Parsimonious Var Models For Air Pollution Dynamic Analysis

Fontanella Lara and Granturco Mariagrazia

Department of Quantitative Methods and Economic Theory, University G.D'Annunzio - Italy

Abstract: We discuss a framework to obtain temporal predictions for an evolving spatial field regularly sampled in time at arbitrary spatial locations. Difficulties caused by large data sets and the modelling of complicated spatio-temporal interactions limit the effectiveness of traditional space-time statistical models. In this study, we propose the use of a flexible approach to deal with large and small time-scale variability of the observed data. The temporal model is applied with respect to both the observed data domain and the common component domain, to achieve a dimensionality reduction.

Key words: VAR models, State-Space, Karhunen- Loève Transform, Trend Surface

INTRODUCTION

Environmental processes are usually monitored over space and time. When dynamical processes involve complicated temporal and spatial features, a simultaneous space-time analysis can provide benefits with respect to a solely temporal or spatial inspection. There are several examples of space-time models applied to environmental problems. Some of them are also based on a Bayesian dynamic linear model[1-3]. In order to obtain a parsimonious model (in parameters), our study addresses directly the problem of dimension reduction, which is a critical issue for implementing the Kalman filter on large data sets.

More precisely, considering a spatial field on a fixed region that is evolving in time and for which we have discrete observations at regular time intervals, we use a simple and parsimonious model, which accounts for the main features of the data to obtain temporal forecasts. We let *Y,* later indexed by both space and time, represents the primary variable of interest. Recognizing the presence of a measurement error component, the space-time dynamic process can be described through a state-space formulation. Thus, we assume that the observational data *Z*, which is an observation of *Y* with error, can be specified by a *measurement equation*. Subtraction of deterministic components, such as temporal trend and seasonal effects, identifies a short-time scale dynamical process *X* which is modelled through the *transition equation*.

The role of the *X*-process is to account for both spatial and temporal dynamics beyond those accounted for in long-term means and seasonal behaviour. The structure of the *transition matrix* Ξ is a key concern. In particular, the crucial issue of modelling spatial structures obviously involves a trade-off between the richness of Ξ and the level of spatial independence of the *transition noise* term. In this context, using the Karhunen- Loève transform, our objective is to obtain a

good interaction between the modelling of Ξ and the covariance matrix of the transition noise vector.

The data: Here we provide an explorative analysis of the data set used in the example. We chose the Milan district as a test bed for our analysis and statistical modelling. The data are daily maximum values of Carbon Monoxide concentrations at 25 monitoring stations. The coordinate system of the monitoring stations refers to the Italian national grid system (Gauss-Boaga) based on the Universal Transverse Mercator (UTM) projection. The map of the monitoring sites is shown in Fig. 1.
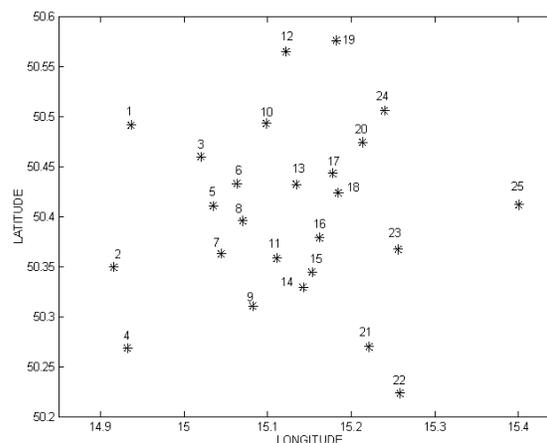
Fig. 1: Map of the Milan monitoring network. The marks correspond to the locations of the CO monitoring stations

The measurement units are micrograms of Carbon per cubic meter ($\mu g(CO)/m^3$). The period of the data covers the years 1998-2001, giving a 25×1461 data matrix. The raw data were provided by the Environmental Agency (ARPA) of Lombardia Region.

Corresponding Author: Department of Quantitative Methods and Economic Theory, University G.D'Annunzio - Italy
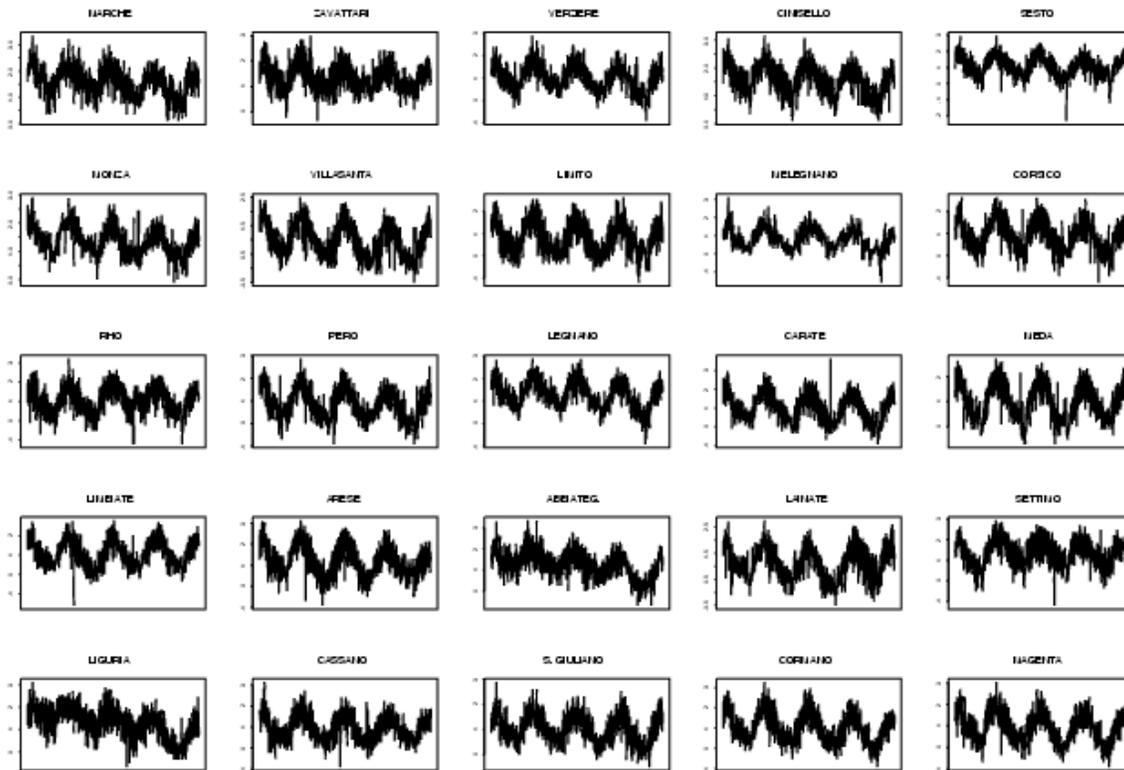
Fig. 2:   CO concentration time profiles at monitoring station locations
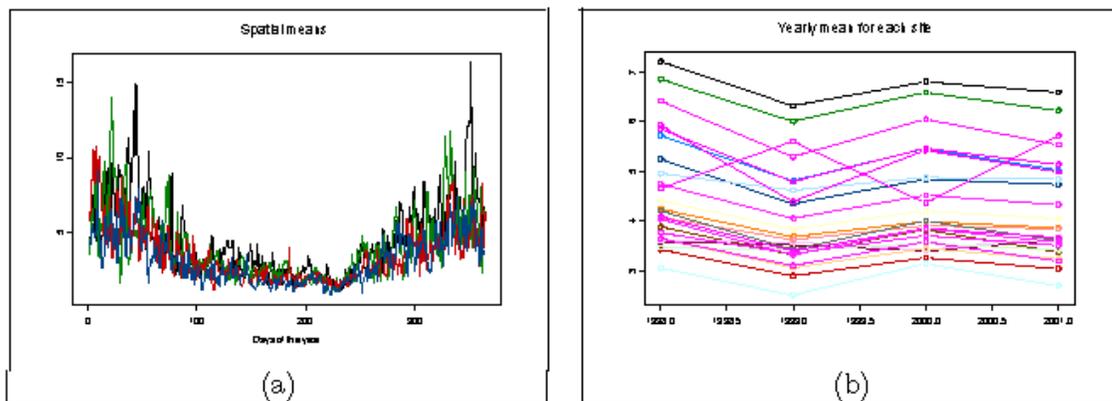
Fig. 3:   (a) Average of CO concentration over 25 sites from 1998 to 2001. (b) Average of CO concentration over days from 1998 to 2001 for 25 sites. Each line indicates one site.

To avoid the effects of a small amount of missing data, they have been estimated by using biharmonic splines[4] for each data time. Because concentration data are always positive, it is convenient to operate on a logarithmic scale, to remove the effect of heteroschedasticity. As shown in Fig. 2, the variability structure of the data exhibits a strong seasonal dependence.

Figure 3a and b are also helpful for an exploratory data analysis. Figure 3a shows the CO average concentration for years 1998-2001, where the average is taken over the 25 monitoring sites. Figure 3b shows the CO average concentration as a function of years, for the 25 monitoring sites, where the average is taken over days. Both figures indicate that the CO concentration pattern is governed by both spatial and temporal features.

The dynamic model: Consider a spatio-temporal field $Z(\mathbf{s},t)$, where $\mathbf{s}$ is a generic location within some geographic region of interest and $\{t=t_1,...,t_N\}$ indexes consecutive times at which monitoring data are collected. Let $Z(\mathbf{s},t)$ have the decomposition

$$Z(\mathbf{s},t) = M(\mathbf{s},t) + \eta(\mathbf{s},t) \qquad (1)$$

Where $M(\mathbf{s},t)$ is a deterministic component modelling a smooth long-term variability of the spatio-temporal process and $\eta(\mathbf{s},t)$ is a stationary residual term, independent of $M(\mathbf{s},t)$, modelling the space-time higher frequency fluctuations around $M(\mathbf{s},t)$. In practical applications, the Z field may be a non linear transform of the directly observable variable field. For example, for non-negative concentration, Z might be obtained from a logarithmic transformation.

The first task is to identify the long-range temporal variability at each monitoring stations $\mathbf{s}_i$, $i = 1,...,n$. This can be done using time or frequency domain procedures[5]. The deterministic component $M(\mathbf{s},t)$ is assumed to be modelled, at each site $\mathbf{s}_i$, as the sum of $(K+1)$ basis functions of time $\varphi_j(t)$

$$M(\mathbf{s}_i,t) = \sum_{j=0}^{K} \alpha_j(\mathbf{s}_i)\varphi_j(t) \qquad (2)$$

Where $\alpha_j(\mathbf{s}_i)$ is the coefficient associated with the j-th function $\varphi_j(t)$, with $\varphi_0(t) = 1$ by convention. Each basis function is independent of the spatial location $\mathbf{s}_i$ and should correspond to some components of variability such as temporal trend and annual, seasonal or monthly periodic effects. Thus, for environmental data, we would typically model these components by letting $\varphi_j(t)$ be a series of sine and cosine functions. Exploiting the temporal information, the coefficients $\alpha_j(\mathbf{s}_i)$ can be estimated via ordinary least-squares. Once the $M(\mathbf{s},t)$ component is determined at station locations by equation (2), the residual term is obtained as

$$\eta(\mathbf{s}_i,t) = Z(\mathbf{s}_i,t) - \sum_{j=0}^{K} \alpha_j(\mathbf{s}_i)\varphi_j(t)$$
$$= X(\mathbf{s}_i,t) + \varepsilon(\mathbf{s}_i,t) \qquad (3)$$

Where $X(\mathbf{s}_i,t)$ is a temporally stationary autoregressive spatio-temporal random field and $\varepsilon(\mathbf{s}_i,t)$ is a zero mean Gaussian measurement noise term.

State-space formulation: The presence of a measurement error naturally leads to a state-space representation of model (3). It is then also natural to consider prediction via the Kalman filter[6]. Let $\eta(\cdot, t)$ the n-vector spatial series at time t. The linear Gaussian state-space model[7] is described by the following *state* and *measurement* equations

$$\zeta(\cdot,t) = \Xi\zeta(\cdot, t-1) + \mathbf{v}(\cdot, t) \qquad (4)$$

$$\eta(\cdot,t) = \mathbf{H}(\cdot, t) + \varepsilon(\cdot, t) \qquad (5)$$

Where $\zeta(\cdot, t)$ is the *state vector*, Ξ is the *transition matrix*, $\mathbf{H}$ is the *measurement matrix* and $\mathbf{v}(\cdot, t)$ is *the state noise*. The output $\eta(\cdot, t)$ is a linear function of the state $\zeta(\cdot, t)$ and the state at one time step depends linearly on the previous state. Both state and measurement noise are mutually independent, zero mean normally distributed random variables, with covariance matrices Σ_v and Σ_ε.

Given the autoregressive assumption for $\mathbf{X}(\cdot, t)$, i.e. $\mathbf{X}(\cdot,t) = \sum_{i=1}^{p} \Phi_i\mathbf{X}(\cdot,t-i) + \tau(\cdot,t)$, the reader may recall that equations (4) and (5) represent the state-space formulation of a VAR(p)+Noise model[8] which is obtained by defining

$$\zeta(\cdot,t) = \begin{bmatrix} \mathbf{X}(\cdot,t) \\ \mathbf{X}(\cdot,t-1) \\ \vdots \\ \mathbf{X}(\cdot,t-p+1) \end{bmatrix}$$

$$\Xi = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

$$\mathbf{v}(\cdot,t) = \begin{bmatrix} \tau(\cdot,t) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

and $\mathbf{H} = [\mathbf{I}_n\ \mathbf{0}\ ...\ \mathbf{0}]$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Using the Kalman filter, our goal is to predict a smoother process, $\mathbf{Y}(\cdot, t) = \mathbf{M}(\cdot, t) + \mathbf{H}\,\zeta(\cdot, t)$, at all spatial locations and time points of interest. In the current case, the problem is well defined if we know or can estimate the parameter matrices Ξ, Σ_v and Σ_ε. In practice, we seldom know these and must either specify or estimate them. However, as pointed out by Wikle[1], the crucial issue of modelling spatial structures clearly involves a balance between the structure of the matrix Ξ and the level of independence among the elements of the state noise. For example, in some cases a simple structure of Ξ, obtained by an autoregressive temporal dependence parameter Φ_i, implies that the covariance matrix for the shocks can be represented by a stationary isotropic covariance model[9].

Alternatively, in other circumstances the choice of a non-diagonal Φ_i matrix may be a natural choice. In these cases, the "richer" is the structure of Φ_i, the higher is the level of independence for the state noise. However a parsimonious model can be reached by using a "nearest-neighbour" VAR model where, as for the *Space-Time Autoregressive Moving Average* (STARMA) models[10], the structure of the parameter matrix Φ_i is fixed, according to a graph which specifies relationships between the sites. We shall come back later on this point.

The common components: In order to achieve a dimension reduction, we exploit here the spatial structure to produce a decomposition of the spatial stochastic process into its more basic constituents. Let $Z(\mathbf{s},t)$, described in (1), have the following decomposition

$$Z(s,t) = \mu(s,t) + \delta(s,t) \qquad (6)$$

Where $\mu(\mathbf{s},t)$ is a spatial trend and $\delta(\mathbf{s},t)$ a residual spatial process for a fixed time t. The deterministic component $\mu(\mathbf{s},t)$ is assumed to be modelled as the sum of $K+1$ spatial *trend fields* $f_j(\mathbf{s}_i)$[11] with dynamic coefficients $\beta_j(t)$

$$\mu(s_i,t) = \sum_{j=0}^{K} \beta_j(t) f_j(s_i) \qquad (7)$$

With $f_0(\mathbf{s}_i)=1$ by convention. Applying the theory of generalized Fourier expansions[12,13] the zero mean residual spatial process $\delta(\mathbf{s}; t)$ can be expressed as

$$\delta(s_i,t) = \sum_{v=1}^{\infty} g_v(t) \psi_v(s_i) \qquad (8)$$

Where the terms $\psi_v(s_i)$ are the eigenfunctions of the following homogeneous integral equation

$$\int_D \mathbf{C}_\delta(\mathbf{s}_i,\mathbf{s}_j) \psi_v(\mathbf{s}_j) d(\mathbf{s}_j) = \lambda_v \psi_v(\mathbf{s}_i) \quad v=1,2,... \qquad (9)$$

These eigenfunctions are also known as *principal fields*[11]. Under the hypothesis that it can be assumed known and common in time, $\mathbf{C}_\delta$ is the covariance matrix of the residual process while λ_v are the eigenvalues. Accordingly, the $g_v(t)$ are the principal components obtained as the projection of $\delta(\mathbf{s}_i,t)$ on the eigenfunctions

$$g_v(t) = \int_D \delta(\mathbf{s}_i,t) \psi_v(\mathbf{s}_i) d(\mathbf{s}_i) \quad v=1,2,... \qquad (10)$$

While equation (8) is known as the *Karhunen-Loève Expansion -KLE-*, equation (10) is known as the *Karhunen- Loève Transform (KLT)*. However, it should be noted that we consider the process observed at a collection of sites, so, in practice, only a finite linear approximation of (8), (9) and (10) is possible. Consequently, if there are n sample points in the domain, only n eigenfunctions can be estimated while, indeed, there are a denumerable infinity for a continuous process. Thus, for a continuous domain the difficulties of the approach are considerable, when data are collected only from a sparse and irregular network, since the geometrical relations involving the domain of integration and the relations between the sites $\mathbf{s}_i$, $i = 1,...,n$, are completely ignored in a "discrete" matrix formulation of (9).

In this study, following Fontanella and Ippoliti[14], a coherent numerical treatment of the problem is obtained considering the spectral decomposition of a weighted covariance matrix, where the weights are given by a Voronoi polygon tessellation of the area of interest.

Within the framework of linear approximations, equation (8) is the most efficient representation of the random process if the expansion is truncated to use fewer than n orthonormal basis functions $\psi_v(\mathbf{s}_i)$. That is, ordering the terms of the expansion in decreasing order

of the variances, λ_v, of the coefficients, $g_v(t)$, gives an optimal expansion in the sense that the series truncated at any point minimises the integrated mean square error between the actual and approximated random function[15]. In other words, it means that if we approximate the random process in terms of some number $m<n$ of basis functions, the optimal basis functions for the truncated expansion correspond to the eigenvectors of $\mathbf{C}_\delta$ with the m largest eigenvalues. Finally, note that one reason for truncating the expansion occurs if the random process consists of a signal in additive noise. In this case, it can turn out that by using a truncated expansion, a significant part of the noise is eliminated, while most of the signal is kept intact[16].

Substituting (7) and (8) into (6) it follows that the spatio-temporal process can be expressed as

$$Z(\mathbf{s}_i,t) = \sum_{j=0}^{K} \beta_j(t) f_j(\mathbf{s}_i) +$$
$$\sum_{v=1}^{m} g_v(t) \psi_v(\mathbf{s}_i) + \sum_{v=m+1}^{\infty} g_v(t) \psi_v(\mathbf{s}_i) \qquad (11)$$

Where the third term represents an error process. As noted by Mardia *et al.*[11], equation (11) spans the process in the space of basic functions, $f_j(\mathbf{s}_i)$ and $\psi_v(\mathbf{s}_i)$, known as *common fields*. The identification of such common fields naturally leads to a parsimonious measurement equation. However, in a different manner from Mardia *et al.*[11], a parsimonious state-space linear model may be directly obtained by transforming the data process in the domain of the *common components*, $\beta_j(t)$ and $g_v(t)$, so that a new and reduced data matrix, $\tilde{Z}$, can obtained as

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \beta_0(t_1) & \beta_1(t_1) & \cdots & \beta_K(t_1) & g_1(t_1) & \cdots & g_m(t_1) \\ \beta_0(t_2) & \beta_1(t_2) & \cdots & \beta_K(t_2) & g_1(t_2) & \cdots & g_m(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_0(t_N) & \beta_1(t_N) & \cdots & \beta_K(t_N) & g_1(t_N) & \cdots & g_m(t_N) \end{bmatrix}$$

$$\tilde{Z} = \begin{bmatrix} \tilde{Z}_0 & \tilde{Z}_1 & \cdots & \tilde{Z}_K & \tilde{Z}_{K+1} & \cdots & \tilde{Z}_{K+m} \end{bmatrix}$$

Where as in (1), $\tilde{Z}_u(t)$, $u = 0,...,(K + m)$, can be expressed as

$$\tilde{Z}_u(t) = \tilde{M}_u(t) + \tilde{\eta}_u(t) \qquad (12)$$

Once the spatio-temporal process has been transformed in the space of the *common components*, one can remove the large scale variation $\tilde{M}_u(t)$ as described earlier and use the linear Gaussian state space model for the transformed process $\tilde{\eta}_u(t)$ expressed in the common field's domain.

From the observed data to the common components: We discuss issues associated with the problem of estimating the common components described before.

The first task is to analyze the common spatial structure; thus the type of trend and the residual covariance matrix $\mathbf{C}_\delta$ must be estimated. In this study, the common spatial trend is recognized as a constant surface, representing one trend fields $f_0(\mathbf{s}_i,t)=1$, giving $\mu(\mathbf{s}_i,t)=\beta_0(t)$. Following Cristakos[17], the empirical variogram was obtained by averaging in time the spatial variograms of the residual process $\delta(\mathbf{s}_i,t)$. To asses the best fit for the empirical variogram, we investigated a variety of variogram transition models[9]. By using the weighted least squares procedure, the Indicative Goodness of Fit (IGF) index[9] was applied as a metric for selecting the best fitting variogram model. According to the IGF statistic, we have chosen an omni directional spherical variogram with range 0.1232, partial sill 0.0802 and nugget 0.0703. The principal fields have been obtained by the decomposition of the covariance matrix weighted by the influence areas of the Voronoi polygons[14]. To achieve a dimension reduction, the choice of a truncation parameter m is essential.

In this study the analysis of the scree graph in Fig. 4 suggests to choose m equal to 13. Consequently, the temporal data set $\tilde{\mathbf{Z}}$ of the common components is constituted by one trend component and 13 principal components.

$$M(\mathbf{s_i},t) = \sum_{j=1}^{5} \left[\alpha_{1j}(\mathbf{s_i})\cos(w_j t) + \alpha_{2j}(\mathbf{s_i})\sin(w_j t) \right]$$

Where w_j, $j=1,...,5$, represent, respectively, the frequencies associated to the cycle-trend component, to the annual harmonic, to a period of 6 and 4 months and finally, to a period of 1 week. The plot of the spatio-temporal series associated to the large scale variation is shown in Fig. 5. Figure 6 provides a representation of the residual short temporal variability of the data set.
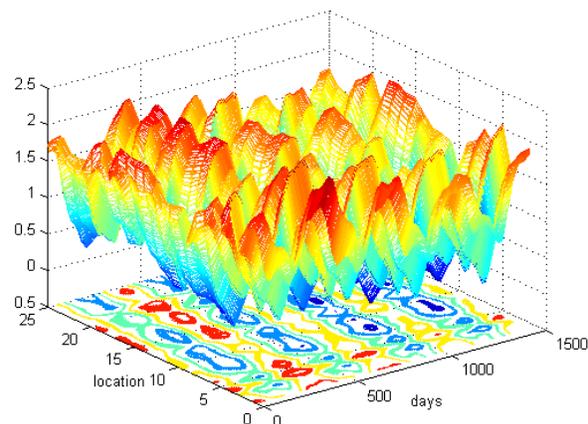
Fig. 5: Surface of the temporal large scale variation for the log-transformed CO concentration

Fig. 4: Scree graph

Modelling the temporal large scale variation: The identification of the temporal dynamic term is performed in the domain of the observed data as well as the domain of the common components. The spectral analysis performed independently on each time series of the observed data suggests the presence of different cycles with different periods. However, from an explorative analysis we have seen that all the time series exhibit the same dominant periodic components. The trend model for the log-transformed *CO* concentration profile at each station $\mathbf{s}_i$, is thus obtained as

Fig. 6: Surface of the residual short temporal variation for the log-transformed CO concentration

Following the same line, the trend model for the common components can be obtained as

$$\tilde{M}_u(t) = \sum_{j=0}^{5} \left[\tilde{\alpha}_{u1_j}\cos(w_j t) + \tilde{\alpha}_{u2_j}(w_j t) \right]$$

Thus, as in the previous case, Fig. 7 and 8 provide the plots of the large and small temporal scale variation for the common components.

MODELLING THE RESIDUAL
SPACE-TIME DYNAMIC TERM

VAR Model in the observed data domain: Once the model for $M(\mathbf{s},t)$ is specified, attention is turned to the $\eta(\mathbf{s},t)$'s. Through the state-space model, the residual
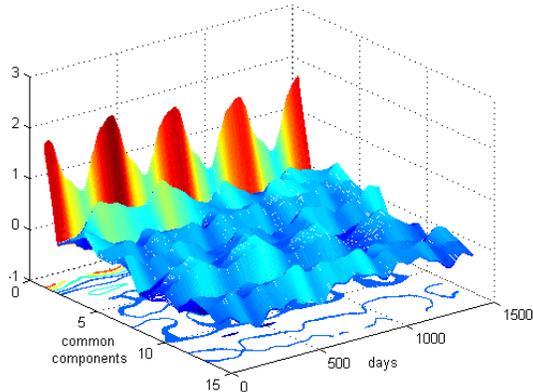
Fig. 7:   Surface of the temporal large scale variation for the common components
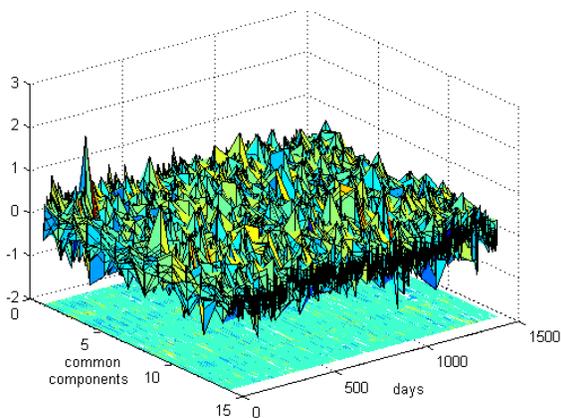
Fig. 8:   Surface of the residual short temporal variation for the common components

space-time dynamic term is modelled as a vector autoregressive (VAR) process. In particular, following the standard diagnostic procedures described in Tiao and Box[18], a first order VAR model, characterized by a 25×25 transition matrix, has been proved to provide the best fit of the observed data.

Nearest-neighbour VAR Model in the observed data domain: In dealing with large spatial dimensions typical of environmental problems, it becomes natural to find a simplified structure of $\boldsymbol{\Xi}=\boldsymbol{\Phi}_1$. In practice, a very simplified version could be achieved considering a scalar structure of $\boldsymbol{\Phi}_1$. This results in the separable spatio-temporal model is described in Huang and Cressie[19]. However, when $\boldsymbol{\Phi}_1$ is assumed diagonal and constant across all spatial locations, the model is not able to capture complicated dynamics. To overcome this problem, a simple extension is to let neighbouring spatial locations at previous time contribute to the process at current time. This model structure is known as *nearest-neighbour* VAR model[1] and constitutes a

particular form of the STAR models[10]. In a different manner from Wikle *et al.*[1], the nearest-neighbour model relies on a *constrained* transition matrix which allows temporal relations only for the *spatially contiguous* sites. Specifically, the spatial contiguity is not specified by a geographical proximity but is directly obtained by looking at the spatial correlation of the observed data. Accordingly, two sites are defined *neighbours* if their distance is less or equal to the variogram range. With this constraint, only 143 parameters, instead of 625, must be estimated. As highlighted in Fig. 9 there is an evident correspondence between the parameters of the constrained transition matrix and the significant ones (at a nominal *p-value* of 5%) of the "complete" transition matrix.

Fig. 9: Transition matrices for the complete VAR model and the nearest neighbour VAR model

Common components-based VAR model: We achieve a dimension reduction of the Markov parameter matrix $\boldsymbol{\Xi}$. In fact, in the common components domain the VAR(1) state-space formulation is characterized by a 14×14 transition matrix. Even if it is not so obvious for this particular example, it should be noted that this approach could lead to a model which results more parsimonious even with respect to the one specified by the *nearest-neighbour* VAR model. This strongly depends on the spatial correlation structure as the larger is the covariance range the smaller is the number of the estimated components. As a guide example, Wikle and Cressie[2], where in their Near-Surface Wind application, the authors achieve a very good approximation of the spatial series, defined over a 17×17 regular grid, using just 20 eigenvectors. However, to provide further insight, Fig. 10 and 11, give some details on the dimension reduction of the parameter matrix $\boldsymbol{\Xi}$ for a (10×10) lattice. Specifically, for an increasing value of the range of three different transition model variograms, Fig. 10a shows the variation of the number (m) of the *principal components*, $g_v(t)$, needed to explain the 75% of the total spatial variability. As is evident, the more the range is large, the lower is the number of components which have to be considered to represent the correlation structure of the process. As expected, the Gaussian model provides a more parsimonious parameterization, which is evident above all for small values of the range. However, as it can be seen in Fig. 10b, the eventual presence of a nugget effect influences the dimensionality reduction and Fig. 10c shows the number of additional principal components which are needed to explain the same percentage of the spatial variation.

Finally, using the Spherical model and the same parameterization resulting from the previous exercise, Fig. 11 compares the number of parameters in $\boldsymbol{\Xi}$ for the *nearest-neighbour* and the common component approaches. To study the influence of a trend component, a quadratic trend surface was also considered in the analysis. As obvious, Fig. 11a shows an inverse relation between the two approaches.
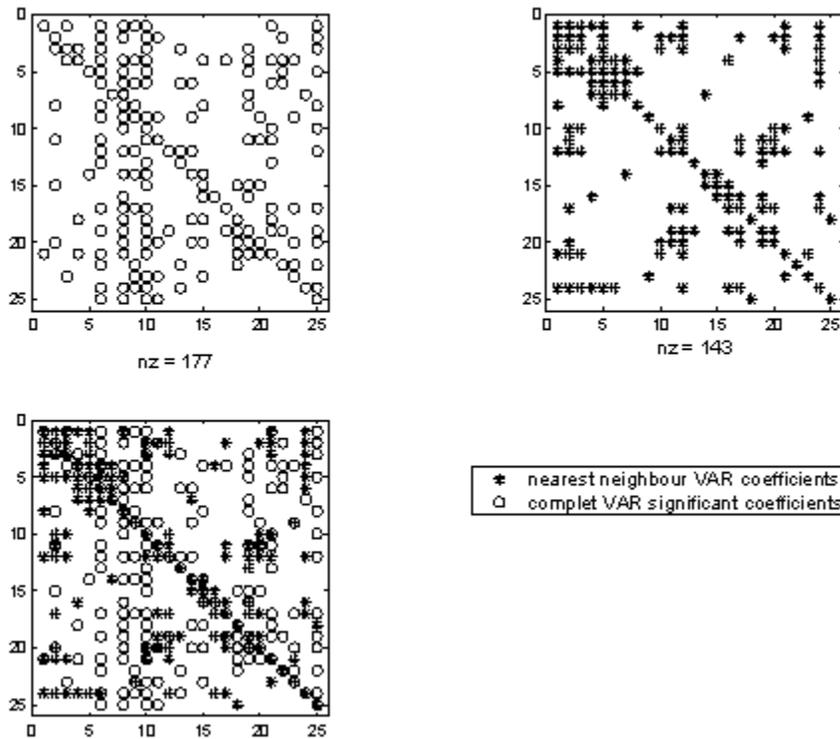
Fig. 9:   Transition matrices for the complete VAR model and the nearest neighbour VAR model
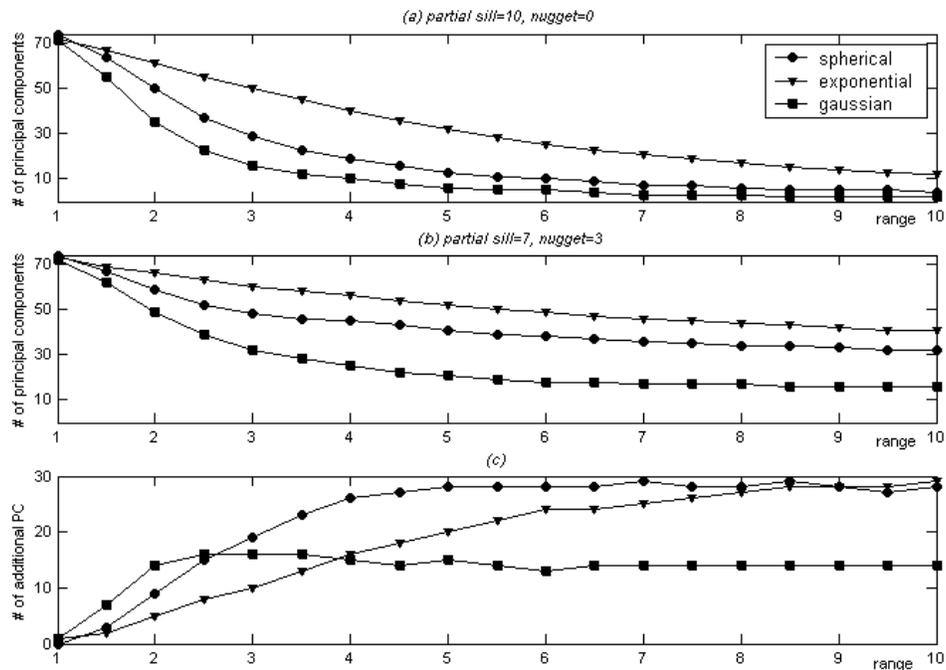
Fig. 10: *(a)* Relation between the number of principal components (PC) and the range of three different transition variogram models with partial sill=10 and nugget=0; *(b)*Relation between the number of PC and the range when a nugget effect is considered; *(c)*Number of additional PC which have to be taken into account when a nugget effect is considered. In all cases, the number of PC explain the 75% of the total spatial variation over a (10×10) lattice
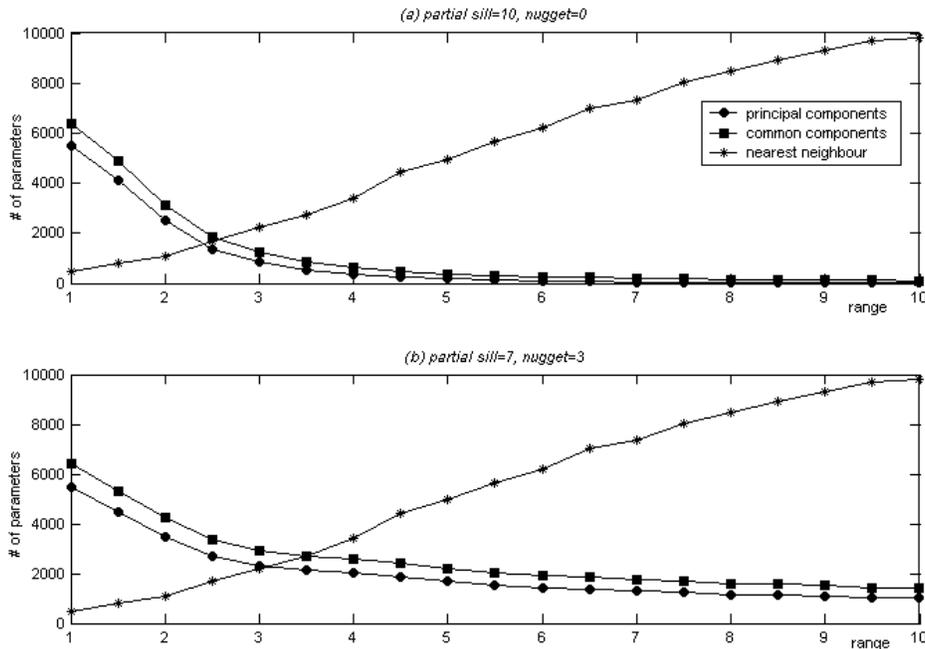
Fig. 11: *(a)*Relation between the number of parameters in Ξ and the range of a spherical variogram model with partial sill=10 and nugget=0; *(b)* Relation between the number of parameters in Ξ and the range of a spherical variogram model with nugget. In both cases, the number of PC explain the 75% of the total spatial variation over a (10×10) lattice
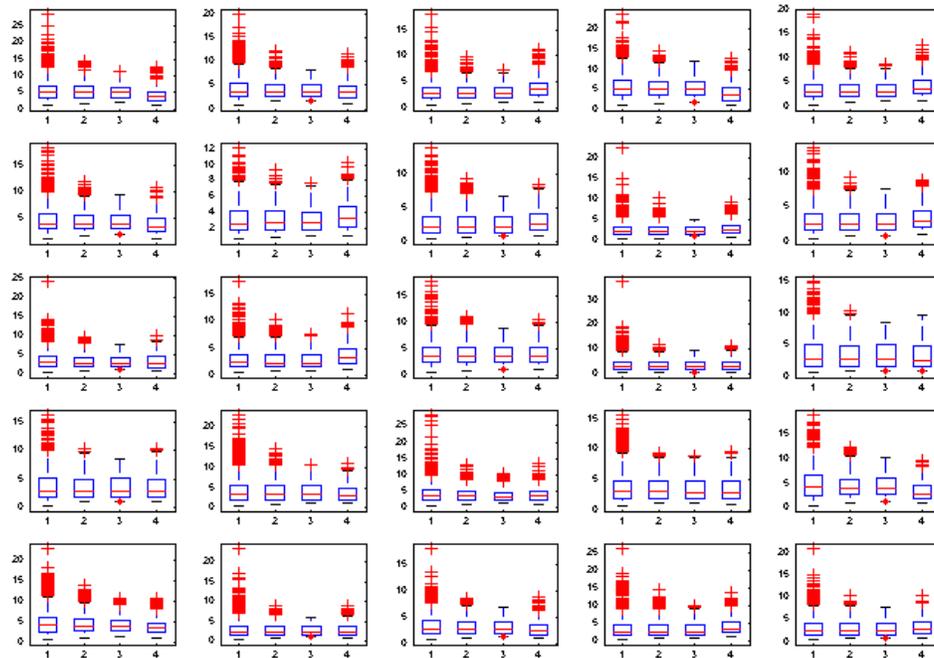
Fig. 12: Boxplot of the temporal series for 1998-2001 –1=observed data; 2=Complete VAR; 3=Nearest Neighbour VAR; 4=Common Components VAR

In particular, we can see that for a (10×10) lattice, both procedures are characterized by (approximatively) the same number of parameters for a range of 2.5.

For larger ranges the common component method leads to a more parsimonious parameterization. On the other hand, the situation slightly improves for the *nearest-neighbour* method when a nugget effect is considered (Fig. 11b) since, in this case, it should be preferred to the common component approach (with a quadratic trend) until the range reaches the point 3.5.
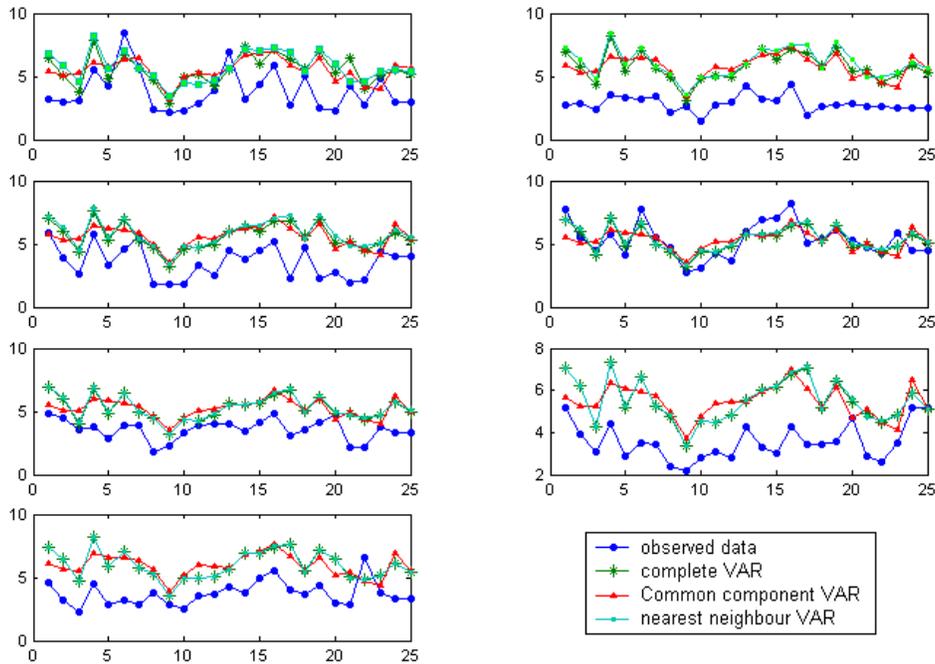
Fig. 13: Predicted spatial series for the last week of December 2001
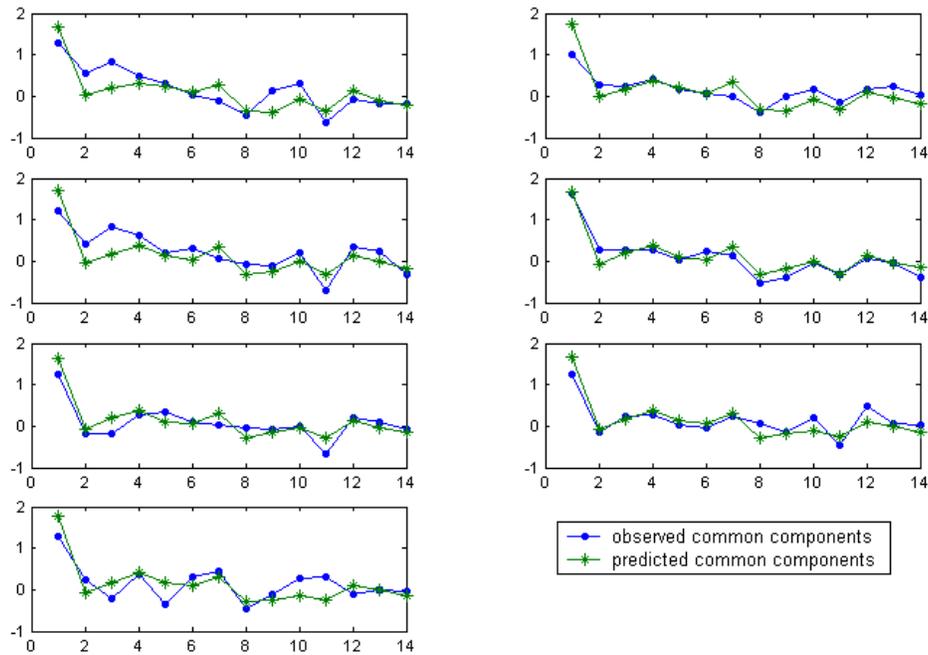
Fig. 14: Predicted common components for the last week of December 2001

FITTING AND FORECASTING RESULTS

To examine the effectiveness of our model we have applied the Kalman filter and compared the fitted series with the observed data. Figure 12 shows the box plot of the distributions of the observed and predicted data for each monitoring station. As it can be noted, all the procedures perform similarly yielding analogous predicted distributions. With respect to the observed data distribution, as expected, the plots highlight a smoothing effect. This could be due to the difficulties of the Kalman filter in fitting extreme values.

Furthermore, to test the model ability to perform temporal predictions, the last week of the observed data set has been taken out from the analysis. Figure 13 compares the observed and predicted spatial series for

the period 25-31 December 2001. As previously noted, even for the temporal forecasts all the procedures lead to similar results.

Focusing the attention to the Common Fields domain, Fig. 14 allows to stress the good performance of the VAR model in forecasting the Common Components for the last week of December 2001.

DISCUSSION

The dynamic model described here offers a flexible approach to modelling a large-class of environmental space-time processes. However, as usual, the model must be tailored to the problem at hand. For example, if our goal is in predicting at a future time, our approach is useful; differently, further details must be provided if we are trying to predict the process at unobserved spatial locations. In particular, given $\tilde{\mathbf{Z}}$, if such prediction is required at time t, $t \leq T$ and at an unmonitored site s_0, a straightforward approach might use equation (11) to obtain

$$\hat{Z}(\mathbf{s}_0, t) = \sum_{j=0}^{K} \beta_j(t)\hat{f}_j(\mathbf{s}_0) + \sum_{v=1}^{m} g_v(t)\hat{\psi}_v(\mathbf{s}_0) \qquad (13)$$

Where $\hat{f}_j(\mathbf{s}_0)$ and $\hat{\psi}_v(\mathbf{s}_0)$ are the estimated *common fields* at site $\mathbf{s}_0$. Their prediction is not a difficult task and since they are non-stochastic, ensuring orthogonality, we could apply some relatively simple interpolation schemes. To that end, Mardia *et al*.[11] and Wikle and Cressie[2], provide two alternative approaches.

We have also seen that the crucial issue of modelling spatial structure involves a trade-off between the structure of the VAR autoregressive parameter matrix and the level of independence among the elements of the driving noise. In order to gain a deeper understanding further work is needed to consider additional parameterizations. In this context, even if one loses the computational efficiency and simplicity realized by the Kalman filter, a full Bayesian model[1,20] provides a useful approach. In this context, spatial predictions can also be obtained following Tonellato[20].

The final point is related to the typology of the data structure. For instance, of particular interest could be those phenomena characterized by a persistent (*long-memory*) temporal autocorrelation. Thus, the investigation of VARFIMA models[21] represents a natural extension of the present approach and is a topic for future work.

REFERENCES

1. Wikle, C.K., L.M. Berliner and N. Cressie, 1998. Hierarchical bayesian space-time models. Environ. and Ecolog. Stat., 5: 117-154.
2. Wikle, C.K. and N. Cressie, 1999. A dimension-reduction approach to space-time kalman filtering. Biometrika, 4: 815-829.
3. Goodall, C. and K.V. Mardia, 1994. Challenges in multivariate spatio-temporal modelling. Proc. of the XVIIth Intl. Biometric Conf., Hamilton, Canada.
4. Sandwell, D., 1987. Biharmonic splines interpolation of GEOS-3 and SEASAT altimeter data. Geophys. Res.Lett., 14: 139-142.
5. Chatfield, C., 1996. The Analysis of Time Series: An Introduction. 5th Edn. Chapman and Hall, London
6. Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. (ASME), 82D: 35-45.
7. West, M. and J. Harrison, 1997. Bayesian Forecasting and Dynamic Models. 2nd Edn. New York: Springer-Verlag.
8. Lutkepohl, H., 1993. Introduction to Multiple Time Series Analysis. 2nd Edn. New York: Springer Verlag.
9. Cressie, N., 1993. Statistics for Spatial Data. New York, Wiley.
10. Pfeifer, P.E. and S.J. Deutsch, 1980. A three-stage iterative procedure for space-time modeling. Technometrics, 22: 35-47.
11. Mardia, K., E. Redfern, C. Goodall and F. Alonso, 1998. The Kriged Kalman Filter. TEST, 7: 217-285.
12. Loève, M., 1945. Functions Aleatoires de Second Ordre. CR Acad. Sci. Paris, pp : 220-469.
13. Karhunen, K., 1947. Uber Linear Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fenn., (AI), 37 : 1-79.
14. Fontanella, L. and I. Ippoliti, 2003. Dynamic models and karhunen- loève expansion for spatio-temporal data set. Statistical methods and application. J. Italian Stat. Soc., 12: 61-78.
15. Freiberger, W. and U. Grenander, 1965. On the formulation of statistical meteorology. Rev. Intern. Statist. Inst., 33: 59-86.
16. Preisendorfer, R.W., 1988. Principal Component Analysis in Meteorology and Oceanography. Elsevier, Amsterdam.
17. Cristakos, G. and D. Hristopulos, 1998. SpatioTemporal Environmental Health Modelling: A Tractatus Stochasticus. Kluwer Academic Publishers, Boston.
18. Tiao, G.C. and G.E.P. Box, 1981. Modeling multiple time series with applications. J. Am. Stat. Assoc., 76: 376.
19. Huang, H.C. and N. Cressie, 1996. Spatio-temporal prediction of snow water equivalent using the Kalman Filter. Computational Statistics & Data Analysis, 22: 159-175.
20. Tonellato, S., 1997. Bayesian dynamic linear models for spatial time series. Tech. Report 5/1997, Dipartimento di Statistica- Università Ca' Foscari di Venezia.
21. Haslett, J. and A.E. Raftery, 1989. Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. Applied Statistics, 38: 1-50.