Nonparametric spatial-temporal analysis of $SO_2$ across Europe

Marco Giannitrapani, Adrian Bowman, Marian Scott, Ron Smith
Dept. of Statistics, University of Glasgow, G12 8QW Scotland, U.K.
Tel: +44(0)141 330 6118 - Fax: +44(0)141 330 4814
email: marco@stats.gla.ac.uk
May 9, 2004

Keywords: additive model, smoother, backfitting algorithm, variogram

1 Abstract

From the 1970’s, a co-ordinated international programme monitoring acidifying air pollution was initiated in direct response to observed acidification. At the same time, several international protocols on the reduction of acidifying and eutrophying emissions ($SO_2$, $SO_4$, etc.) were also agreed.

This work presents an evaluation of the observed spatial and temporal trends in $SO_2$ in Europe for the last quarter of the twentieth century, on the basis of data from EMEP (Co-operative Programme for Monitoring and Evaluation of the Long Range Transmission of Air Pollutants in Europe). The policy question of interest is whether the protocols have resulted in a real improvement in environmental quality and a real change in the acidifying environment. In the first part of the work, we report on nonparametric modelling of the temporal trends, accounting for the effect of meteorological covariates using generalized additive models (GAMs). The second part of the work considers the spatial patterns in the $SO_2$ field and their temporal evolution.

2 Additive Models

Modelling spatial and temporal trends, accounting for the effect of meteorological covariates, can be carried out with additive models proposed by Hastie and Tibshirani [4]. These models are based on the idea of modelling a response variable $y$ through one dimensional smoothers as building
blocks for a restricted class of nonparametric multiple regression models. The additive model that they proposed has the followed form:

\[ y = \alpha + \sum_{j=1}^{p} m_j(x_j) + \varepsilon \]  

where the errors \( \varepsilon \) are independent of the \( x_j \), \( E(\varepsilon) = 0, \) \( Var(\varepsilon) = \sigma^2 \) and the \( m_j \) are arbitrary univariate smooth functions, one for each predictor. It is important to note that one of the major advantages of such a model is that the nature of the effects of a predictor on the response does not depend on the values of the other predictors. This means that the smoothing is always one dimensional and consequently any dimensionality problems we avoided, at the possible cost of an approximation of the surface.

Estimation of additive models can be carried out through the backfitting algorithm, which is a general regression-type fitting mechanisms. Conditional expectations provide a simple intuitive motivation for the backfitting algorithm. If the additive model is correct, then for any \( k \), \( E(y - \alpha - \sum_{j \neq k} m_j(x_j|x_k) = m_k(x_k) \). This immediately suggests an iterative algorithm for computing all the \( m_j \) (Hastie and Tibshiriani [4]). Local linear regression has been used to estimate each component of the additive model (1). Bowman & Azzalini [1] gives an introduction to this form of smoothing.

3 Modelling time trend accounting for meteorology

The data analyzed are the concentrations of weekly \( SO_2 \) monitored at Eskdalemuir (GB02), Westerland (DE01), Waldhof (DE02), Schauinsland (DE03), Deuselbach (DE04), Brotjacklriegel (DE05), and Kosetice (CZ03), from 1973 up to 2001. These stations are all part of the EMEP monitoring network. The predictors used are: years, weeks of the year, weekly amount of precipitation, weekly temperature mean, weekly humidity mean, weekly mean of wind direction weighted by wind speed. Because of the skewness of \( SO_2 \) and of the amount of rainfall, these two variables were transformed to the log scale. The effect of predictors such as “weeks of the year” and “wind direction” have been estimated using circular smoothers to reflect this feature of the recorded data. A version of local linear regression smoother which incorporates correlated errors in the response variable has been used.

\[ \log(\text{SO}_2) = \mu + m_y(\text{years}) + m_w(\text{weeks}) + m_r(\text{rain}) + m_t(\text{temperature}) + m_h(\text{humidity}) + m_{wds}(\text{wind.direction.speed}) + \varepsilon \]  

Fig. 1 shows the fit of the model (2) (red line) for the seven sites, including all meteorological variables. The blue lines display the years components and they show generally decreasing patterns across the seven sites. However, these trends are not always monotonic and have different shapes. Fig. 2
shows the effect of each component of the model (2) on $SO_2$, at Deuselbach (DE04). At this site, the components indicate that higher $SO_2$ concentrations (over the years) have been associated with the period November to January and with very low mean temperatures. Rainfall decreases the concentration but so also does low humidity, and the higher concentrations are associated with wind directions towards the east. The covariate effect of year indicates that concentrations were steadily decreasing from 1988 to 2001.

4 Modelling spatial trends

A spatial analysis for $SO_2$ in Europe has been carried out. Monthly means of $SO_2$ from 1990 to 2001 at 125 sites across Europe have been analyzed using kriging procedures. Fig.3 a) shows the observed spatial pattern in $SO_2$ in August 2001, and estimated spatial trends based on a simple plane and universal kriging. For a given time point $i$, the plane $y_{(i)} = \beta_0(i) + \beta_1(i) \text{lat}_{(i)} + \beta_2(i) \text{long}_{(i)} + \varepsilon_{(i)}$ has been fitted in order to estimate the trend. Three different theoretical variograms have been fitted to the residuals: the spherical, the exponential and the Gaussian. Each of them has been fitted using a weighted least squares procedure (Cressie N. [3]), that is not dependent on a particular sample estimator. To minimize Cressie’s weighted least squares, the Gauss-Newton algorithm was used. Using a Gaussian variogram, Fig.3 c) & d), show respectively the kriging predictions and the standard errors for kriging predictions.

The spatial plane with a Gaussian variogram was then fitted to the $SO_2$ concentrations at each time point. The temporal pattern of the coefficients for latitude and longitude across time and the variogram estimates were studied. Figures 4 a), b), c) are the plots of the estimated parameters of the variograms (range, sill, and nugget effect) across time. From these plots, the variogram parameters appear approximately constant across time. Figures 4 d), e), f) show that the coefficients are roughly constant, while the intercept values seems to be decreasing, suggesting that level of $SO_2$ are on average changing. The package S+SpatialStats has been mostly used for this analysis.

5 Modelling spatial temporal trends

Having seen that the spatial structure is approximately constant over time, the following additive models have been fitted.

\[
\begin{align*}
\log SO_2 &= \alpha + m_1(\text{years}) + m_2(\text{months}) + m_3(\text{long}, \text{lat}) + \varepsilon \quad (3) \\
\log SO_2 &= \alpha + m_{12}(\text{years, months}) + m_3(\text{long}, \text{lat}) + \varepsilon \quad (4)
\end{align*}
\]
Figure 1: fits of the additive model (2) for log(SO$_2$) at: a) Eskdalemuir (GB02), b) Westerland (DE01), c) Waldhof (DE02), d) Schauinsland (DE03), e) Deuselbach (DE04), f) Brotjackkriegel (DE05), g) Kosetice (CZ03).
Figure 2: fit of each component of model (2) at Deuselbach (DE04). a) $m_y$(years); b) $m_w$(weeks); c) $m_r$(rain); d) $m_t$(temperature); e) $m_h$(humidity); f) $m_wds$(wind.direction.speed).
Figure 3: Contour plots of: a) observed values, b) estimated trend (plane), c) kriging predictions, d) standard errors for kriging predictions for $SO_2$ in August 2001.

Figure 4: Temporal plots of: a) Ranges, b) Sills, c) Nuggets, d) Intercept, e) Coefficient for Longitude, f) Coefficient for Latitude.
Assuming a separable model, it is possible to define the full correlation matrix $\Sigma$ for both models, as the direct product (known also as the Kronecker product) of the temporal correlation matrix $\Theta$ with the spatial correlation one $\Gamma$, namely $\Sigma = \Theta \otimes \Gamma$.

It is also necessary to note that, because of the amount of temporal data across space, fitting this model with the backfitting algorithm is computationally extremely expensive. In order to avoid this dimensionality problem, a procedure called binning has been used (Bowman & Azzalini [2]). Indicating by $b$ the number of bins, and denoting the local mean estimator as $\hat{y} = Sy$, where $S$ denotes the smoothing matrix, with binning the dimensionality of the smoothing matrices is controlled by $b$ rather then by $n$, which is a very considerable reduction. In fact the smoothing matrices can be written as $\hat{y} = BSDy$, where $S$ is a smoothing matrix of dimension $b \times b$, $D$ is the matrix that reduces the response variable $y$ to the binned data $\bar{y}$ ($y = D\bar{y}$), and $B$ is the matrix that expands the binned values $\bar{y}$ back to the full vector $y$. The backfitting algorithm has been amended for the binned case. Both models, (3) and (4), have been fitted to the monthly means of $SO_2$ from 1990 to 2001 at 125 sites across Europe using binning. The fits of both models are shown in Fig.5. and Fig.6. The difference between the two models is that model (3) assume the same seasonal structure across time, while model (4) allows the seasonality to change across years.

These models have been tested, applying the approximate $F$ test (Bowman & Azzalini [1]). The $p$ value of the test is 0.011. This significant $p$ value means that model (4) is a better fit compared to model (3). In other words it means that there has been a significant change in seasonality across years.

Acknowledgement Marco Giannitrapani acknowledges the support from the Chancellors fund of the University of Glasgow, and the Centre for Ecology and Hydrology in Edinburgh.

References


Figure 5: Fits of the components (a) $m$ (years), (b) $m$ (months), and (c) $m$ (latitudes, longitudes) for model 3.

Figure 6: Fits of the components (a) $m$ (years, months), and (b) $m$ (latitudes, longitudes) for model 4.