Predictive spatio-temporal models for spatially sparse environmental data

Xavier de Luna* and Marc G. Genton†

October 17, 2003

Abstract

We present a family of spatio-temporal models which are geared to provide time-forward predictions in environmental applications where data is spatially sparse but temporally rich. That is measurements are made at few spatial locations (stations), but at many regular time intervals. When predictions in the time direction is the purpose of the analysis, then spatial-stationarity assumptions which are commonly used in spatial modeling, are not necessary. The family of models proposed does not make such assumptions and consists in a vector autoregressive (VAR) specification, where there are as many time series as stations. However, by taking into account the spatial dependence structure, a model building strategy is introduced, which borrows its simplicity from the Box-Jenkins strategy for univariate autoregressive (AR) models for time series. As for AR models, model building may be performed either by displaying sample partial correlation functions, or by minimizing an information criterion. Two environmental data sets are studied. In particular, we find evidence that a parametric modeling of the spatio-temporal correlation function is not appropriate because it rests on too strong assumptions. Moreover, we propose to compare model selection strategies with an out-of-sample validation method based on recursive prediction errors.

Keywords: Accumulated prediction errors; Spatio-temporal correlation; Partial correlation; Vector autoregression.

*Department of Statistics, Umeå University, S-90187 Umeå, Sweden. E-mail: xavier.deluna@stat.umu.se

†Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. E-mail: genton@stat.ncsu.edu

1 Introduction

In this article we present a model building strategy designed to work within a family of vector autoregressive models for time series being recorded at specific spatial locations. The methodology has been developed with environmental applications in mind where measurements on a variable are made at regular time intervals and at several stations located within a specific area. More specifically, we focus on situations were measurements are available at a few stations −the spatio-temporal data is sparse in space but rich in time. We put the discussion into concrete form with two examples treated previously in the literature.

The first data set that we consider consists in average daily wind speeds measured at 11 synoptic meteorological stations located in Ireland during the period 1961-78, making up for 6,570 observations per location. Gneiting (2002) used this data set to illustrate the lack of realism implied by the separability assumption, where the correlation function is written as a product of a purely spatial correlation function and a purely temporal correlation function. Modeling directly the space-time correlation by an appropriate parametric family of functions is inspired from the geostatistical tradition where spatial dependences are of main interest. Such models make strong spatial stationarity assumptions: typically isotropy is assumed where spatial correlations are a function only of the distance between stations. When spatio-temporal correlation functions are considered, an isotropy-like assumption is typically made, where the relative location of the stations is again considered as irrelevant; see, e.g., Mardia and Goodall (1993), Host, Omre, and Switzer (1995). These spatial stationarity assumptions are mainly appropriate for data sets rich in space (so that they can be checked) and when the purpose is to obtain predictions at spatial locations where there are no measurements available. In order to apply the geostatistical approach to environmental space-time data, many efforts have been made resulting in the introduction of transformations achieving spatial stationarity (e.g., Guttorp, Meiring, and Sampson, 1994) or by constructing flexible enough covariance models, e.g., by allowing non-separable covariance functions, see Gneiting (2002). A geostatistical inspired analysis for spatio-temporal data which does not make neither spatial stationarity nor separability assumptions was recently proposed by Stroud, Müller, and Sansó (2001). Their approach is flexible but is relevant for data which is rich both in time and in space.

A spatio-temporal data set with measurements available only at a few stations was analyzed in Tonellato (2001). Carbon monoxide atmospheric concentrations were observed at four stations located in the city of Venice. For each station 300 hourly records made in September 1995 are available. The purpose of Tonellato's analysis was to provide forecasts based only on past observations made at the different locations. This was achieved with a multivariate time series state space model together with the Bayesian inferential paradigm. Tonellato's model assumes isotropy. This was avoided with the Kalman filter proposed in Wikle and Cressie (1999).

When data sets are sparse in space, multivariate time series models are indeed most flexible to provide time-forward predictions taking into account spatio-temporal dependencies. We introduce in this paper a modeling approach which avoids making any spatial stationarity assumptions. We use vector autoregressive (VAR) models, and propose a model building strategy which borrows its simplicity from the widely applied model selection methods used for autoregressive (AR) models for univariate time series (Box and Jenkins, 1976). For AR models, the selection problem is crucially simplified because of the natural nestedness of the models: time lags for the time series are introduced in the model sequentially and only $p+1$ models need to be visited, where p is the maximum time lag. In a general vector autoregressive model, this nesting does not exist, but can be retrieved by considering a space-time hierarchy which is often almost as natural as the purely temporal hierarchy.

The article is set up as follows. Section 2 describes the vector autoregressive models with a spatial structure. The theory related to the inference on the parameters is available in the literature for general VAR models. The necessary stationarity assumptions are described. Crucially, spatial stationarity is not a necessary assumption and a different model can be built for each station. Temporal trends present in the data can be estimated and/or removed in a classical manner. The model building strategy that we then introduce is the main originality of this paper, and is essential since it makes the VAR family of models relatively simple and natural to use for the environmental applications of interest. Our modeling approach is also applied on the Venician carbon monoxide data in Section 2. In Section 3, we perform a correlation analysis on the Irish wind data set which allows us to investigate isotropy and spatio-temporal correlation symmetry assumptions. Our analysis indicates that the latter assumption is not fulfilled for this application. We also show the flexibility of our models compared to the direct modeling of the space-time correlation function. Both modeling approaches are useful when used on the appropriate type of data. In Section 4 we focus on predictive performance and use an out-of-sample validation technique to compare different modeling strategies, both on the Venician carbon monoxide and the Irish wind data. The article is concluded with a discussion in Section 5.

2 Vector autoregressive models with spatial structure

The models developed in this section are specifically designed for the analysis of spatio-temporal data sets with the purpose of providing time-forward predictions at given spatial locations. Our aim is to provide predictions based on a minimum of assumptions.

2.1 Time-stationary models

We assume that observations $z(\mathbf{s}_i, t)$ are made at $\mathbf{s}_i, i = 1, \ldots, N$ locations (stations) and $t = 1, \ldots, T$ times. A predictive model for $\boldsymbol{z}_t = (z(\mathbf{s}_1, t), \ldots, z(\mathbf{s}_N, t))'$ is

$$\boldsymbol{z}_t - \boldsymbol{\beta} = \sum_{i=1}^{p} R_i(\boldsymbol{z}_{t-i} - \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{\beta} = (\beta(\mathbf{s}_1), \ldots, \beta(\mathbf{s}_N))'$ is a vector of spatial effects (spatial trend), R_i, $i = 1, \ldots, p$, are unknown $N \times N$ parameter matrices, and $\boldsymbol{\varepsilon}_t$ is an N-dimensional white noise or innovation process, that is, $\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Sigma_{\boldsymbol{\varepsilon}}$, and $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_u') = \mathbf{0}$ for $u \neq t$. This model is appropriate once temporal trends have been removed, see the discussion in the next section.

The model (1) is a vector autoregressive (VAR) model commonly used in multivariate time series analysis (e.g., Lütkepohl, 1991; Peña, Tiao, and Tsay, 2001). When $\boldsymbol{z}_t$ is univariate (only one time series is observed) then (1) is simply called an autoregressive (AR) model. The deterministic dynamic system defined by (1) where the error term $\boldsymbol{\varepsilon}_t$ is dropped −also called skeleton of the model−, is said to be stable if its reverse characteristic polynomial has no roots in and on the complex unit circle, i.e.

$$\det(I - R_1 x - \ldots - R_p x^p) \neq 0, \qquad \text{for } x \in \mathbb{C},\ |x| \leq 1. \tag{2}$$

The stability property ensures that the iteration of the dynamic system yields a converging path towards a constant. Stability is an important property since it implies time-stationarity of the stochastic process (e.g., Lütkepohl, 1991; Peña et al., 2001), i.e., $\mathrm{E}(\boldsymbol{z}_t) = \boldsymbol{\mu}$, for all t, and $\mathrm{Cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-\tau}) = \Gamma_{\mathbf{z}}(\tau)$, i.e. is a function of τ only, for all t and $\tau = 0, 1, 2, \ldots$. The covariance matrix $\Gamma_{\mathbf{z}}(\tau)$ can then be computed from the parameter matrices $R_1, \ldots, R_p$ and $\Sigma_{\boldsymbol{\varepsilon}}$. For instance, when $p = 1$, we have $\mathrm{vec}(\Gamma_{\mathbf{z}}(0)) = (I_{N^2} - R_1 \otimes R_1)^{-1}\mathrm{vec}(\Sigma_{\boldsymbol{\varepsilon}})$ and $\Gamma_{\mathbf{z}}(\tau) = R_1^{\tau}\Gamma_{\mathbf{z}}(0)$, where $\mathrm{vec}(\cdot)$ denotes the operator vectorizing a matrix.

Estimation of the parameters in model (1) can be carried out with maximum likelihood (if distributional assumptions are made), with least squares or with moments estimators (Yule-Walker type). The theory related to such estimators is exposed in Lütkepohl (1991, Chap.

3) and Peña et al. (2001, Chap. 14), therefore it is not reproduced here. It is, however, worth mentioning that robust estimation of the parameters may be obtained by using robust estimators of moments (Ma and Genton, 2000) together with Yule-Walker estimating equations as it was proposed in de Luna and Genton (2002). Estimation of the parameters can be carried out for all stations simultaneously or station-wise in an equivalent manner. However, model building must be performed separately for each station, see Section 2.3.

An important property of the presented models is that they do not assume any spatial stationarity, for instance that spatial correlations depend only on the distance between stations (isotropy). Such assumptions have often been made in the spatio-temporal literature, for instance to develop space-time ARMAX models (Pfeifer and Deutsch, 1980; Stoffer, 1986). For the applications of interest in this paper, spatial-stationarity is an over-restrictive assumption which is, moreover, difficult to check in practice.

2.2 Spatio-temporal trends

In the univariate time series literature two types of trends are usually considered, deterministic trends −typically functions of time− and stochastic trends, see, e.g., Dagum and Dagum (1988) and references therein. Deterministic trends can often be removed by differencing with ∇^d, the classical time series difference operator of order d. For instance, we have $\nabla z(\mathbf{s}_i, t) = z(\mathbf{s}_i, t) - z(\mathbf{s}_i, t-1), \nabla^2 z(\mathbf{s}_i, t) = (z(\mathbf{s}_i, t) - z(\mathbf{s}_i, t-1)) - (z(\mathbf{s}_i, t-1) - z(\mathbf{s}_i, t-2))$, and so on. Model (1) can, therefore, be used after very general types of spatio-temporal trends have been removed. For instance, a natural deterministic trend specification is the following:

$$z(\mathbf{s}, t) = \mu + g(\mathbf{s}, t) + y(\mathbf{s}, t),$$

where $y(\mathbf{s}, t)$ is a time stationary process. Let us now take differences in time:

$$\nabla z(\mathbf{s}, t) = g(\mathbf{s}, t) - g(\mathbf{s}, t-1) + \nabla y(\mathbf{s}, t)$$

where $\nabla y(\mathbf{s}, t)$ is a time stationary process by definition. Thus, $\nabla z(\mathbf{s}, t)$ can be modeled by (1) if $\beta(\mathbf{s}) = g(\mathbf{s}, t) - g(\mathbf{s}, t-1)$ is a function of $\mathbf{s}$ only. This, we believe, will happen most often in practice, at least approximately. Indeed, it will happen as soon as $g(\mathbf{s}, t)$ is a polynomial function in t with coefficients possibly function of $\mathbf{s}$. A simple example, for which a difference of order one will suffice is: $g(\mathbf{s}, t) = \gamma_1(\mathbf{s}) + \gamma_2(\mathbf{s})t$. In other words, differencing eliminates deterministic polynomial time trends interacting with a spatial trend. Any spatio-temporal trend $g(\mathbf{s}, t)$ which is well approximated by a polynomial function in t can be, at least approximately, handled by time differencing. Moreover, the process $y(\mathbf{s}, t)$ above could be thought of as having a stochastic trend, for instance, $z(\mathbf{s}, t)$ may be an integrated (in time) process.

In the spatial dimension, differencing has also been suggested in order to remove trends, see e.g. the intrinsic random functions of order k, IRF-k, proposed by Matheron (1973). However, they can be used only if the spatial stations are located on a regular grid, which is seldom the case in practice.

Periodic time trends or cycles may also be tackled by taking differences. For instance, observations taken monthly will typically have to be stationarized with the seasonal difference operator $\nabla_{12} z(\mathbf{s}, t) = z(\mathbf{s}, t) - z(\mathbf{s}, t - 12)$. When other variables are observed at the same locations and times, these may also be used to model trends by regressing on them.

Another popular approach consists in modeling a deterministic trend with a weighted sum of known basis functions, where the weights are typically estimated by regression. For example, periodic functions can be used to account for seasonal effects along the time axis and polynomials can be used to model smooth variations in space. Further discussions can be found in the review article by Kyriakidis and Journel (1999).

2.3 Model building and checking

Although the model presented in Section 2.1 is essentially a VAR model, the spatial structure of the data is informative and, in particular, the parameter matrices R_i's will typically have a specific structure.

We, therefore, introduce a model building strategy to identify zeroes in the R_i's matrices of model (1). The N rows of these matrices correspond to the N locations at which time series are observed. These N stations must be modeled separately if no spatial-stationarity assumption is to be made. For each station $\mathbf{s}$, we have a covariate selection problem, where, for the response $z(\mathbf{s}, t)$, the available predictors are the time-lagged values at all stations, i.e. $z(\mathbf{s}_i, t - j)$, $i = 1, \ldots, N$ and $j = 1, \ldots$. This would be a complex model selection problem if all lagged values at all stations had to be considered as potential predictors, since then the number of potential models would rapidly explode with the number of stations and the time frequency. However, by taking into account the spatio-temporal structure, it becomes possible to define an ordering with which to sequentially introduce the predictors in the model for $z(\mathbf{s}, t)$. Such an ordering is used in univariate time series models where for explaining x_t (a variable observed at time t) the lag one variable x_{t-1} is considered first, then the lag two x_{t-2}, and so forth.

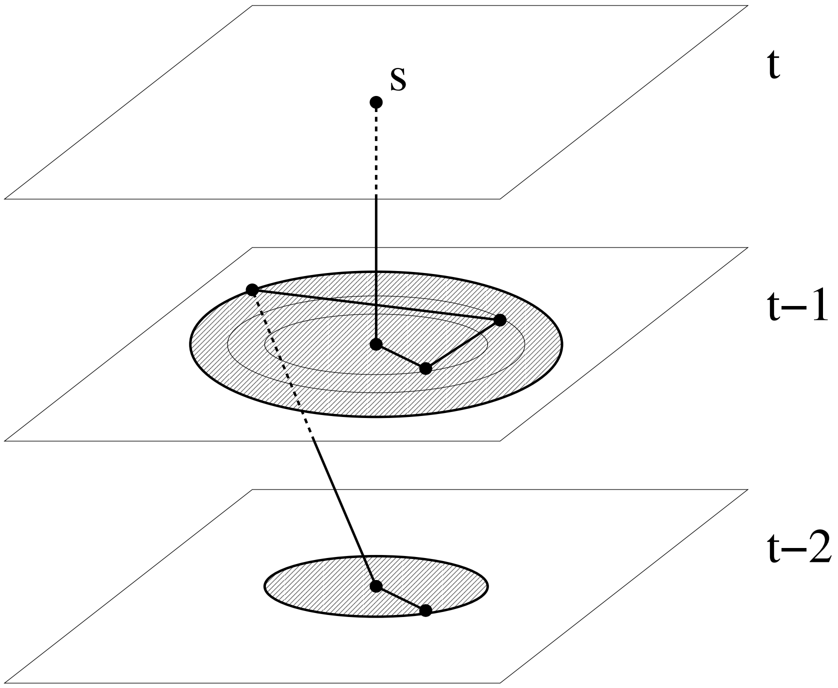

Figure 1: A schematic representation of the ordering of the stations (3).

In the spatio-temporal context we propose to use a natural ordering defined as follows. To explain $z(\mathbf{s}, t)$, consider predictors in the following order

$$z(\mathbf{s}, t-1), z(\mathbf{s}(1), t-1), z(\mathbf{s}(2), t-1), \ldots, z(\mathbf{s}(N-1), t-1), z(\mathbf{s}, t-2), z(\mathbf{s}(1), t-2), \ldots \quad (3)$$

where $\mathbf{s}(1), \mathbf{s}(2), \ldots, \mathbf{s}(N-1)$ is an ordering of the stations, for instance, in ascending order with respect to their distance (using a given metric) to $\mathbf{s}$. This ordering of the stations, and thereby the covariates, is most clearly explained graphically, see Figure 1. Other orderings can be considered as well, for instance orderings motivated by physical knowledge about the underlying process, see the Irish wind speed application in Section 3.

The fact that predictors can be entered in the model sequentially simplifies the model building stage and several strategies may be followed to know how many of these predictors should be used. A popular technique in time series modeling is to look at partial autocorrelations. For the spatio-temporal models under consideration and with a given ordering of the predictors we can straightforwardly generalize this time series technique by looking at partial correlation along the ordering of predictors. For three random variables x, z and $\boldsymbol{y}$, the latter possibly vector valued, the partial correlation of x and z given $\boldsymbol{y}$ is defined as

$$\mathrm{Corr}(x, z|\boldsymbol{y}) = \mathrm{Corr}(x - P(x|\boldsymbol{y}), z - P(z|\boldsymbol{y})),$$

where $\text{Corr}(x, z)$ denotes the correlation between x and z, and $P(x|\boldsymbol{y})$ is the best linear predictor of x given $\boldsymbol{y}$. This partial correlation has the property that it is equal to zero when x and z are independent conditional on $\boldsymbol{y}$.

Renaming the sequence (3) as follows

$$\begin{aligned} x_1 &= z(\mathbf{s}, t-1), x_2 = z(\mathbf{s}(1), t-1), \ldots, x_N = z(\mathbf{s}(N-1), t-1), \\ x_{N+1} &= z(\mathbf{s}, t-2), x_{N+2} = z(\mathbf{s}(1), t-2), \ldots, x_{2N} = z(\mathbf{s}(N-1), t-2), \\ & \ldots, \end{aligned}$$

we can define a partial correlation function (PCF) for station $\mathbf{s}$ as

$$\rho_{\mathbf{s}}(h) = \text{Corr}(z(\mathbf{s}, t), x_h | x_1, \ldots, x_{h-1}).$$

The usefulness of the PCF for model selection is now clarified. Let us define h_1 to be such that $\rho_{\mathbf{s}}(h_1) \neq 0$ and $\rho_{\mathbf{s}}(h) = 0$ for $h_1 < h \leq N$. Similarly, a value h_i can be defined for each time lag i, such that $\rho_{\mathbf{s}}(h_i) \neq 0$ and $\rho_{\mathbf{s}}(h) = 0$ for $h_i < h \leq iN$. The orders h_i's can be identified by looking at the sample partial correlation function $\hat{\rho}_{\mathbf{s}}(h) = \widehat{\text{Corr}}(z(\mathbf{s}, t) - \hat{P}(z(\mathbf{s}, t)|x_1, \ldots, x_{h-1}), x_h - \hat{P}(x_h|x_1, \ldots, x_{h-1}))$. An approximate test for $\rho_{\mathbf{s}}(h) = 0$ is obtained by noting that, under joint normality of the variables, $\hat{\rho}_{\mathbf{s}}(h)\sqrt{(n-h)/(1-\hat{\rho}_{\mathbf{s}}(h)^2)}$ is t-distributed with $n-h$ degrees of freedom, where n denotes the sample size utilized; see, e.g., Krzanowski (1988, Sec. 14.4). The normality assumption is fairly natural when linear models are utilized. This test statistics can be used to derive confidence intervals for $\rho_{\mathbf{s}}(h) = 0$, see Section 2.4. We propose the following strategy.

Identification strategy

Step 0: Choose one of the observed sites, $\mathbf{s}$.

Step 1: Identify h_1 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = 1, \ldots, N$.

Step 2: Identify h_2 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = N+1, \ldots, 2N$ when $x_{h_1+1}, \ldots, x_N$ have been put to zero since they have been shown to be unhelpful in explaining $z(\mathbf{s}, t)$ in the previous step.

Step 3: Identify h_3 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = 2N+1, \ldots, 3N$ when $x_{h_1+1}, \ldots, x_N$ and $x_{h_2+1}, \ldots, x_{2N}$ have been put to zero since they have been shown to be unhelpful in explaining $z(\mathbf{s}, t)$ in the previous steps.

Step 4: Step 3 is repeated in a similar manner for all necessary time lags in order to identify h_4, h_5, etc.

Step 5: Repeat the previous steps for all observed sites.

The deletion of uninteresting predictors at each time lag improves on the efficiency by avoiding the estimation of zero coefficients. However, this procedure does not avoid the estimation of all zero coefficients, since, for instance, the predictors included at later stages may put to zero coefficients that were not so when considering partial models. This problematic is, however, also present in the identification of the order of linear AR models. For such models, all parameters are estimated up to a given time-lag by convention. This is convenient since it avoids chasing zero coefficients by using a sequence of t-tests on the coefficients.

An alternative to the use of the PCF is the use of an automatic model selection criterion in each step of the identification strategy. AIC (Akaike information criterion, Akaike, 1974) or BIC (Bayesian information criterion, Schwarz, 1978) may be used, the former being usually preferred for predictive purposes.

Finally, model checking is commonly done by looking at residuals for signs of deviation from model assumptions. A feature that can be controlled for is, for instance, whether they are correlated in time.

2.4 Carbone monoxide in Venice

The data set available has $T = 300$ hourly observations of atmospheric concentration (micrograms per cubic meter) of carbon monoxide (CO) recorded in September 1995 at $N = 4$ monitoring stations located in Mestre (Venice, Italy). This data set has been analyzed by Tonellato (2001) in a Bayesian dynamic linear model framework.

Our methodology is particularly well suited for this application for several reasons. First, Italian law requires public authorities to produce short-term forecasts of air pollutant concentrations at locations where monitoring stations are present, that is, spatial interpolation is not of prime interest. Second, our model does not make spatial stationarity assumptions by modeling each station separately, in contrast with Tonellato (2001) who used a spatial stationary isotropic exponential correlation function. Such assumptions are arbitrary because with only four stations it is not possible to assess their validity. Moreover, wind speed and direction can be expected to influence air pollutant concentrations in a nonstationary and anisotropic way.

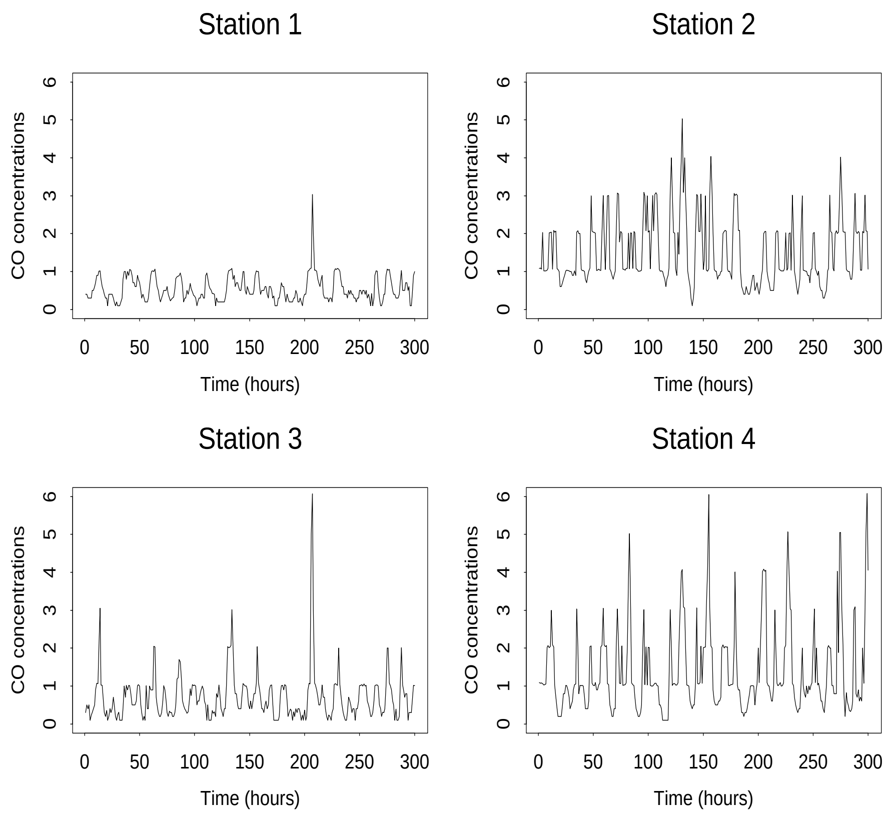

Figure 2: The time series of CO concentrations at the 4 stations.

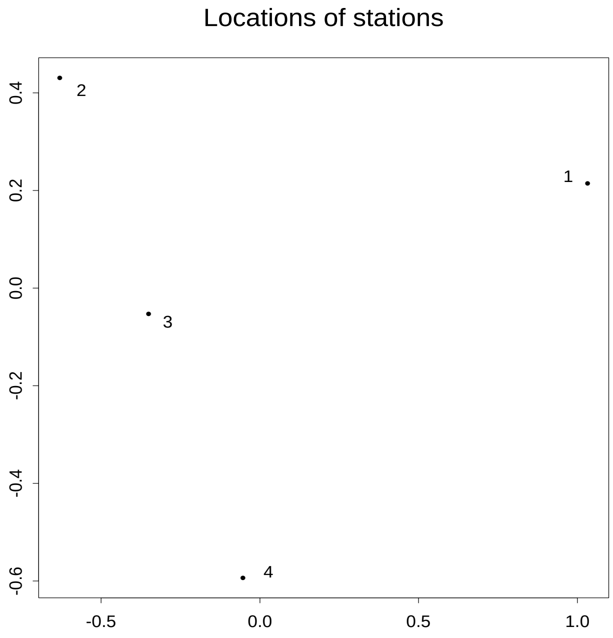

Figure 3: The locations of the 4 stations (distance in kilometers).

Table 1: Univariate autoregressive models identified with BIC for the time series observed on each station. The maximum lag allowed for with BIC was 30. Parameters are estimated with least squares. The models are written in matrix form as a VAR(2) model to facilitate the comparison with the VAR model with spatial structure identified in Table 2. The residual correlation matrix assumes that the time series are independent of each other.

$\hat{R}_1$	$\hat{R}_2$	$\hat{\Sigma}_\varepsilon$
$\begin{pmatrix} 0.64 & 0 & 0 & 0 \\ 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0.48 & 0 \\ 0 & 0 & 0 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.19 & 0 \\ 0 & 0 & 0 & 0.17 \end{pmatrix}$	$\begin{pmatrix} 0.13 & 0 & 0 & 0 \\ 0 & 0.11 & 0 & 0 \\ 0 & 0 & 0.22 & 0 \\ 0 & 0 & 0 & 0.13 \end{pmatrix}$

Table 2: The VAR(2) model identified for the CO data with BIC. The maximum temporal lag was fixed at six because of the results in Table 1. Parameters in model (1) are estimated with least squares. The residuals covariance matrix is the classical variance estimator.

$\hat{R}_1$	$\hat{R}_2$	$\hat{\Sigma}_\varepsilon$
$\begin{pmatrix} 0.59 & 0 & 0 & 0.16 \\ 0 & 0.67 & 0 & 0 \\ 0.14 & 0.15 & 0.32 & 0.25 \\ 0 & 0 & 0 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.17 \end{pmatrix}$	$\begin{pmatrix} 0.13 & 0.01 & 0.03 & 0.03 \\ 0.01 & 0.11 & 0.03 & 0.01 \\ 0.03 & 0.03 & 0.20 & 0.04 \\ 0.03 & 0.01 & 0.04 & 0.13 \end{pmatrix}$

There are a few missing values in the data set. Following Tonellato (2001), they are replaced by the mean of the values at the same station and hour over the sample period. The data is plotted in Figure 2 according to the four different locations as depicted in Figure 3. We take the logarithm of the observations, thereby stabilizing the variance. Stations 2 and 4 are located along streets with high intensity of traffic, whereas station 1 is located in a garden and station 3 in a pedestrian area. Therefore, there are differences in CO levels among the stations. This spatial trend is readily included in our models since each station is allowed to have its own level, $\boldsymbol{\beta}$ in model (1). Time trends must, on the other hand, be accounted for previous to fitting the time-stationary model (1). We follow Tonellato (2001) and estimate a trend for each station by regressing on a family of daily harmonics. We then subtract these temporal trends from the data.

As a benchmark we perform first an analysis ignoring the spatial structure of the data set.

That is, the four time series are analyzed separately, by fitting an autoregressive model to each of them. Using the Bayesian information criterion we obtain the models described in Table 1.

The univariate time series approach ignores the spatial dependence structure. We, therefore, apply the model building strategy described in the previous section, where the station ordering (3) is defined by considering ascending distances, see Figure 3. We start by looking at the partial correlations to explore the existing dependences. Plots of these correlations are given in Figure 4 for station 3 as illustration, as a function of the ordered stations, along with 95% confidence intervals for $\rho_{\mathbf{s}}(h) = 0$ (horizontal lines). We see that all the partial correlation at time lag one are significant except for station 2, and that for time lag two, only the partial correlation corresponding to station 2 is strongly significant. The partial correlations at time lag three are not significant.

The model selection may be performed automatically and we use the algorithm described in the previous section associated to the BIC criterion. The models obtained are described in Table 2. For instance, for station 3 the model identified corresponds to the partial correlation pattern observed in Figure 4 at time $t-1$ and $t-3$, but differs slightly at time $t-2$.

Tables 1 and 2 can be compared directly. The model selection strategy we use identifies spatio-temporal dependencies as relevant for time-forward forecasting since the coefficient matrices R_i's are not diagonal. Note, however, that for stations 2 and 4 a univariate time series model is selected. The latter are located along streets with high intensity traffic. Results in Table 2 are in this respect reasonable since the stations with high level of pollution are influential for those with low levels. Identification strategies are further evaluated in Section 4.1

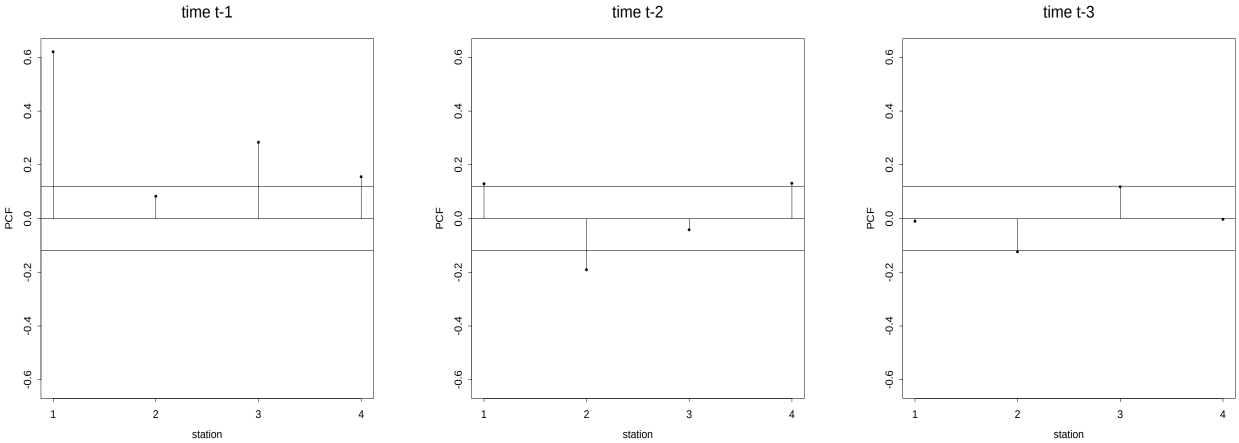

Figure 4: The PCF for the station 3 up to time $t-3$.

3 Spatio-temporal correlation analysis

In this section, we aim at showing the flexibility of our approach in modeling the correlation structure of the Irish wind data set of Haslett and Raftery (1989). The data consist of daily averages of wind speeds recorded at $N = 11$ synoptic meteorological stations in Ireland between 1961 and 1978. A map of the locations of the stations can be found in Haslett and Raftery (1989). Predictions of wind speeds may be useful for various reasons. An example is the production of renewable energies, see Alexiadis, Dokopoulos, and Sahsamanoglou (1999).

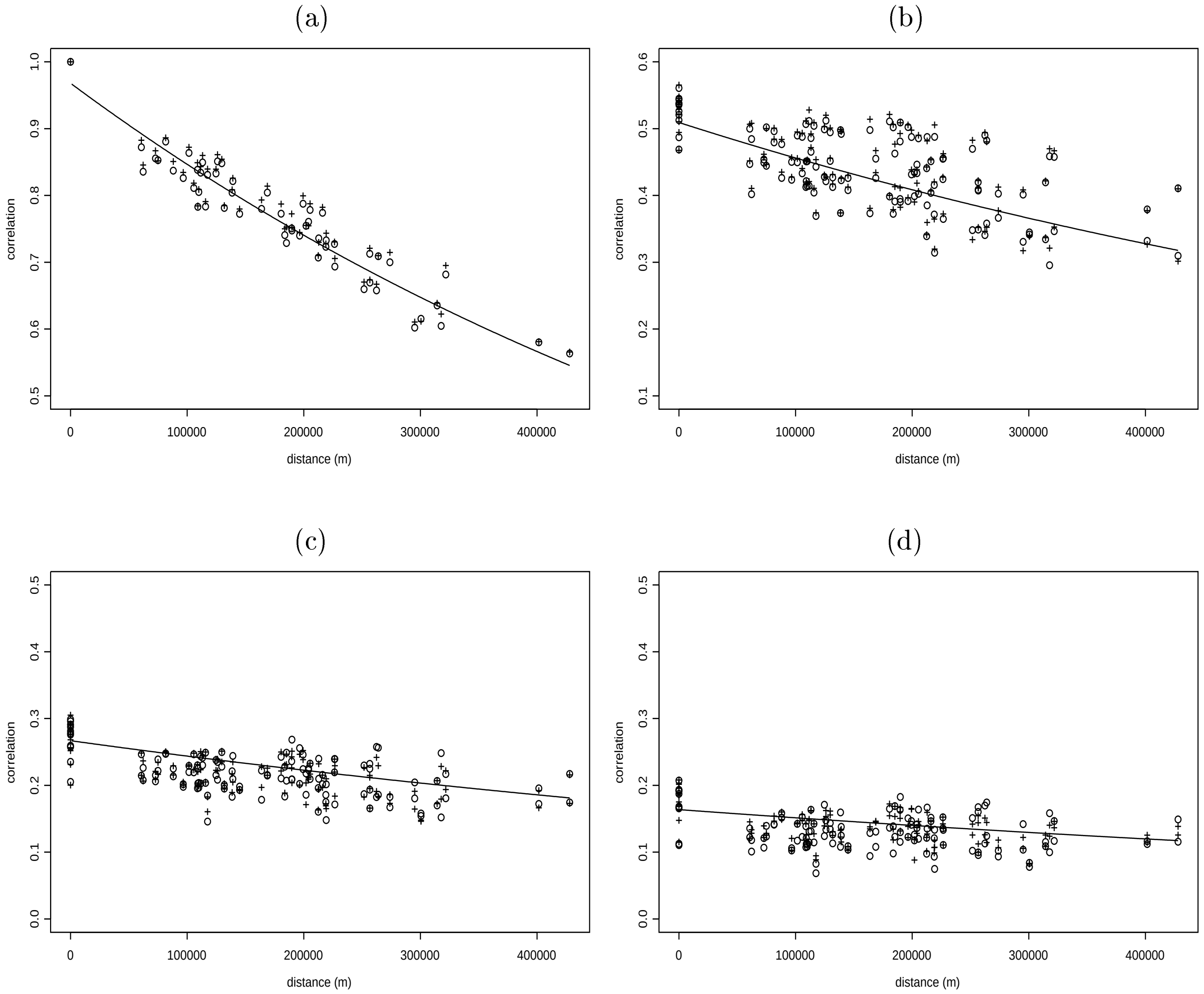

Figure 5: Spatial correlation as a function of distance (in meters) for the Irish wind dataset (open circles: from fitted VAR(3); pluses: empirical) at various temporal lags: (a) $\tau = 0$; (b) $\tau = 1$; (c) $\tau = 2$; (d) $\tau = 3$. The solid curve is the stationary correlation function fitted by Gneiting (2002).

Following Haslett and Raftery (1989) and Gneiting (2002), we take a square root transformation to stabilize the variance over stations and time periods, and subtract an estimated seasonal effect and spatially varying mean from the data. The data set initially included one additional station, but it was omitted because its spatial correlation structure with respect to the other stations was not consistent with the stationarity assumptions (see Haslett and Raftery, 1989). Our analysis below indicates that some spatial stationarity assumptions are still being violated by the remaining stations.

Gneiting (2002) provided evidence that the assumption of full symmetry of the spatio-temporal correlation structure is not realistic, because wind patterns are predominantly westerly over Ireland. We incorporate this physical information about general wind directions in our model by defining a special ordering of the stations. Specifically, for each station $\mathbf{s}$, we define an ordering of the stations to the west of $\mathbf{s}$ in ascending order of distances, followed by the stations to the east of $\mathbf{s}$. This ordering reflects the pattern of Irish winds to blow mainly from west to east. Note that this ordering does not imply any assumption of stationarity or even isotropy, but helps in improving the selection of a model. An inappropriate ordering would only imply that fewer coefficients in the matrices R_i of model (1) would be identified as zeros, thereby implying larger prediction errors (see Section 4.2).

Next, we fit our model (1) to the modified data above. We use the BIC model selection criterion to identify a model for each station separately. The maximum temporal lag allowed for was three to make our results comparable to those in Gneiting (2002), and because univariate AR analyses of the different time series confirmed that higher temporal lags were not necessary. The resulting model is a VAR(3), where each matrix R_1, R_2, and R_3 is of dimension 11×11. The covariance matrix $\Sigma_{\boldsymbol{\varepsilon}}$ is estimated from the residuals of the model and yields estimates of the matrices $\Gamma_{\mathbf{z}}(\tau)$, $\tau = 0, 1, 2, 3$, from straightforward formulae, see Lütkepohl (1991, p. 23-24). The estimated spatial correlation matrices corresponding to $\Gamma_{\mathbf{z}}(\tau)$ from our model are plotted in Figure 5 as open circles for $\tau = 0, 1, 2, 3$. The empirical spatial correlations are plotted as pluses along with the fitted stationary correlation functions (solid curve) proposed by Gneiting (2002). The estimated correlation from our model (open circles) are very close to the empirical correlations (pluses), indicating that the spatial VAR(3) model can capture the correlation structure well. Although the fitted stationary correlation function summarizes the correlation structure rather well at the temporal lag $\tau = 0$ (Figure 5, panel (a)), this is less so at higher lags (Figure 5, panel (b)-(d)).

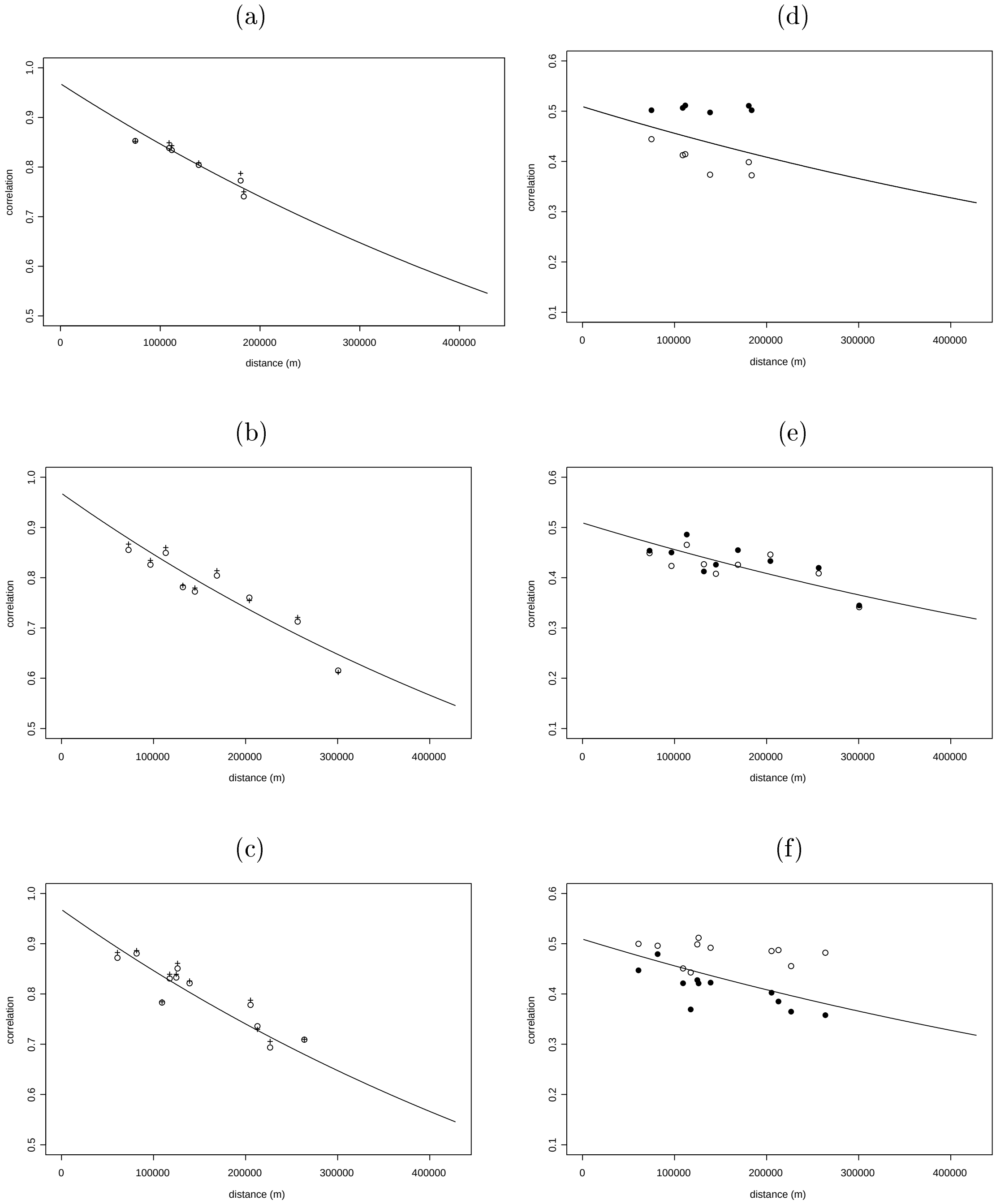

Figure 6: Panels (a)-(c): Spatial correlation (open circles: fitted VAR(3); pluses: empirical) as a function of distance (in meters) for the Irish wind dataset at temporal lag $\tau = 0$, for the directions WE (a), NS (b), and SW-NE (c). Panels (d)-(f): Spatial correlation from the fitted VAR(3) model at temporal lag $\tau = 1$, for the directions WE, NS, and SW-NE (closed circles in panels (d), (e) and (f) respectively) and EW, SN, and NE-SW directions (open circles in panels (d), (e) and (f) respectively). In the right-hand-side panels empirical correlations are omitted for readability. The solid curve is the stationary correlation function fitted by Gneiting (2002).

We investigate next the assumptions of isotropy and full symmetry of the correlation structure of the Irish wind data. For this purpose, we consider three main directions: horizontal west-east (WE), vertical north-south (NS), and diagonal (SW-NE). The panels (a)-(c) in Figure 6 depict the spatial correlation at temporal lag $\tau = 0$ in the WE, NS, and SW-NE directions respectively. Open circles indicate the correlations from our fitted model VAR(3), pluses indicate empirical correlations, and the solid curve is Gneiting's (2002) fit. We can see that the fitted correlation function (solid line) represents the correlation structure rather well in each of the three directions. Thus, the assumption of isotropy at temporal lag $\tau = 0$ seems adequate.

The panels (d)-(f) in Figure 6 depict the spatial correlation from our fitted VAR(3) model at temporal lag $\tau = 1$ in the WE, NS, and SW-NE directions (closed circles) and EW, SN, and NE-SW directions (open circles) respectively. We can see that although the correlations are rather symmetric in the NS and SN direction (panel (e)), there are strong asymmetries in the WE and EW direction and in the SW-NE and NE-SW direction. This asymmetric behavior, which is caught by the VAR(3) specification, cannot be modeled by a single correlation function (solid line).

4 Prediction performance analysis

In Section 2 we have proposed a strategy to identify a spatio-temporal model. Model selection strategies are seldom unique and it is necessary to evaluate them. The availability of different model selection strategies may be due to the availability of different classes of models. Even for a given family of models, different model strategies may be available, which are optimal under different circumstances; see de Luna and Skouras (2003). In the latter paper, a framework was advocated to compare model selection strategies through out-of-sample validation based on recursive prediction errors. This can be adapted to the spatio-temporal data studied herein. For a given station located at $\boldsymbol{s}$, a model selection strategy can be evaluated with the accumulated prediction error criterion

$$\sum_{t=M}^{T} (z(\boldsymbol{s},t) - \hat{z}^{t-1}(\boldsymbol{s},t))^2, \tag{4}$$

where $\hat{z}^{t-1}(\boldsymbol{s},t)$ is the prediction of $z(\boldsymbol{s},t)$ obtained by applying the model selection strategy on the sub-sample $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{t-1}$. That is, a model is chosen and fitted to all sub-samples $t = M, \ldots, T$. The prediction errors $z(\boldsymbol{s},t) - \hat{z}^{t-1}(\boldsymbol{s},t)$ are also called recursive residuals. Other loss functions than the squared prediction error may be used. The criterion (4) can be computed

Table 3: Root accumulated mean squared errors (square root of (4) with $M = 50$) for four strategies: AR models chosen with BIC and AIC (AR-BIC and AR-AIC respectively) and spatio-temporal autoregressions chosen with BIC and AIC (SVAR-BIC and SVAR-AIC respectively). The maximum temporal lag allowed for was six. The smallest root accumulated mean squared errors per station are represented in boldface.

strategy	AR-BIC	SVAR-BIC	AR-AIC	SVAR-AIC
Station 1	**0.371**	0.380	**0.371**	0.382
Station 3	0.480	**0.466**	0.480	0.494

for several model selection strategies, possibly based on different model sets. In de Luna and Skouras (2003, Theorem 1) it was shown that choosing the model strategy minimizing (4) would eventually identify the best strategy with probability one.

We now apply this out-of-sample framework to the two data sets studied earlier in this paper. This allows us to evaluate model selection strategies associated to the VAR models and to compare them with a benchmark univariate time series modeling.

4.1 Carbone monoxide in Venice

The carbon monoxide data was described in Section 2.4, where we used BIC to identify a univariate time series model and a spatio-temporal model. These models need to be evaluated in order to decide which is most appropriate. We evaluate them here together with two other model selection strategies. We use both BIC and AIC to select a model within the AR family (univariate time series models) and the VAR family with spatial structure. Only station 1 and 3 are considered because the models in Table 1 and 2 were equivalent for station 2 and 4.

From Table 3, we note that BIC provides the best prediction performance. Moreover, for station 3, the univariate time series model was outperformed by the model with spatial structure found with BIC. On the other hand, for station 1 the univariate time series model has lowest accumulated prediction error indicating that the spatial structure found with BIC may be superfluous for predictive purposes.

4.2 Irish winds

The Irish wind data set was described in Section 3, where stationarity hypotheses were investigated. The spatial structure (ordering of the stations) of the VAR models used was defined by taking into account that winds are predominantly westerly over Ireland. The spatial ordering of the stations is part of the model specification and can also be investigated with a prediction performance analysis. We, therefore, consider here six model identification strategies based on different families of models. As in the previous section two strategies based on univariate time series models: AR-BIC and AR-AIC. Moreover, AIC and BIC are used to select between two families of VAR models having different spatial structure: the first family uses an ordering of stations (3) defined solely on ascending distances between stations (SVAR-BIC and SVAR-AIC with dist. order), while the second family is based on an ordering which takes into account the information on wind directions as described in Section 3 (SVAR-BIC and SVAR-AIC with wind order).

The root accumulated mean squared errors (square root of (4)) are presented in Table 4 for the eleven stations and the six model selection strategies. Two main general comments can be given. Firstly, the spatio-temporal models outperform the univariate time series models. Moreover, the differences in prediction performances between the different spatio-temporal modeling strategies are much less important than the differences observed between the spatio-temporal modeling and the univariate time series strategies. Note, however, that AIC performs better than BIC in many cases and that the models whose spatial structure takes into account the predominant wind direction (wind order) has as good or better prediction performance than the models ignoring this physical information (dist order) in all but one case (station 3).

The out-of-sample prediction performances summarized in Table 4 may be analyzed in more detail by displaying the accumulated prediction errors graphically. More precisely, for two strategies S_1 and S_2 it is informative to plot the sums

$$\sum_{t=M}^{i} \left((z(\boldsymbol{s},t) - \hat{z}^{t-1}(\boldsymbol{s},t))^2 - (z(\boldsymbol{s},t) - \tilde{z}^{t-1}(\boldsymbol{s},t))^2 \right) \tag{5}$$

against $i = M, \ldots, T$, where $\hat{z}^{t-1}$ and $\tilde{z}^{t-1}$ are obtained with S_1 and S_2, respectively. Such cumulative sums plots allows us to understand whether a strategy which has best root accumulated mean squared error is really performing better recursively through time. This is illustrated in Figure 7 where for station 3 (Kilkenny), the best strategy SVAR-AIC(dist order) for this station is compared with the other strategies. These comparisons provide strong evi-

Table 4: Root accumulated mean squared errors (square root of (4) with $M = 1,000$) for six strategies: AR models chosen with BIC and AIC (AR-BIC and AR-AIC respectively) and spatio-temporal autoregressions chosen with BIC and AIC (SVAR-BIC and SVAR-AIC respectively). For the latter, two different ordering of the stations are examined. One based solely on distances between stations (dist order) and the other taking also into account the dominating wind direction over Ireland (wind order). The maximum temporal lag allowed for was three, see Section 3. The smallest root accumulated mean squared errors per station are represented in boldface.

Station	Strategy					
	AR-BIC	AR-AIC	SVAR-BIC		SVAR-AIC	
			dist order	wind order	dist order	wind order
Roche's Pt. (1)	0.492	0.492	0.474	**0.473**	**0.473**	**0.473**
Valentia (2)	0.503	0.503	**0.498**	**0.498**	0.499	0.499
Kilkenny (3)	0.446	0.446	0.424	0.424	**0.423**	0.424
Shannon (4)	0.461	0.460	0.455	0.454	**0.452**	**0.452**
Birr (5)	0.481	0.480	0.470	0.470	0.469	**0.468**
Dublin (6)	0.461	0.461	0.445	0.445	**0.444**	**0.444**
Claremorris (7)	0.488	0.488	0.485	0.485	0.483	**0.482**
Mullingar (8)	0.446	0.445	**0.433**	**0.433**	**0.433**	**0.433**
Clones (9)	0.477	0.477	**0.467**	0.468	**0.467**	**0.467**
Belmullet (10)	0.490	0.490	0.491	0.490	**0.489**	**0.489**
Malin Head (11)	0.499	0.499	0.494	**0.492**	0.493	0.493

dence for the fact that the best spatio-temporal strategy performs better than the univariate time series models. The difference in performance when comparing spatio-temporal strategies to each other is much less conclusive, indicating that the different spatio-temporal strategies provide similar prediction performances.

The patterns observed in Figure 7 could also be observed for the other stations although we do not include the graphs. Notable exceptions were that for station 2 and 10, for which univariate time series and spatio-temporal strategies had similar performances. This can be explained by the fact that these two stations are located on the west coast. Winds coming predominantly from this direction make these stations fairly independent from winds measured at stations located eastwards.

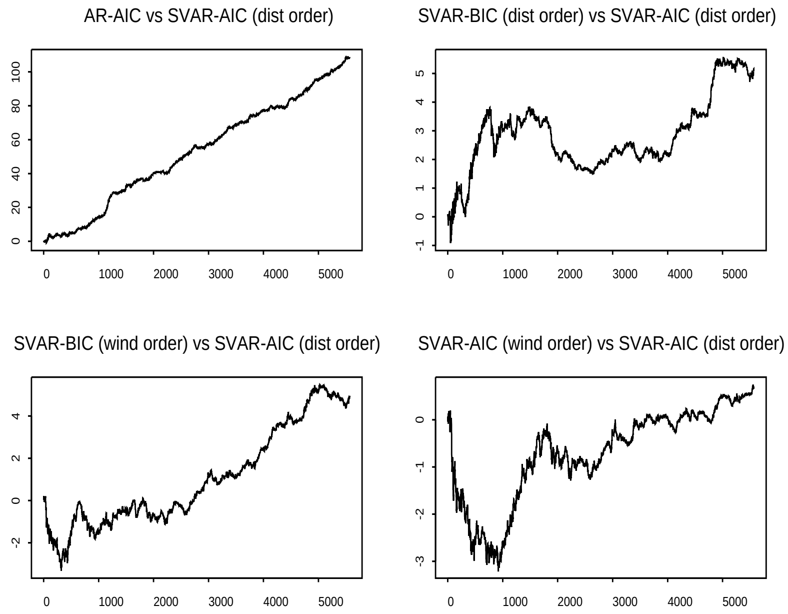

Figure 7: The graphs plot the differences in accumulated squared prediction errors (5) between the two model selection strategies (given in titles) for station 3 (Kilkenny). Upward trends mean that the second strategy is performing better while downward trends show that the first strategy is performing better.

5 Conclusion

In this paper, we have advocated the use of VAR models when the available spatio-temporal data is rich in the time dimension but sparse in the spatial dimension, and when the purpose of the analysis is to provide time-forward predictions at the spatial location where historical data is available. VAR modeling is well understood and widely applied in many disciplines. Our main contribution is to propose a model identification strategy which takes advantage of the spatial location of the different time series.

We believe that the proposed modeling strategy will find wide applicability in the environmental sciences where time forward prediction and monitoring of spatio-temporal processes is a major activity. This wide applicability should be eased by the fact that the models and identification strategy (which can be automated) are both easy to implement and general. The simplicity in implementation is a consequence of the consideration of linear models which are nested with respect to a natural spatio-temporal hierarchy. *Splus* codes are available from the authors upon request. The generality of the method is implied by treating each spatial location separately in the modeling process, thereby avoiding restrictive and often difficult-to-verify spatial-stationarity assumptions.

We have aimed at showing both the simplicity and generality of the method with two real data sets. For instance, the estimation of temporal trends has been done with harmonic functions of time. As we mentioned in Section 2.3, our modeling strategy may readily be used together with other trend estimation techniques such as those using measurements on other variables when these are available.

The modeling framework is, moreover, open to further generalizations, although at the cost of its simplicity and probably robustness. For instance, non-linear autoregressive models may be entertained. This would be much more difficult to do in the context of spatial-stationary processes. State-space representations could also be considered in a similar setting allowing us to deal optimally with missing observations at certain locations and times.

Acknowledgments

The authors thank Dennis Boos, Dongseok Choi, and Tilmann Gneiting for helpful comments.

6 References

Akaike, H. (1974), "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, **19**, 716-723.

Alexiadis, M. C., Dokopoulos, P. S., Sahsamanoglou, H. S. (1999), "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, **14**, 836-842.

Box, G.E.P., Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Dagum, C., Dagum, E.B. (1988), "Trend," in *Encyclopedia of Statistical Sciences*, S. Kotz, N.L. Johnson, eds, New York: Wiley & Sons, 321-324.

de Luna, X., Genton, M.G. (2002), "Simulation-based inference for simultaneous processes on regular lattices," *Statistics and Computing*, **12**, 125-134.

de Luna, X., Skouras, K. (2003), "Selecting a model selection strategy," *Scandinavian Journal of Statistics*, **30**, 113-128.

Gneiting, T. (2002), "Nonseparable, stationary covariance functions for space-time data," *Journal of the American Statistical Association*, **97**, 590-600.

Guttorp, P., Meiring, W., Sampson, P. (1994), "A space-time analysis of ground-level ozone data," *Environmetrics*, **5**, 241-254.

Haslett, J., Raftery, A.E. (1989), "Space-time modeling with long-memory dependence: assessing Ireland's wind power resource," *Applied Statistics*, **38**, 1-50.

Host, G., Omre, H., Switzer, P. (1995), "Spatial interpolation errors for monitoring data," *Journal of the American Statistical Association*, **90**, 853-861.

Krzanowski, W.J. (1988), *Principles of Multivariate Analysis: A User's Perspective*, Oxford: Oxford University Press.

Kyriakidis, P.C., Journel, A.G. (1999), "Geostatistical space-time models: a review," *Mathematical Geology*, **31**, 651-684.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Ma, Y., Genton, M.G. (2000), "Highly robust estimation of the autocovariance function," *Journal of Time Series Analysis*, **21**, 663-684.

Matheron, G. (1973), "The intrinsic random functions and their applications," *Advances in Applied Probability*, **5**, 439-468.

Mardia, K.V., Goodall, C.R. (1993), "Factorized models in spatial modeling," in *Multivariate Environmental Statistics*, N.K. Bose, G.P. Patil, C.R. Rao, eds., Amsterdam: North Holland.

Peña, D., Tiao, G. C., Tsay, R. S. (2001), *A Course in Time Series Analysis*, Wiley.

Pfeifer, P.E., Deutsch, S.J. (1980), "A three-stage iterative procedure for space-time modeling," *Technometrics*, **22**, 35-47.

Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, **6**, 461-464.

Stoffer, D.S. (1986), "Estimation and identification of space-time ARMAX models in the presence of missing data," *Journal of the American Statistical Association*," **81**, 762-772.

Stroud, J.R., Müller, P., Sansó, B. (2001), "Dynamic models for spatiotemporal data," *Journal of the Royal Statististical Society, Series B*, **63**, 673-689.

Tonellato, S.F. (2001), "A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations," *Applied Statistics*, **50**, 187-200.

Wikle, C.K., Cressie, N. (1999), "A dimension reduction approach to space-time Kalman filtering," *Biometrika*, **86**, 815-829.