Non-separability of space-time covariance models in environmental studies *

Francesca Bruno†, Peter Guttorp‡, Paul D. Sampson‡, Daniela Cocchi†
†University of Bologna. Department of Statistics "P. Fortunati". Bologna, Italy
‡University of Washington. Department of Statistics. Seattle, USA

Abstract

In the last two decades space-time models have been studied with increasing interest. The main reason for these developments is their importance to problems of spatio-temporal prediction for environmental processes. The space-time covariance model is the fundamental tool for characterizing both the temporal and spatial components of the phenomena under study. The empirical estimation of space-time covariance models can be very complex without simplifying hypotheses. For this reason many authors assume separability between spatial and temporal components. This assumption is not realistic from an empirical point of view. In this paper we propose a model in which non-separability arises from daily variability changing seasonally. The final model is applied to tropospheric ozone data from the Emilia-Romagna region in Italy.

Keywords: non-separability, non-stationarity in scale, ozone data, space-time model

1 Introduction

For most analyses of environmental processes that are spatially and temporally indexed, it is often necessary to specify a covariance structure in space and time. Much work in geostatistics has made the simplifying assumption that the space-time covariance structure of a temporally and spatially stationary process is separable, i.e., can be written as the product of a purely spatial and a purely temporal covariance function. We propose a new model for space-time covariance structure in which space-time non-separability arises because of daily variability changing seasonally. We apply the model to tropospheric ozone data from 32 sites in the Emilia-Romagna region of Italy.

The research leading to this paper has been partially funded by Consiglio Nazionale delle Ricerche (short-mobility project) and by 2002 MIUR grant (protocol n. 2002134337-001). The NRCSE provided the structure and space to work on this project.
Cressie and Huang (1999) introduced classes of nonseparable stationary covariance functions to model space-time interactions. They based their approach on Fourier transforms. Recently, Gneiting (2002) proposed another general class of nonseparable, stationary covariance functions for spatio-temporal random processes directly in the space-time domain (i.e., it is a construction not based on the inversion of a Fourier transformation). In both these papers the space-time processes are assumed stationary in time and space. In many environmental applications these assumptions are not satisfied. Meiring et al. (1999) illustrated non-separability of space-time covariance in hourly ozone due to diurnal variation in the spatial correlation structure, with strong spatial correlation in the afternoon, but hardly any at night.

In this paper we consider ozone data on a daily time scale over 5 years where the seasonal characteristics are important to both the mean and the space-time covariance structure. Following a brief consideration of formal definitions of separable space-time covariance models in Section 2, we propose in Section 3 a simple model in which space-time non-separability arises because of seasonal non-stationarity in the short-scale temporal variability. In Section 4 we describe the tropospheric ozone data from Emilia-Romagna, Italy, we discuss simple marginal time and space correlation functions, and we describe the final temporal and spatial model. We conclude with discussion in Section 5.

2 Separable space and time covariance models

The definition of space-time separability is given in Cressie and Huang (1999). In particular, it is possible to establish separability as a condition that satisfies the two following statements (see Cressie and Huang, 1999):

1. the temporal covariance function is the same at all monitoring locations, regardless of the relative displacement of the locations:

   \[ \text{Cov}(Y(t, s), Y(t', s')) = \text{Cov}(Y(t, s'), Y(t', s)) \]  

   for all \((s, s'), t, t' = 1, ..., T;\)

2. the spatial covariance function is constant in time, so that

   \[ \text{Cov}(Y(t, s), Y(t, s')) = \text{Cov}(Y(t', s), Y(t', s')) \]  

   for all \((t, t'), s, s' \in \mathbb{R}^2.\)
If one assumes spatial and temporal stationarity, we may consider just space-time correlation functions, in which case we have the simple product of a spatial correlation function and a temporal correlation function:

\[ \text{Corr}(Z(t, x), Z(s, y)) = \rho_1(|t - s|)\rho_2(||x - y||) \]  

(3)

for \( t, s \in \mathbb{R}; x, y \in \mathbb{R}^2 \). In equation (3) \( \rho_1 \) is the purely temporal autocorrelation function, and \( \rho_2 \) is the purely spatial correlation function. This framework extends readily to accommodate nonstationarity in the spatial correlation structure using the spatial deformation model of Damian et al. (2001). This model is just a stationary model after an appropriate transformation of the geographic coordinate system, so that (3) becomes:

\[ \text{Corr}(Z(t, x), Z(s, y)) = \rho_1(|t - s|)\rho_2(||f(x) - f(y)||) \]  

(4)

for \( t, s \in \mathbb{R}; x, y \in \mathbb{R}^2 \). In equation (4) \( f \) is a smooth and bijective function which permits a deformation of the spatial coordinates from a “G-plane”, the original geographic coordinate system to a “D-plane” coordinate system that encodes spatially varying local anisotropy commonly recognized in most spatio-temporal environmental processes (depending on the spatial and temporal scale at which the process is observed).

3 The Model

Let \( \{Y(t, x); t \in [0, \infty), x \in S \subset \mathbb{R}^2\} \) represent a general spatio-temporal random process. The process could be represented as

\[ Y(t, x) = \mu(t, x) + Z(t, x) + \epsilon(t, x); t = 1, ..., T, x = 1, ..., X \]  

(5)

where \( \mu(t, x) \) is the spatio-temporal mean field or trend, \( Z(t, x) \) denotes a smooth spatio-temporal underlying process, and \( \epsilon(t, x) \) is a Gaussian noise process, representing spatio-temporal measurement error (and small scale variability). When considering annual processes one frequently must take into account features of seasonal variability beyond that encoded in the spatio-temporal trend. The data we discuss below suggest a simple but useful approach to this is to write

\[ Y(t, x) = \mu(t, x) + \sigma_t(x)Z^*(x, t) + \epsilon(t, x); t = 1, ..., T, x = 1, ..., X. \]  

(6)

In some cases the temporal correlation structure of \( Z^*(x_i, t) \) can be modeled in terms of a further decomposition,

\[ Z^*(x_i, t) = \alpha_iZ_1(t) + Z_2(x_i, t) \]  

(7)
where $i$ denotes the site at location $x_i$, $\alpha_i$ is a site-specific coefficient for the (large scale) temporal dependence process $Z_1(t)$, and $Z_2(x_i, t)$ is a temporally uncorrelated space-time process.

4 Tropospheric ozone in the Emilia Romagna region of Italy

4.1 The data

The data set consists of tropospheric ozone measurements from 32 monitoring stations across the Emilia-Romagna region of Italy (Figure 1). The area is in the North-east part of Italy with the Adriatic sea to the East and the Appennini mountains to the South. The farthest monitoring sites are about 280 km far from each other. Some monitoring sites are located close to the sea, and only a few are at altitude higher than 100 m. Our analyses concern ozone concentrations at a daily time scale expressed in terms of daily maximum 8-hours moving average computed from the hourly ozone concentration data recorded in micrograms per cubic meter, $\mu g/m^3$, for the period 1998-2002. Eight-hour moving average time series are suggested by Italian Law (D.M.A. 16/05/1996) as well as U.S. EPA air quality standards. The

![Figure 1: Map of the monitoring sites (Emilia Romagna)](image)

stations shown in Figure 1 are unequally spaced, most of them are along the main
route (via Emilia) that crosses the entire area. In the original time series some missing values are present in a small percentage. These are mainly due to instrumental problems. To detect the trend in the data we performed a two-way decomposition of temporal and spatial effects on the original data. In particular, by means of the Median Polish procedure it was possible to describe each observation as composed of several effects. In particular, we found four main effects: a grand mean effect, a site-specific effect, a year-specific effect and a day-specific effect. These effects enter

Figure 2: Trend decomposition by means of Median polish in the trend component of the model. The residuals from the spatio-temporal trend determination are represented in Figure 3 for some of the 32 monitoring sites.

Figure 3: Ozone time series (after removing the trend)
4.2 Temporal covariance structure

To investigate the time structure we used Box-Jenkins time series analysis (Box and Jenkins, 1976). The analysis was performed separately for each monitoring station. Almost all the monitoring stations show a non-stationary component. In particular, the best model, selected by means of residuals analysis and AIC index, is an ARIMA(1,1,1) model. However, the ARIMA(1,1,1) process is a non-stationary temporal model. Therefore we suggest using model (6) instead of model (5) to represent the process. In practical terms, the first step consists of standardizing the temporal series with moving window means and standard deviations. We set the width of each moving window to be approximately 30 days in order to represent smooth seasonal variation. In this way we identify $\sigma_t(x)$, for each $x \in S$. The role of $\sigma_t(x)$ in this paper is really fundamental. This component represents the part of the non-stationary temporal variability that, when removed, leaves separable space and time correlation components. After we estimate this component, we use it to standardize our data, and we can use classical separable space-time correlation models. In Figure 4 we report the behavior of the estimated $\sigma_t(x)$ for some sites. These represent the aspect of daily variability that changes seasonally.

![Figure 4: Behaviour of σₜ(x) for some monitoring sites](image)

All the series of $\sigma_t(x)$ present the same behavior in time with a clear seasonal-
ity. The new space-time standardized series were again analyzed by means of Box-Jenkins techniques for time series analysis. The time series process is now stationary in time and the analysis suggests an AR1 model. Since each time series seemed to have a similar structure we decided to fit an underlying unique AR1 model. To find the right AR coefficient estimate we computed a principal component analysis of the space-time data matrix and then fitted an AR1 model to the time series defined by the first principal component. The maximum likelihood estimate for the coefficient of the AR1 model is 0.48.

4.3 Spatial covariance structure

A spatial analysis is used to identify the spatial covariance structure (Cressie, 1993). In particular, the empirical correlations for $Z^*(x, t)$ for temporal lag 0 are reported in the top left panel of Figure 5. These correlation could be fitted with an exponential model with a nugget effect. However, the correlation decreases to a non-zero value for distances represented in our data. This suggested consideration of the model (7) with a large scale process $Z_1(t)$. In this case the spatial correlation is shown in the center panel of Figure 5. Moreover, an analysis of spatial non-stationarity is performed by using the Damian, Sampson and Guttorp (2001) approach to estimating heterogeneous spatial covariance by means of Markov Chain Monte Carlo simulation. The new correlations are then presented in the right panel of Figure 5. One can see that removal of the large scale process $Z_1(t)$ resulted in spatial correlations tending to zero within the spatial range of the data (Figure 5b), while the spatial deformation clearly results in a better (tighter) empirical correlogram (Figure 5c).

In Figure 6 we show the spatial deformation (right panel) of a regular grid in the geographic plane (left panel). It is characterized mainly by compression, meaning
high spatial correlation, along the northwest-southeast axis of the principal route through the region, especially in the northwest.

![Spatial deformation for process \( Z_2 \)](image)

**Figure 6: Spatial deformation for process \( Z_2 \).**

### 5 Discussion

Most of the nonseparable space-time correlation models are difficult to understand and apply. The model proposed in this paper is quite simple. It is based on the idea that the non-separability of time and space component due to non-stationarity in time can be managed by a new parameterization of a general model. If the seasonal effect does not change over years, we can think of it as \( \sigma_t(x) = \sigma_{<t>}(x) \), where \( <t> = t - [t] \) is the time within the year. This can easily be estimated by averaging over years. In addition, missing stretches of data that affect estimated variances (such as around time 500 in the CITT plot in Figure 4) would be eliminated by averaging over years. In this analysis we have ignored meteorological covariates. Some preliminary analysis (Bruno, 2003) indicates a slight difference in the correlation structure between different wind regimes.
References


