded cells (at 5 km^2 resolution) over successive time periods up to 72 hours in the future. Also, the weather numerical model Cosmo is available. This is a non-hydrostatic atmospheric model developed by the Consortium for Small-scale Modeling. Cosmo model is run every day producing 72 hours forecasts at 7 km^2 spatial resolution for several meteorological variables. Here, we consider the hourly temperature forecasts produced by Cosmo at the grid cells spanning the region, since the temperature influences directly the kinetics of reactions producing ozone [2].

2 Modeling

Let $Y_t(\mathbf{s})$ denote the hourly ozone concentration at a location $\mathbf{s}$, $A(\mathbf{s})$ be the altitude of site $\mathbf{s}$ and $C_t(B)$ define the numerical model output over the grid cell B. Following the downscaling approach developed by [1], we address the spatial misalignment by associating to each site $\mathbf{s}$ the grid cell $A(\mathbf{s})$ that contains $\mathbf{s}$. Then, the model links the observational data and the numerical model output as follows:

$$Y_t(\mathbf{s}) = \beta_0 + \beta_{0,t}(\mathbf{s}) + \big(\beta_1 + \beta_{1,t}(\mathbf{s})\big)C_t(B) + \beta_2 A(\mathbf{s}) + \epsilon_t(\mathbf{s}) \tag{1}$$

where $\epsilon_t(\mathbf{s})$ is a white noise process with nugget variance τ^2. The spatio-temporally varying coefficients $\beta_{0,t}(\mathbf{s})$ and $\beta_{1,t}(\mathbf{s})$ are modeled as independent zero-mean processes with a separable covariance structure. The orography of the region is also taken into account since the ozone level changes according to the elevation.

Model (1) is fitted using a Gibbs sampler, using both the air quality model output (Chimere) and the weather model output (Cosmo). We use data of 48-hour running windows starting at each hour from 10AM on August 10th to 9AM on August 11th, 2012.

3 Results

We perform a site-one-out cross-validation experiment producing 3-hours ahead forecasts of ozone level for 24 consecutive windows.

Table 1 provides a comparison of out-of-sample 3-hours ahead ozone predictions. In terms of Mean Squared Prediction Error (MSPE), model (1) with Cosmo output as covariate outperforms the one fitted using Chimere output. This is due to the high degree of smoothness of the air quality model output. Also, the MSPE tends to increase as the length of the forecast period increases, as we expected. The inclusion of the altitude, $A(\mathbf{s})$, in model (1) improves the ozone forecasting.

	1-hour	2-hour	3-hour
Chimere	364.81	568.92	825.93
Cosmo	278.97	350.47	461.14
Cosmo without $A(\mathbf{s})$	394.26	479.18	624.99

Table 1: MSPE for 3-hours ahead ozone forecasts obtained from model (1) with Chimere, with Cosmo and with Cosmo excluding elevation.

Figure 1 (left panel) shows the 1-hour ahead forecast map obtained from the model using Cosmo output. The posterior standard deviation map in the right panel gives a measure of the uncertainty associated with the predictions.

Model (1) is simple, very flexible and computationally efficient. Accurate ozone forecasts are obtained in short-time frames along with associated uncertainty.

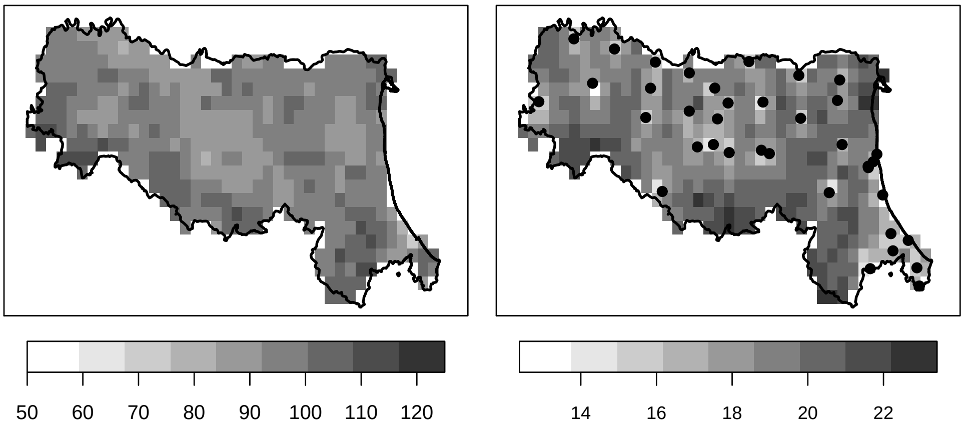

Figure 1: The 1-hour ahead ozone forecast map (left panel) and the standard deviation map (right panel) at 12PM on 12th August. Black dots represent monitoring sites.

References

[1] V. J. Berrocal, A. E. Gelfand, D. M. Holland. **A Spatio-temporal Downscaler for Output from Numerical Model**. *Journal of Agricultural, Biological and Environmental Statistics*; 2010; 14; pp. 176-197.

[2] D. Cocchi, C. Trivisano. **Ozone**. In *Encyclopedia of Environmetrics*, A.-H. El-Shaarawi and W. Piegorsch (eds), John Wiley & Sons Ltd: Chichester, UK. DOI: 10.1002/9780470057339.vao022.pub2; 2013.

[3] A. E. Gelfand, H-J Kim, C. F. Sirmans and S. Banerjee. **Spatial Modeling with Spatially Varying Coefficient Processes**. *Journal of the American Statistical Association*; 2003; 98; pp. 387-396.

[4] L. Paci, A. E. Gelfand, D. M. Holland. **Spatio-temporal modeling for real-time ozone forecasting**. *Spatial Statistics*; 2013; Submitted.

[5] S. K. Sahu, S. Yip, D. M. Holland. **A fast Bayesian method for updating and forecasting hourly ozone levels**. *Environmental and Ecological Statistics*; 2009; 18; pp. 185-207.