A simple spatio-temporal procedure for the prediction of air pollution levels

Jorge M. Mendes¹ and Kamil F. Turkman²*

¹Higher Institute of Statistics and Information Management, New University of Lisbon, Lisbon, Portugal
²CEAUL Faculty of Sciences, University of Lisbon, Lisbon, Portugal

Received 12 February 2001; Revised 3 September 2002; Accepted 3 October 2002

In this paper we study the spatio-temporal behaviour of air pollutants measured daily over the city of Lisbon, Portugal. Our specific aim is to predict air pollutant levels in time and space over a fine grid of locations based on observations from a small number of monitoring sites. Our suggested prediction procedure is based on the simple and intuitive idea of first making predictions in time at the monitoring sites and then extending these predictions in space to locations other than the monitoring sites using kriging methods. Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: spatio-temporal models; kriging; multivariate time series

1. INTRODUCTION

The city of Lisbon is the biggest urban area in Portugal. It is located by the estuary of the river Tagus at 38°44’ N, 9°8’ W. Its total area is 84 km², corresponding to a population density of 7.272 inhabitants per km².

Owing to its location on the Atlantic coast and the north winds that often blow over the city, high concentrations of pollutants are rare. However, under adverse meteorological conditions, weak winds and thermic inversions, especially in autumn and winter, high concentrations of pollutants can occur.

Cars and public transportation are the main source of atmospheric pollution in the city. Emissions from industrial facilities do not have a great influence, since over the last few years the main industries have been deactivated. At present the monitoring network is composed of eight monitoring stations, but there are plans to expand the network to 14 stations. Several pollutants are monitored over the network, such as carbon monoxide (CO), nitrogen oxides (NO and NO₂), sulphur dioxide (SO₂), nitrogen O₃ and particulate matter (PM₁₀). Owing to the relative importance of traffic emissions, three of these pollutants, CO, NO and NO₂, are monitored at all eight stations.

The spatial distribution of the network was set up by empirical guidelines. Because of the expected high pollutant concentration levels, locations with a high intensity of traffic and a high density of pedestrians were chosen, as well as the eastern residential areas of the city where, until some years ago, the major industries were concentrated. Figures 1 and 2 show the geographical distribution of the eight monitoring stations over a map of Lisbon.

At each monitoring station, air samples are collected continuously. Several devices then analyse the sample, and the concentration level of each pollutant is determined based on optical characteristics of the air sample. Electrical signals are generated and converted to numerical values that are sent, by a phone line, to a central computer where the data are validated and processed.

In this paper our objective is to study the spatio-temporal...

E-mail: kamil.turkman@fc.ul.pt
Contract/grant sponsor: Portuguese Science Foundation (FCT); Contract/grant number: PRAXIS/P/1/1001/98.
behaviour of air pollutants measured daily over the city of Lisbon. Although data are available for daily concentrations of many different pollutants, the study will be restricted to nitrogen dioxide (NO₂) owing to the fact that this pollutant is known to have the most adverse effect on human health. Specifically, our basic objective is to make short-term daily predictions of NO₂ concentrations over Lisbon in space and time using a very simple model for our data, which permits the use of existing time series models and kriging methods to obtain multivariate predictions of NO₂ concentrations over a fine grid of points. At present the number of monitoring stations is being increased and it is hoped that this study will help in deciding the number of new monitoring sites as well as their positions. The model suggested in this paper can also be used as a basis for devising optimal alarm systems for issuing warnings of high pollution levels [1].

The suggested procedure is based on the simple and intuitive idea of first making predictions in time at the monitoring locations and then extending these predictions in space to locations where there are no observations, using appropriate kriging methods. Almost all environmental processes show variability over space and time, involving complicated spatio-temporal structures and interactions. Bayesian hierarchical models [see e.g. References [2,3]] seem to be particularly suited for the modelling of spatio-temporal processes. Such models provide strategies for incorporating complicated space-time interactions at different stages of the hierarchy, thus making it relatively feasible to implement in high dimensions and for large data sets. In some cases, when the number of monitoring sites as well as the data sets are very large and high resolution of predictions is needed, then the implementation of Bayesian hierarchical models may not be easy or even feasible, as these methods require complete specification of the joint spatio-temporal covariance structures. In such cases, dimension-reducing approaches (see e.g. Reference [4]) may be the most appropriate way of handling the data. However, all these methods need advanced statistical knowledge and computational skills. We believe that the suggested simple procedure may serve as a reference to modelling spatio-temporal data for those practitioners who do not have easy access to more sophisticated methodologies such as Bayesian hierarchical models and MCMC (Markov Chain Monte Carlo) methods.

The outline of the paper is as follows. In Section 2 we suggest a model, clearly indicating the assumptions on the dependence structure and on the distributions which give rise to the suggested prediction procedure. The preliminary data analysis together with the practical modelling aspects will be given in Section 3, where we respectively look at the problem of missing observations, the estimation of spatio-temporal trend and seasonal components as well as the estimation of spatial and temporal dependence structures. Section 4 is dedicated to spatio-temporal predictions, and in Section 5 we present a simple model validation procedure. Section 6 is dedicated to the discussion of prediction error of the model. In Section 7 we give some conclusions together with possible improvements in the model and the data analysis.

2. THE MODEL

We assume that \(X(s, t), s \in \mathcal{D}, t=0, 1, \ldots \) is the spatio-temporal process generating the NO₂ data, which are observed daily at \(n=8\) observation sites over 5 years, so that the data are in the form

\[ x(s_i, t), \quad i = 1, \ldots, 8, \quad t = 1, 2, \ldots, 1826 \]

Here \(\mathcal{D} \subset \mathbb{R}^2\) is some region representing the geographical area of Lisbon. Let

\[ X(t) = (X^{(1)}(t), X^{(2)}(t)) \]  

where

\[ X^{(2)}(t) = \{X(s_1, t), X(s_2, t), \ldots, X(s_n, t)\} \]

is the \(n\)-dimensional time series representing the values of pollution concentration at the \(n\) observation sites and

\[ X^{(1)}(t) = \{X(s_{i+1}, t), X(s_{i+2}, t), \ldots, X(s_{i+m}, t)\} \]

is an \(m\)-dimensional time series of pollution concentrations at \(m\) locations in \(\mathcal{D}\) where there are no observations and we want spatial predictions. Thus the data are of the form \(x(t) = (X^{(1)}(t), X^{(2)}(t), t = 1, 2, 1826)\), where \(X^{(1)}(t)\) is the unobservable vector. Since the observed time series \(X^{(2)}(t)\) show indications of non-constant variance, a logarithmic transformation is applied to the data. Hence, from now on, our process \(X(s, t)\) represents the transformed pollution levels.

Now we will write our likelihood, underlying clearly the assumptions.

**Assumption 1**

For any \(t\),

\[ p(x^{(1)}(t) | x^{(2)}(t), x(t-1), \ldots, x(1)) = p(x^{(1)}(t) | x^{(2)}(t)) \]  

In (3) we assume that at any time \(t\) the pollution levels at unobserved sites, conditional on observation sites at time \(t\) and all the past information contained up to time \(t-1\), depend on the values of pollution level at the observed sites at the same time point \(t\).
Assumption 2

\[ p(x^{(2)}(t)|x(t-1), \ldots, x(1)) = p(x^{(2)}(t)|x^{(2)}(t-1), \ldots, x^{(2)}(t-p)) \]

for some integer \( p \). In (4) we assume that the process at observed sites, conditional on all the information up to \( t-1 \), depends only on the observations taken on the same sites at time points going back to some time lag \( p \).

These two assumptions are an over simplification of the reality, separating the spatial and temporal dependence structures in a very specific manner, facilitating inference. Note that Equation (3) defines the spatial dependence whereas Equation (4) defines the temporal dependence.

Based on these two assumptions, the likelihood for the data

\[ L = p(x(t), x(t-1), \ldots, x(1)) = p(x(t)|x(t-1), \ldots)p(x(t-1)|x(t-2), \ldots)p(x(1)) \]

separates into two products

\[ \prod_{t=p+1}^n p(x^{(1)}(t)|x^{(2)}(t)) \]

and

\[ \prod_{t=p+1}^n p(x^{(2)}(t)|x^{(2)}(t-1), \ldots, x^{(2)}(t-p)) \times p(x^{(2)}(p), \ldots, x^{(2)}(1)) \]

representing separately the spatial and temporal structures. Hence the respective sets of model parameters can be estimated separately.

We now make our assumptions on the probability distributions given in (3) and (4).

1. \( p(x^{(1)}(t), x^{(2)}(t)) \sim N(\mu(s, t), \Sigma(s)) \)  
   \[ \mu(s, t) = m(s) + v(s, t) \]  
   Here \( m(s) \) is the spatial trend and \( v(s, t) \) is the space-dependent time seasonal component to be specified later.

2. The detrended, deseasonalized time series \( Y^{(2)}(t) = (X(s, t) - \mu(s, t), t = 1, \ldots, 8) \) is a multivariate Gaussian process with mean zero and covariance matrix

\[ \text{cov}(Y^{(2)}(t + h), Y^{(2)}(t)) = \Sigma^{(2)}(h) \]  

Here the covariance matrix will be uniquely defined by assuming that the detrended, deseasonalized series is a stationary multivariate autoregressive (AR) process.

These assumptions clearly suggest how to make predictions of the pollution levels \( (X^{(1)}(t + h), X^{(2)}(t + h)) \) at origin \( t \) in two stages.

- Estimate the multivariate \( AR(p) \) structure in (7) from the deseasonalized series \( Y^{(2)}(t) \) to obtain the predictor \( Y^{(2)}(t + h) \) for \( Y^{(2)}(t + h) \).
- Estimate the spatial structure in (5) from the detrended, deseasonalised series \( Y^{(2)}(t) \) and then use an appropriate kriging method to predict the values \( Y^{(1)}(t + h) \) for \( Y^{(1)}(t + h) = X^{(1)}(t + h) \) using the predicted values \( Y^{(2)}(t + h) \) instead of \( Y^{(2)}(t + h) \).
- The predictions for \( X^{(1)}(t + h) \) and \( X^{(2)}(t + h) \) can then be obtained by adding the trend and seasonal components calculated at respective locations on \( Y^{(1)}(t + h) \) and \( Y^{(2)}(t + h) \).

3. The data and the estimation

The data we use hereafter are the daily mean concentration levels of NO₂ between 1 January 1994 and 31 December 1998 at the eight observation sites shown in Figures 1 and 2. However, we do not consider the observations obtained at the Benfica station (observation site \( s_5 \); see Figure 2) in this analysis for several reasons. First, the series obtained at this site has too many missing observations. Second, we believe that this site will distort the spatial continuity owing to its separation from the other stations by Monsanto park, which is of considerable size and negligible NO₂ pollution.

The daily data obtained at the eight stations show a large number of missing observations. For instance, out of 1826 daily observations, the station at Rua da Prata (observation site \( s_7 \); see Figure 2) has 376 missing observations, of which 223 are consecutive. The missing observations at each site are filled in by a model which takes into consideration the strong seasonal component, moving averages (MAs) of adjacent daily, weekly and yearly observations and appropriate random components. As such, the missing observation \( X_{m}(s_i, t) \) at time \( t \) of the series observed at location \( s_i \) is calculated from the model

\[ X_{m}(s_i, t) = S(s_i, t) + MA(s_i, t) + e(s_i, t) \]  

The seasonal component at each site is estimated by a trigonometric regression of the form

\[ \sum_{l=1}^{3} f_l(s_i) \cos(\lambda_l t) + \sum_{l=1}^{3} g_l(s_i) \sin(\lambda_l t) \]

with frequencies \( \lambda_l \) at 2πt/7, 2πt/90 and 2πt/365. The moving average component at each site is made up of the weighted average of three separate moving averages:

\[ MA_1(s_i, t) = 0.75 MA_1(s_i, t) + 0.25 MA_2(s_i, t) + 0.25 MA_3(s_i, t) \]

where \( MA_1(s_i, t), MA_2(s_i, t) \) and \( MA_3(s_i, t) \) are the moving averages of 1, 7 and 365 day lagged observations respectively. The residual component \( e(s_i, t) \) in (8) is generated from i.i.d. normal distributions fitted to deseasonalized data at each observation site.

Figures 3 and 4 show respectively the original data sets with missing observations and the data sets with estimated missing observations.
Figure 3. Daily mean concentration (log-transformed) of NO$_2$ in Lisbon from 1/1/94 to 31/12/98: (a)–(h) $s_1$ (Olivais), $s_2$ (Entrecampos), $s_3$ (Beato), $s_4$ (Chelas), $s_5$ (Rua da Prata), $s_6$ (Casal Ribeiro), $s_7$ (Benfica) and $s_8$ (Av. Liberdade).

Figure 4. Daily mean concentration (log-transformed) of NO$_2$ in Lisbon from 1/1/94 to 31/12/98 after filling in missing observations: (a)–(h) $s_1$ (Olivais), $s_2$ (Entrecampos), $s_3$ (Beato), $s_4$ (Chelas), $s_5$ (Rua da Prata), $s_6$ (Casal Ribeiro), $s_7$ (Benfica) and $s_8$ (Av. Liberdade).
3.1. Spatio-temporal trend and seasonal components

We assume that at any point \( s \) in \( D \) and time \( t \) the log-transformed process \( X(s, t) \) is of the form

\[
X(s, t) = \mu(s, t) + Y(s, t)
\]

where

\[
\mu(s, t) = m(s) + v(s, t)
\]

\[
m(s) + \sum_{k=1}^{3} f_k(s) \cos(\lambda_k t) + \sum_{k=1}^{3} g_k(s) \sin(\lambda_k t)
\]

In order to be able to extrapolate the spatial trend function \( m(s) \) and the temporal seasonal component \( \sum_{k=1}^{3} f_k(s) \cos(\lambda_k t) + \sum_{k=1}^{3} g_k(s) \sin(\lambda_k t) \) estimated from the observation sites to any point in space, we assume simple functional relationships of the form

\[
m(s) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_1^2 + a_5 x_1 x_2
\]

\[
f_k(s) = b_0 k + b_1 k x_1 + b_2 k x_2
\]

\[
g_k(s) = c_0 k + c_1 k x_1 + c_2 k x_2
\]

where \( x_1 \) and \( x_2 \) are the east and north co-ordinates of the point \( s \), i.e. \( s = (x_1, x_2) \), and \( \lambda_k \) are the fixed frequencies \( 2\pi/7 \), \( 2\pi/90 \) and \( 2\pi/365 \) respectively. The coefficients are then estimated by least squares.

Once the data at the eight locations are detrended and deseasonalized by \( X(s, t) - \mu(s, t) \), then the covariance structures \( \Sigma(s) \) and \( \Sigma^{(2)}(t) \) given in (5) and (7) respectively can be estimated from the detrended and deseasonalized data.

3.2. Estimation of the spatial dependence structure

The sample covariance between values at stations \( s_i \) and \( s_j \)

\[
i = j = 1, \ldots, 7,
\]

is

\[
c_{ij} = \frac{1}{1826} \sum_{t=1}^{1826} y(s_i, t)y(s_j, t)
\]

The spatial dispersion \( d_{ij}^2 \) can be calculated by

\[
d_{ij}^2 = \text{var} [y(s_i, t) - y(s_j, t)] = c_{ij} + c_{ji} - 2c_{ij}
\]

Note that if the spatial process is second-order stationary, then the spatial dispersion and the sample variogram are identical [5]. Figure 5 shows the spatial configuration of the seven stations and the graph of the corresponding spatial dispersions as a function of the geographical distance between the stations, while Table 1 gives the values of spatial dispersion between the seven stations.

It is quite clear from Figure 5 that the spatial dispersion is not a function of the geographical distance. This may be due to the fact that the pollution concentrations are higher along the large avenues.

In order to estimate the spatial dependence structure under such conditions, we use a methodology suggested in Reference [5]. The basic idea is to relocate the observation sites according to the new distance of dispersion, as

<table>
<thead>
<tr>
<th>Station</th>
<th>( s_1 )</th>
<th>( s_2 )</th>
<th>( s_3 )</th>
<th>( s_4 )</th>
<th>( s_5 )</th>
<th>( s_6 )</th>
<th>( s_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>26.62</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>15.09</td>
<td>23.97</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_4 )</td>
<td>20.68</td>
<td>32.64</td>
<td>27.85</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_5 )</td>
<td>31.97</td>
<td>38.92</td>
<td>32.26</td>
<td>26.94</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_6 )</td>
<td>27.99</td>
<td>31.85</td>
<td>24.73</td>
<td>34.78</td>
<td>34.81</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( s_7 )</td>
<td>31.20</td>
<td>32.97</td>
<td>29.18</td>
<td>28.35</td>
<td>34.77</td>
<td>36.95</td>
<td>—</td>
</tr>
</tbody>
</table>
Figure 6. (a) Spatial configuration in D-plane: $s_1$ (Olivais), $s_2$ (Entrecampos), $s_3$ (Beato), $s_4$ (Chelas), $s_5$ (Rua da Prata), $s_6$ (Casal Ribeiro) and $s_7$ (Av. Liberdade). (b) Spatial dispersion versus MSI distance. The curve represents the exponential semivariogram fitted to the MSI distance.

compared to the original locations positioned according to geographical distance. This can be achieved by using multidimensional scaling, which generates a two-dimensional co-ordinate representation of the sampling stations with interpoint distances representing the sample spatial dispersion. Once the sampling stations are repositioned, a thin plate spline interpolation for each co-ordinate can be made to provide a mapping between the representations of the stations in the two co-ordinate systems. Other coordinates where predictions are needed can easily be transformed from one co-ordinate system to another using this fitted thin plate spline interpolation. As in Reference [5], we call the original co-ordinate system the geographical plane (G-plane) and the transformed co-ordinate system the dispersion plane (D-plane). Let $s_i = (x_{i1}, x_{i2})$ be a point in the G-plane. The corresponding D-plane point $s'_i = (y_{i1}, y_{i2})$ is calculated by fitting the thin plate splines:

$$ s'_i = (y_{i1}, y_{i2}) = (f_1(s_i), f_2(s_i)) $$

where

$$ f_l(s_i) = z_l^{(0)} + z_l^{(1)} x_{i1} + z_l^{(2)} x_{i2} + \sum_{j=1}^{7} b_{lj}^{(l)} k_l(s_i), \quad l = 1, 2 $$

and $k_l(s) = |s - s_i|^2 \log(|s - s_i|)$. Figures 6(a) and 6(b) show respectively the configuration of the seven stations in the new D-plane and the plot of the spatial dispersion as a function of the new distance, which we call the multidimensional scaling interpoint (MSI) distance. Table II shows the fitted thin plate splines.

The strong relation between the spatial dispersion and the new MSI distance is now very clear from Figure 6(b). An isotropic omnidirectional exponential semivariogram is fitted to the dispersion in Figure 6(b):

$$ \gamma(h) = 0.007 + 0.3899 \left[ 1 - \exp \left( -\frac{3h}{1.087} \right) \right] $$

Spatial prediction or kriging can now be performed for any location using this estimate, by first transforming the coordinates of the new location in the G-plane to the D-plane using the estimated spline functions. We return to this point later after modelling the temporal component.

3.3. Estimation of the temporal component

We assume that the deseasonalized series $Y(s_i, t) = (X(s_i, t) - \mu(s_i, t), t = 1, \ldots, 1826, i = 1, \ldots, 7)$ follow a stationary multivariate AR(p) process. However, in order to validate the spatio-temporal predictions, we will model only the five stations at $s_2$ (Entrecampos), $s_3$ (Beato), $s_4$ (Chelas), $s_6$ (Casal Ribeiro) and $s_7$ (Av. Liberdade), leaving out the time series observed at locations $s_1$ (Olivais) and $s_5$ (Rua da Prata) for

<table>
<thead>
<tr>
<th>Table II. Estimated thin plate splines</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_l^{(0)}$</td>
</tr>
<tr>
<td>$s_1$</td>
</tr>
<tr>
<td>$s_2$</td>
</tr>
<tr>
<td>$s_3$</td>
</tr>
<tr>
<td>$s_4$</td>
</tr>
<tr>
<td>$s_5$</td>
</tr>
<tr>
<td>$s_6$</td>
</tr>
<tr>
<td>$s_7$</td>
</tr>
<tr>
<td>$s_1$</td>
</tr>
<tr>
<td>$s_2$</td>
</tr>
<tr>
<td>$s_3$</td>
</tr>
</tbody>
</table>
validation. The geographical positions of these stations are shown in Figure 7.

Point predictors for

\[ Y(s_2, t + h), Y(s_3, t + h), Y(s_4, t + h), \]
\[ Y(s_5, t + h), Y(s_6, t + h) \]

will form the basis of kriging to predict the values of pollution level over an area within the co-ordinates 111E–116.5E and 195N–200N, as well as predicting the observed values at locations \( s_1 \) (Olivais) and \( s_2 \) (Rua da Prata) for validating the model. The grid locations as well as the stations used for validation are shown in Figure 7.

A multivariate AR(4) is fitted to the deseasonalized time series at the five locations using the software Pest [6]. The analysis of the residuals showed that the fitted model is adequate.

4. SPATIO-TEMPORAL PREDICTIONS

The spatio-temporal prediction is made using the following procedure.

1. Predict \( h \)-step-ahead pollution levels \( Y(s_2, t + h), Y(s_3, t + h), Y(s_4, t + h), Y(s_5, t + h) \) and \( Y(s_6, t + h) \) for different values of \( h \) using the fitted multivariate AR(4) model.

2. Based on the estimated semivariogram (11) and using ordinary kriging, predict the values of \( Y(s_0, t + h) \) based on the predicted values \( Y(s_2, t + h), Y(s_3, t + h), Y(s_4, t + h), Y(s_5, t + h) \) and \( Y(s_6, t + h) \).

3. The values of actual pollution level \( X(s_0, t + h) \) at \( s_0 \) can now be calculated for time \( t + h \) by adding to \( \hat{Y}(s_0, t + h) \) the spatial mean function \( m(s_0) \) and the spatio-temporal seasonal component \( \bar{v}(s_0, t + h) \) given in (10), and then log-transforming back the series.

The plot of one-step-ahead spatio-temporal predictions for the whole grid for \( t = 731 \) is shown in Figure 8; other one-step- as well as \( h \)-step-ahead prediction surfaces for different values of \( t \) can also be calculated without any additional difficulty.

5. MODEL VALIDATION

In order to validate the model, we fitted AR(4) models to the five locations. Then we predicted the first 7 days of January 1999 (one-step-ahead predictions only) with this model and extended the spatial predictions explained above to the two observation sites left for validation, namely stations \( s_1 \) (Olivais) and \( s_2 \) (Rua da Prata). Table III and Figure 9 show these predictions together with the observed pollution levels at each of these two locations for those 7 days. We also applied a leave-two-out approach to all combinations of stations, in which the pollution levels at two stations are predicted as a function of the data at the remaining stations for all combinations. Table IV gives a summary of the quality of predictions for this leave-two-out validation. Results show that the predictions are not always good across the whole region. One possibility for explaining these mixed results is that the monitoring stations are installed at locations where more pollution is expected, thus introducing bias in the predictions.

Assumptions 1 and 2 in Section 2 need to be validated. However, this may not be an easy task owing to the fact that these conditions relate to the dependence structure of the process between the observation sites and other sites where there are no observations. Assumptions 1 and 2 can be validated by looking at the partial cross-correlations and partial autocorrelations of the process at observed and unobserved sites. Since an estimate of this function cannot be obtained owing to the lack of observations at unobserved

---

**Table III.** One-step-ahead spatio-temporal predictions and actual values of daily mean concentration (log-transformed) of NO\(_2\) for \( s_1 \) (Olivais) and \( s_2 \) (Rua da Prata)

<table>
<thead>
<tr>
<th>( t )</th>
<th>( \hat{Y}(s_1, t) )</th>
<th>( \hat{Y}(s_2, t) )</th>
<th>( Y(s_1, t) )</th>
<th>( Y(s_2, t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>731</td>
<td>3.38</td>
<td>3.38</td>
<td>4.48</td>
<td>4.09</td>
</tr>
<tr>
<td>732</td>
<td>3.83</td>
<td>3.87</td>
<td>4.70</td>
<td>3.95</td>
</tr>
<tr>
<td>733</td>
<td>3.86</td>
<td>3.82</td>
<td>3.41</td>
<td>3.41</td>
</tr>
<tr>
<td>734</td>
<td>3.21</td>
<td>3.31</td>
<td>3.70</td>
<td>3.84</td>
</tr>
<tr>
<td>735</td>
<td>3.94</td>
<td>3.79</td>
<td>3.88</td>
<td>4.00</td>
</tr>
<tr>
<td>736</td>
<td>3.90</td>
<td>3.70</td>
<td>4.66</td>
<td>4.03</td>
</tr>
<tr>
<td>737</td>
<td>4.34</td>
<td>4.07</td>
<td>5.09</td>
<td>4.39</td>
</tr>
</tbody>
</table>
and actual values of daily mean concentration (log-transformed) of NO₂ for s₁ (Olivais) and s₅ (Rua da Prata).

Figure 9. One-step-ahead spatio-temporal predictions, confidence intervals and actual values of daily mean concentration (log-transformed) of NO₂ for s₁ (Olivais) and s₅ (Rua da Prata).

sites, we looked at the various forms of partial cross-correlations and autocorrelations of the process at observed sites. Estimated partial cross-correlations and autocorrelations die off exponentially fast in most cases, indicating that Assumptions 1 and 2 reasonably capture the dependence structure of the data. However, owing to the complicated distributional forms of these empirical partial cross-correlations, it is not possible to give formal statistical tests.

6. PREDICTION ERRORS

In the previous section we showed how one can obtain the spatio-temporal prediction for the detrended and deseasonalized pollution concentrations Y(s₀, t + h) for some fixed point in space, s₀, and time t + h. If observations at five locations are available at time t + h, then the predictor of Y(s₀, t + h) is given by

\[ E[Y(s₁, t + h)|y_{1,5}] = \sum_{i=1}^{5} \lambda_i y(s_i, t + h) \]

where \( \lambda = (\lambda_i, i = 1, \ldots, 5) \) are kriging weights to be estimated from the data. However, observations \( y(t + h) = (y(s_i, t + h), i = 1, \ldots, 5) \) are not available and our suggested predictor is

\[ \hat{Y}(s₀, t + h) = E[Y(s₀, t + h)|\bar{Y}(t + h)] \]

where

\[ \bar{Y}(t + h) = E[Y(t + h)|y(t), y(t - 1), \cdots] \]

and \( Y(t) \) denotes the five-dimensional time series. If the observations \( y(t + h) = (y(s_i, t + h), i = 1, \ldots, 5) \) are available,

Table IV. Summary of prediction results for leave-two-out validation

<table>
<thead>
<tr>
<th>Combination number</th>
<th>Stations in</th>
<th>Stations out</th>
<th>Standard mean errors</th>
<th>% predicted values inside 95% confidence limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>s₂ s₄ s₅ s₆ s₇</td>
<td>s₁ s₆</td>
<td>1.9, -1.3</td>
<td>14.3, 85.7</td>
</tr>
<tr>
<td>3</td>
<td>s₂ s₃ s₄ s₅ s₆</td>
<td>s₁ s₃</td>
<td>2.3, 0.8</td>
<td>28.6, 100.0</td>
</tr>
<tr>
<td>4</td>
<td>s₂ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>2.2, 0.2</td>
<td>28.6, 100.0</td>
</tr>
<tr>
<td>5</td>
<td>s₂ s₃ s₄ s₅ s₆</td>
<td>s₁ s₅</td>
<td>1.9, 1.3</td>
<td>42.9, 71.4</td>
</tr>
<tr>
<td>6</td>
<td>s₂ s₃ s₄ s₅ s₆</td>
<td>s₁ s₂ s₃ s₄ s₅</td>
<td>2.2, -0.1 s₁ s₆ s₇</td>
<td>28.6, 100.0</td>
</tr>
<tr>
<td>7</td>
<td>s₂ s₃ s₄ s₅ s₆</td>
<td>s₁ s₂ s₃ s₄ s₅</td>
<td>0.1, 1.5 s₁ s₂ s₃ s₄</td>
<td>85.7, 42.9</td>
</tr>
<tr>
<td>8</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.1, 0.0</td>
<td>100.0, 85.7</td>
</tr>
<tr>
<td>9</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₅</td>
<td>0.0, 0.6</td>
<td>100.0, 100.0</td>
</tr>
<tr>
<td>10</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₆</td>
<td>-0.3, -1.3</td>
<td>85.7, 85.7</td>
</tr>
<tr>
<td>11</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₆</td>
<td>0.2, 0.8</td>
<td>100.0, 100.0</td>
</tr>
<tr>
<td>12</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>1.7, 0.8</td>
<td>42.9, 85.7</td>
</tr>
<tr>
<td>13</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.4, 0.6</td>
<td>57.1, 100.0</td>
</tr>
<tr>
<td>14</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.2, -1.4</td>
<td>71.4, 85.7</td>
</tr>
<tr>
<td>15</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.8, 0.8</td>
<td>57.1, 100.0</td>
</tr>
<tr>
<td>16</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.2, 0.4</td>
<td>85.7, 100.0</td>
</tr>
<tr>
<td>17</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>-0.1, -1.2</td>
<td>85.7, 85.7</td>
</tr>
<tr>
<td>18</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.2, 0.6</td>
<td>85.7, 100.0</td>
</tr>
<tr>
<td>19</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.2, -1.2</td>
<td>100.0, 85.7</td>
</tr>
<tr>
<td>20</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₄</td>
<td>0.3, -0.2</td>
<td>77.6, 93.9</td>
</tr>
<tr>
<td>21</td>
<td>s₁ s₃ s₄ s₅ s₆</td>
<td>s₁ s₆</td>
<td>-1.2, 0.5</td>
<td>85.7, 100.0</td>
</tr>
</tbody>
</table>

then the prediction error variance
\[
\hat{\sigma}^2(s_0, t + h) = \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
\[= \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h)) y(t + h)]^2 \] (13)

can be calculated using the standard arguments of kriging. However, \(y(s, t + h)\) are not available and are predicted from the observed data up to time \(t\). What is the effect of using \(\hat{Y}(s_0, t + h) = \mathbb{E}[Y(s_0, t + h)|Y(t + h)]\) on the prediction error? Let
\[
\hat{\sigma}^2(s_0, t + h) = \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
Then, by arguments given in the Appendix,
\[
\hat{\sigma}^2(s_0, t + h) = \sigma^2(s_0, t + h) + 2\Sigma(s, t + h)\hat{\mu}^T\] (14)
where \(\Sigma(s, t + h)\) is the \(h\)-step-ahead prediction error obtained from the multivariate autoregressive model fitted to the data. Figure 10 shows the one-step-ahead prediction error surfaces calculated from (14). 95% confidence intervals for the one-step-ahead predictions at the two validation sites are given in Figure 9.

7. CONCLUSIONS
We believe that the simple spatio-temporal model suggested captures the essential spatio-temporal dependence structure in the data and gives reasonable predictions. Owing to the specific conditional independence conditions, the model permits us to calculate the spatial and temporal predictions as well as their errors in an additive way. Although by no means the best way of modelling the data owing to the restrictive assumptions made on the dependence structure, this simple model permits us to use widely available software on kriging and multivariate time series. In principle, the quality of the short-term predictions should be drastically increased by incorporating covariates such as wind direction and intensity, temperature and precipitation. However, at present, no such data were available for the authors in conjunction with the pollution data. These covariates would also be highly valuable in reducing the effect of non-stationarity on the predictions.

Acknowledgements
This work was supported by the Portuguese Science Foundation (FCT; project PRAXIS/P/1001/98).

APPENDIX
Let
\[
\hat{\sigma}^2(s_0, t + h) = \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
and
\[
\sigma^2(s_0, t + h) = \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
Then
\[
\hat{\sigma}^2(s_0, t + h) = \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
\[= \mathbb{E}[(Y(s_0, t + h) - \hat{Y}(s_0, t + h)) y(t + h)]^2 \] (15)
The first expression on the right-hand side of (15) is \(\hat{\sigma}^2(s_0, t + h)\), whereas
\[
E[(\hat{Y}(s_0, t + h) - \hat{Y}(s_0, t + h))^2]
\]
\[= E \left[ \sum_{i=1}^{5} \lambda_i Y(s_i, t + h) - \sum_{i=1}^{5} \lambda_i \hat{Y}(s_i, t + h) \right]^2 \]
\[= E \left[ \sum_{i=1}^{5} \lambda_i \varepsilon(s_i, t + h) \right]^2 \]
\[= \Sigma(s, t + h)\] (16)
where \(\varepsilon(s, t + h) = (Y(s_i, t + h) - \hat{Y}(s_i, t + h), i = 1, ..., 5)\) is the \(h\)-step-ahead prediction error obtained from the five-dimensional AR(4) model at origin \(t\), and \(\Sigma(s, t + h)\) is the variance-covariance matrix of \(\varepsilon(s, t + h)\). The third expression on the right-hand side of (15) can be written as
\[
2E[(Y(s_0, t + h) - \hat{Y}(s_0, t + h)) \times (\hat{Y}(s_0, t + h) - \hat{Y}(s_0, t + h))] \]
\[= 2E \left[ \varepsilon(s_0, t + h) \sum_{i=1}^{5} \lambda_i \varepsilon(s_i, t + h) \right] \]
\[= \sum_{i=1}^{5} \lambda_i E[\varepsilon(s_i, t + h) \varepsilon(s_i, t + h)] \] (17)
where \(\varepsilon(s_0, t + h) = Y(s_0, t + h) - \hat{Y}(s_0, t + h)\) is the spatial prediction error. Note that the spatial and the \(h\)-step-ahead temporal prediction errors \(\varepsilon(s_0, t + h)\) and \(\varepsilon(s_i, t + h), i = 1, ..., 5\) are not independent, but they are independent conditional on \(Y(s_0, t + h)\). Hence
\[
E[\varepsilon(s_0, t + h) \varepsilon(s_i, t + h)]
\]
\[= \int E[\varepsilon(s_0, t + h) \varepsilon(s_i, t + h) | Y(s, t + h) = y] dF_{Y(z, t + h)}(y) \]
\[= \int E[\varepsilon(s_0, t + h) | Y(s, t + h) = y] E[\varepsilon(s_i, t + h) | y] dF_{Y(z, t + h)}(y) \]
Since
\[
E[\varepsilon(s_0, t + h) | Y(s_0, t + h) = y] = Y(s_0, t + h) + E[Y(s_0, t + h) | Y(s, t + h) = y] = 0
\]
we see that the third term on the right-hand side of (15) is zero and
\[
\hat{\sigma}^2(s_0, t + h) = \sigma^2(s_0, t + h) + 2\Sigma(s, t + h)\hat{\mu}^T \] (18)
REFERENCES