

Comparing spatio-temporal hierarchical models for air quality data

Michela Cameletti and Rosaria Ignaccolo

Abstract We propose a comparison approach for hierarchical spatio-temporal models, considering both goodness of fit and model complexity, as well as computational load. Here we implement different models on particulate matter, but the proposed “set of criteria” can be used for general environmental phenomena where a hierarchical structure is suitable, a large scale trend is included and a choice of spatio-temporal covariance function needs.

Key words: Spatio-temporal covariance functions, Prediction performance indexes, Spatio-temporal models, Environmental statistics.

1 Introduction

Particulate matter (PM) is one of the most important air pollutants in Europe and, despite the improvements due to the European Union (EU) legislation, it still has a heavy toll on human life ([3]). From the statistical perspective, many articles have been proposed for modelling PM concentrations and understanding their complex underlying spatio-temporal dynamics. The common characteristic of many models in literature is their hierarchical structure. This means that they are constructed by putting together conditional sub-models which are defined at each hierarchical stage. With reference to the likelihood, this corresponds to take a conditional point of view for which the joint probability distribution of a spatio-temporal process can be expressed as the product of some simpler distributions. This property makes it possible to cope with the complexity of spatio-temporal processes in a straightfor-

Michela Cameletti
University of Bergamo, Via dei Caniana 2, 24127 Bergamo e-mail: michela.cameletti@unibg.it

Rosaria Ignaccolo
University of Torino, Corso Unione Sovietica, 218 bis, 10134 Torino e-mail: ignaccolo@econ.unito.it

ward way and this is the reason why hierarchical models have become so popular for modelling environmental phenomena ([7]), especially from a Bayesian perspective.

In this context an interesting question arises about which is the suitable hierarchical structure to combine with the spatio-temporal covariance function for catching the spatial and temporal dynamics of the considered process. Is it preferable to have a two-level model with a complex nonseparable covariance function or a model with a more complicated hierarchical structure but a simpler covariance function? Finding an answer to this question could be useful in order to supply the Environmental Agencies with a statistical models for building reliable PM concentration maps, provided with the corresponding uncertainty measure.

The goal of this paper is to answer this question, so that we propose some as objective as possible “criteria” to compare air quality spatio-temporal models. Such criteria consider both goodness of fit and model complexity as well as computational load. Actually a model comparison is proposed by Huang et al. ([4]) through AIC and BIC criteria as well as mean squared prediction error at a fixed time. Here we compare models on real data concerning PM, but the proposed comparison approach holds for general environmental phenomena where a hierarchical structure is suitable, a large scale trend is included and a choice of spatio-temporal covariance function needs. The paper is organized as follows. Section 2 presents and discusses hierarchical models different in their dynamics and covariance structure. Such models are implemented on PM data in Piemonte, and their performances are compared in Section 3. Section 4 concludes.

2 Hierarchical spatio-temporal models

Let $z(s_i, t)$ be the scalar spatio-temporal process measured at site s_i and at time t where $i = 1, \dots, d$ and $t = 1, \dots, T$. We assume the following *measurement equation* as the first level of all the hierarchical models we consider,

$$z(s_i, t) = u(s_i, t) + \varepsilon(s_i, t) \quad (1)$$

where $\varepsilon(s_i, t) \sim N(0, \sigma_\varepsilon^2)$ is the measurement error defined by a Gaussian white-noise process serially and spatially uncorrelated. In the following we consider some specifications for the so-called state process $u(s_i, t)$ giving rise to different hierarchical models.

MODEL A

Model A is a two-level model characterized by a large-scale term and a spatio-temporal process for the residual small-scale component. In particular, the quantity $u(s_i, t)$ of Eq.1 is given by the sum of a trend $\mu(s_i, t)$ and a random process $\omega(s_i, t)$, as follows

$$u(s_i, t) = \mu(s_i, t) + \omega(s_i, t). \quad (2)$$

The trend $\mu(s_i, t)$ is a function of k covariates and is given by

$$\mu(s_i, t) = X(s_i, t) \beta \quad (3)$$

where $X(s_i, t) = (X_1(s_i, t), \dots, X_k(s_i, t))$ denotes the covariate vector for site s_i at time t and $\beta = (\beta_1, \dots, \beta_k)'$ is the coefficient vector.

The zero-mean Gaussian process $\omega(s_i, t)$ of Eq.2 is the residual small-scale process whose spatio-temporal covariance function depends on the parameter vector θ , namely

$$\text{Cov}(\omega(s_i, t), \omega(s_j, t')) = \sigma_\omega^2 C_\theta(h, l) \quad \forall i \neq j, t \neq t' \quad (4)$$

where $h = \|s_i - s_j\|$ is the Euclidean distance between site i and j , $l = |t - t'|$ is the temporal lag between t and t' and σ_ω^2 is the constant in time and space variance of the process. As the covariance function of Eq.4 depends only on h and l , the process is second-order stationary and isotropic. We consider 3 different forms for the correlation function $C_\theta(h, l)$ defining the following models.

Model A1 Under the hypothesis that the process $\omega(s, t)$ is i.i.d. over time, it holds that

$$\text{Cov}(\omega(s_i, t), \omega(s_j, t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C_\theta(h) & \text{if } t = t' \end{cases} \quad (5)$$

where $C_\theta(h)$ is a purely spatial correlation function. To this regard many spatial covariance functions defining isotropic second-order stationary processes can be found in [1, Ch.1].

Model A2 Adopting a separable approach, the space-time covariance function factors into a purely spatial and a purely temporal component and is given by

$$C_\theta(h, l) = C_{\theta_1}(h) C_{\theta_2}(l)$$

with $\theta = (\theta_1, \theta_2)$. In this case there is no interaction between space and time.

Model A3 Supposing a nonseparable relation between space and time, a space-time correlation function, chosen from the general class introduced by Gneiting in [2], is defined by

$$C_\theta(h, l) = \frac{1}{\psi(|l|^2)} \varphi\left(\frac{h^2}{\psi(|l|^2)}\right). \quad (6)$$

The function $\psi(x), x \geq 0$, is any completely monotone function while $\varphi(x), x \geq 0$, is any positive function with a completely monotone derivative. In particular, we consider the two cases described in Table 1, named respectively **Model A3-1** and **Model A3-2**, where $0 < b \leq 1$, $v > 0$, the smoothness parameters α and γ take values in $(0, 1]$, the space-time interaction parameter τ is defined in $[0, 1]$ and the scaling parameters of time and space a and c are positive. It follows that $\theta = (b, \alpha, \gamma, a, c)$ for Model A3-1 and $\theta = (v, \alpha, \gamma, \tau, a, c)$ for Model A3-2.

Table 1 Function $\psi(x)$ and $\varphi(x)$ that define the nonseparable spatio-temporal correlation function of Model A3-1 and A3-2.

	ψ	φ
Model A3-1	$\psi(x) = (ax^\alpha + b)/(b(ax^\alpha + 1))$	$\varphi(x) = \exp(-cx^\gamma)$
Model A3-2	$\psi(x) = (ax^\alpha + 1)^\tau$	$\varphi(x) = (1 + cx^\gamma)^{-\nu}$

MODEL B

Differently from Model A, Model B has a three-level hierarchical structure due to an additional equation introduced for modelling the temporal dynamics of a latent process constant in space. This means that for Model B the term $u(s_i, t)$ of Eq.1 is given by the sum of a trend $\mu(s_i, t)$, of an unidimensional latent process $y(t)$ and of a purely spatial process $\omega(s_i)$, as follows:

$$u(s_i, t) = \mu(s_i, t) + y(t) + \omega(s_i). \quad (7)$$

In Eq.7 the trend $\mu(s_i, t)$ is defined as in Eq.3, while the term $y(t)$ refers to an unidimensional process constant in space with a temporal dynamics given by

$$y(t) = \rho y(t-1) + \eta(t) \quad (8)$$

where $y(0) \sim N(0, \sigma_0^2)$ and $\eta(t) \sim N(0, \sigma_\eta^2)$. Finally, the spatial process $\omega(s_i)$ of Eq.7 follows a Gaussian distribution $N(0, \sigma_\omega^2)$ whose covariance function is defined as $Cov(\omega(s_i), \omega(s_j)) = \sigma_\omega^2 C_\theta(h)$. Note that from Eq.7 and 8 it follows that Model B is characterized by an additively separable spatio-temporal covariance function.

MODEL C

Model C is a three-level hierarchical model described by a spatio-temporal residual process with a temporal dynamics. Specifically, the term $u(s_i, t)$ of Eq.1 is given by the trend $\mu(s_i, t)$ and a spatio-temporal process $y(s_i, t)$, namely:

$$u(s_i, t) = \mu(s_i, t) + y(s_i, t). \quad (9)$$

The trend $\mu(s_i, t)$ has the same structure as for Model A and B (see Eq.3) while $y(s_i, t)$ is a spatio-temporal process which changes in time according to the following dynamics

$$y(s_i, t) = \rho y(s_i, t-1) + \omega(s_i, t) \quad (10)$$

where $y(s_i, 0) \sim N(0, \sigma_0^2)$ and $\omega(s_i, t) \sim N(0, \sigma_\omega^2)$ is a spatio-temporal process i.i.d. over time whose spatio-temporal correlation function is given by Eq.5. Note that it can be shown that also Model C is characterized by a separable spatio-temporal covariance function, but in multiplicative form.

MODEL DISCUSSION

All the considered models are characterized by the same measurement equation given by Eq.1 and by the trend term of Eq.3 defined as function of some covariates which may change in space and time (e.g. meteorological variables) or may be constant in time (e.g. spatial coordinates).

Model A has a very simple structure characterized by a unique residual spatio-temporal process for which 3 different covariance functions are considered. Model A3 is the most complex model due to its nonseparable spatio-temporal correlation function. Considering that space and time interact together, it is reasonable to use a nonseparable approach from a physical point of view even if it has some computing drawbacks related to the massive matrix dimensions. Thus some simplifications are introduced for Model A1 and Model A2 where we respectively adopt the i.i.d. over time assumption and the separability hypothesis.

Model B and C differ from Model A for their three-level hierarchical structure: in both the case in fact a stage is introduced for modelling the temporal dynamics of a latent process. In particular, for Model B this is given by a purely temporal process (see Eq.8) which is supposed to be constant in space while for Model C the latent process corresponds to a purely spatial process changing in time through Eq.10. It is important to point out that when all the data are collected together, the dimension of the latent process is always unidimensional for Model B while for Model C it depends on the number of sites d . From the computational perspective this means that the convergence of the estimation procedure is lower for Model C even if for Model B there is an extra parameter to estimate, namely σ_η^2 .

3 Model comparison for PM₁₀ in Piemonte

We implement the 6 models presented in Sect. 2 with reference to PM₁₀¹ concentration data measured in Piemonte region in year 2006 ($T = 365$). In particular, we consider $d = 25$ Low Volume Gravimetric (LVG) monitoring stations for PM₁₀ and 13 extra validation stations for model performance assessing.

Exploratory analyses of the measured PM₁₀ concentrations at the different sites show skewed distributions and standard deviation correlated with mean, so we choose to transformate data by the logarithm to stabilize the variances and make the distributions approximately normal. The covariates are chosen by a preliminary regression analysis using the Akaike's information criteria (AIC) and parameter significance. The variables included in the models' trend are: intercept, altitude, spatial coordinates, daily mean wind speed, daily maximum mixing height, amount of precipitation and daily PM emissions. Finally, we adopt the exponential form for the spatial covariance function of Model A1, A3, B and C. In particular, $C_\theta(h) = \exp(-\theta h)$ holds for Model A1, B and C while for Model A3 we have that

¹ Particulate Matter with a diameter lower than 10 μm .

$C_\theta(h, l) = \exp(-\theta_1 h) \exp(-\theta_2 l)$. For Model A1 and A2 the nonseparable covariance functions reported in Table 1 are used.

As regards inference, i.e. parameter estimation and spatial prediction, we adopt a fully Bayesian approach via Monte Carlo Markov Chain (MCMC) methods implemented in R software [6]. In particular, we use the Metropolis-within-Gibbs algorithm together with Normal independent priors $N(0, 100)$ for each trend coefficients and independent Inverse Gamma distributions $IG(2, 1)$ for the variance parameters.

In order to compare the estimated models we firstly consider their intrinsic complexity evaluated by means of the number of parameters to be estimated. Moreover, we take into account the computational complexity quantified considering the dimension of the biggest matrix to be inverted and the time required for computing the spatial predictions over the validation sites. From the results reported in Table 2, we can see that the nonseparable models A3-1 and A3-2 are the most complex because they are characterized by a larger parameter vector and $dT \times dT$ dimensional matrices. Considering that matrix inversion is of order n^3 in computation (where n is the total number of data), this gives rise to a massive computational load for the nonseparable models. This complexity does not find a counterpart in the corresponding prediction computing times reported in Table 2 that are expected to be higher. This is due to the fact that for the moment all the inferential procedures related to Model A2 and A3 have been implemented, for computational feasibility reasons, with `RandomFields` R package while for the other models ad-hoc written code is used.

With reference to the parameter estimates, which are not reported here for brevity, we point out that the trend regressor coefficients are robust with respect to the model specification. In particular, the intercept is estimated around 3.5 (on the log scale) and, as expected, daily PM emissions is the only regressor with a positive coefficient.

Finally we assess the model prediction performance considering five indexes based on the differences between predicted and observed data. More precisely, together with the usual RMSE and the Pearson correlation coefficient ρ , we adopt the Normalised Mean Bias Factor (NMFB) recently introduced in [8] and two indexes WNNR and NNR, proposed in [5], defined on the Normalized Ratios between the predicted and the observed values. For a fixed site s_i , let O_t be the observed time series and P_t the predicted time series, with $t = 1, \dots, T$, and denote $\bar{O}$ and $\bar{P}$ the corresponding mean values. The Normalised Mean Bias Factor

is defined by $NMBF = \begin{cases} \frac{\sum_t P_t}{\sum_t O_t} - 1 & \text{if } \bar{P} \geq \bar{O} \\ 1 - \frac{\sum_t O_t}{\sum_t P_t} & \text{if } \bar{P} < \bar{O} \end{cases}$. NMBF, which takes values in

$(-1, +\infty)$, has the advantage of both avoiding inflation due to low values of observations and overcome the asymmetry problem between overestimation and underestimation. The Weighted Normalised mean square error of the Normalised Ratios is defined by $WNNR = \frac{\sum_t s_t^2 (1-k_t)^2}{\sum_t s_t k_t}$ while the not weighted one is $NNR = \frac{\sum_t (1-k_t)^2}{\sum_t k_t}$ where $s_t = O_t / \bar{O}$ and $k_t = \exp[-| \ln(P_t / O_t) |]$. Both WNNR and NNR assume values in $(0, +\infty)$, and WNNR takes more into account the peaks of observed data.

The distributions over the prediction sites of these performance indexes are visualized by boxplots in Fig. 1. First of all, considering all the indexes but especially

Table 2 Models' performance and complexity.

Model	Number of parameters	Max Matrix dimension	Prediction computing time (sec.)	Prediction capability
A1	11	$d \times d$	1906	**
A2	14	$T \times T$	2064 ^a	***
A3-1	15	$dT \times dT$	2062 ^a	***
A3-2	16	$dT \times dT$	2074 ^a	**
B	13	$d \times d$	3600	**
C	12	$d \times d$	4922	*

^a Referred to a different estimation and prediction procedure carried out using RandomFields R package.

RMSE and the Pearson coefficient, it results that Model C is overall the worst one. Moreover, NMFB seems not able to discriminate clearly between models while the other indexes lead to conclude that Model A1 and B have similar performances. The highest Pearson coefficients and the lowest RMSE can be observed for Model A2 and A3-1. All these considerations are summarized in a qualitative index named prediction capability and reported in the last column of Table 2 where the number of stars increases with the model performance. From a comprehensive overview of Table 2 it is possible to conclude that a richer model structure, like the one of Model B and C with an additional stage for temporal dynamics, does not assure a better performance. Even Model A1, which is the simplest one, is equivalent to Model B in terms of prediction capability. The best performance is awarded to Model A2 and A3-1 which, in turn, are more computational demanding due to their more complex covariance structure and the high-dimensional matrices involved.

4 Conclusions

In this work we compare various spatio-temporal models that differ for the hierarchical structure and the covariance function form. The comparison procedure takes into account the intrinsic model complexity and the prediction capability, evaluated using some performance indexes, as well as the computational complexity which is strictly related to the dimension of the involved matrices. With reference to Piemonte PM₁₀ data, a global evaluation, based on this set of criteria, leads to conclude that for modelling the complex spatio-temporal dynamics of air pollution, it is better, in terms of prediction performance, to consider the interaction between space and time (in a separable or nonseparable form) instead of modelling them separately with different hierarchical stages. This choice however has a drawback in terms of computational complexity. In particular, the nonseparable models are computationally demanding due to the massive spatio-temporal matrices involved. Actually we are working on the development and the optimization of our ad-hoc written code to estimate and predict these models in a computationally feasible framework.

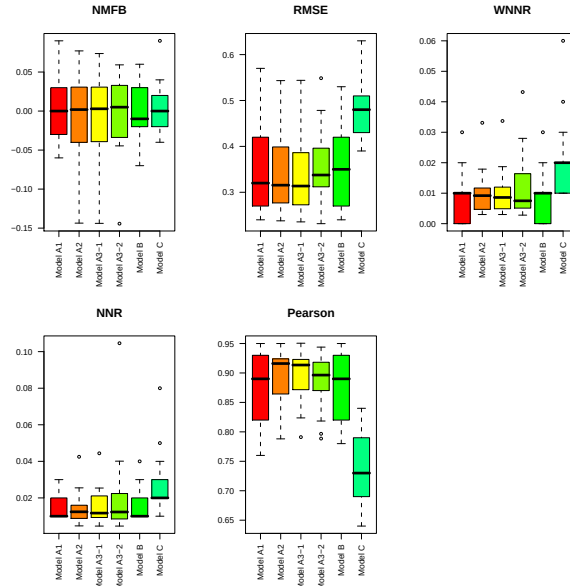


Fig. 1 Indexes boxplot over the validation sites for the 6 models.

References

1. Banerjee, S., Carlin, B., Gelfand, A.: Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability. Chapman and Hall, New York (2004)
2. Gneiting, T.: Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97**, 590–600 (2002)
3. Harrison, R., Stedman, J., Derwent, D.: New directions: Why are PM_{10} concentrations in Europe not falling? *Atmos. Environ.* **42**, 603–606 (2008)
4. Huang, H., Martinez, F., Mateu, J., Montes, F.: Model comparison and selection for stationary space-time models. *Comput. Stat. Data An.* **51**, 4577–4596 (2007)
5. Poli, A., Cirillo, M.: On the use of the normalized mean square error in evaluating dispersion model performance. *Atmos. Environ.* **27**, 2427–2434 (1993)
6. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2010)
7. Wikle, C.K.: Hierarchical models in environmental science. *Int. Stat. Rev.* **71**, 181–199 (2003)
8. Yu, S., Eder, B., Dennis, R., Chu, S., Schwartz, S.: New unbiased symmetric metrics for evaluation of air quality models. *Atmos. Sci. Lett.* **7**, 26–34 (2006)