Environmetrics Research Article

Received: 10 September 2014, Revised: 1 April 2015, Accepted: 6 April 2015, Published online in Wiley Online Library: 5 May 2015

(wileyonlinelibrary.com) DOI: 10.1002/env.2345

Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy

Crescenza Calculli[a]*, Alessandro Fassò[b], Francesco Finazzi[c], Alessio Pollice[d] and Annarita Turnone[e]

Multivariate spatio-temporal statistical models are deserving for increasing attention for environmental data in general and for air quality data in particular because they can reveal dependencies and spatio-temporal dynamics across multiple variables and can be used to obtain dynamic concentration maps over specified areas.
In this frame, we introduce a multivariate generalization of a known spatio-temporal model referred to as the hidden dynamic geostatistical model. Maximum likelihood parameter estimates are obtained implementing the expectation maximization algorithm and suitably extending the `D-STEM` software, recently introduced for alternative model specifications, allowing to handle multiple variables with heterogeneous spatial support, covariates, and missing data.
A case study illustrates some of the statistical issues typical of a medium complexity problem related to air quality data modeling. Considering air quality and meteorological data over the Apulia region, Italy, the usual approach using meteorological variables as regressors is not possible because these data are observed on different monitoring networks, and preliminary spatial interpolation to co-locate the data is to be avoided. Hence, an eight-variate model is considered for understanding the relations between daily concentrations of particulate matters (PM_{10}) and nitrogen dioxides (NO_2) and six non co-located meteorological variables. Model interpretation is given, and its use for dynamic map construction, uncertainty included, is illustrated. Moreover, some preliminary evidence of the model capability to detect a Saharan dust event is presented. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: multivariate hierarchical models; spatio-temporal random effect models; D-STEM; particulate matters; nitrogen dioxides

1. INTRODUCTION

The European Union has defined a complete set of rules, acknowledged by environmental legislations of all member countries, that seeks to standardize effective control methodologies toward the quantification of the spatial distribution of pollutant concentrations and the evaluation of air quality. Recently, following the adoption of the new Directive 2008/50/EC of the European Parliament and of the Council of the European Union, the Italian Legislative Decree n. 155/2010 has defined new criteria for the use of evaluation methods different from measurements in fixed sites, with particular reference to modeling techniques. Moreover, this Decree envisages the possibility to neglect air pollution excesses due to the transboundary transport of pollutants. In fact, air pollution is not only a local phenomenon, and the importance of Saharan air masses contributions to particulate air pollution in Southern Europe has been largely recognized (Querol *et al.*, 2004; Querol *et al.*, 2004; Amodio *et al.*, 2011). In particular, considering southeastern Italy, Contini *et al.* (2014) acknowledge the influence of Saharan dust on particulate matter concentrations and estimate a 22% increase of the probability of exceeding the daily standard threshold in the presence of the so-called Saharan dust events.

The previous considerations push toward the investigation of advanced statistical models aimed at characterizing and predicting air quality events and assessing policies over specified areas. In this work, we consider a case study based on concentration data of PM_{10} and NO_2

* *Correspondence to: Crescenza Calculli, Italian Institute for Nuclear Physics, Istituto Nazionale di Fisica Nucleare – Bari, Via E. Orabona 4, 70125 Bari, Italy. E-mail: crescenza.calculli@ba.infn.it*

a *Italian Institute for Nuclear Physics, Istituto Nazionale di Fisica Nucleare - Bari, Via E. Orabona 4, 70125 Bari, Italy*

b *Department of Engineering, University of Bergamo, viale Marconi 5, 24044 Dalmine, Bergamo, Italy*

c *Department of Management, Economics and Quantitative Methods, University of Bergamo, via dei Caniana 2, 24127 Bergamo, Italy*

d *Department of Economics and Mathematical Methods, University of Bari, Largo Abbazia S. Scolastica 53, 70124 Bari, Italy*

e *Agenzia Regionale per la Prevenzione e la Protezione dell'Ambiente – ARPA Puglia, Corso Trieste 27, 70126 Bari, Italy*

and on the measurements of six meteorological variables over the Apulia region, Italy, coming from monitoring networks differing in the number and location of monitoring stations and affected by a non negligible amount of missing data. Spatio-temporal statistical modeling aims at revealing dependencies and spatio-temporal dynamics and obtaining daily concentration maps over the study region.

In the last decade, several univariate models for air pollution data have been developed within the hierarchical modeling framework, facing different methodological and applied issues (Sahu *et al.*, 2006; Calder, 2008; McMillan *et al.*, 2010). As far as spatio-temporal pollutant concentration data are concerned, univariate hierarchical models were used in several case studies. In the context of maximum likelihood estimation, the expectation maximization (EM) algorithm has been largely used, for example, Fassò (2013) considers traffic policy assessment, while Smith *et al.* (2003) deal with high percentages of missing particulate matter concentration data. In the Bayesian framework, Cameletti *et al.* (2011) compare six models for PM_{10} in Piedmont (Italy), featuring different levels of complexity either in the hierarchical structure or in the spatio-temporal covariance function. Moreover, univariate spatio-temporal hierarchical models were proposed to account for rural/background and urban/suburban random effects on PM_{10} concentrations (Sahu *et al.*, 2006), to combine monitoring data and the output from a local-scale air pollution model for health risk assessment (Pirani *et al.*, 2013) and to assess the effects of human activity on nitrogen dioxide (NO_2) pollution in European urban areas (Shaddick *et al.*, 2013), to mention a few.

In particular, the well-known univariate spatio–temporal model introduced by Huang and Cressie (1996) is referred in this paper as the hidden dynamic geostatistical (HDG) model because it may be interpreted as a dynamic random field plus a measurement error. Its importance, both under the frequentist and Bayesian paradigm, is assessed in the comparative studies of Huang *et al.* (2007) and Cameletti *et al.* (2011). Cameletti *et al.* (2013) provide the integrated nested Laplace approximation–stochastic partial differential equations Bayesian estimation of the HDG model, while Katzfuss and Cressie (2011) and Katzfuss and Cressie (2012) consider a special case of the HDG model coupled with fixed rank smoothing to handle large datasets, using both the EM algorithm and a fully Bayesian estimation framework.

Multivariate space-time data are frequently available in the analysis of air quality and meteorological data. For example, in the frame of the EM algorithm, Fassò and Finazzi (2011) and Finazzi *et al.* (2013) consider heterotopic networks. Moreover, De Iaco *et al.* (2012) implement space-time cokriging based on a space-time linear coregionalization model for mapping three airborne pollutants. Using the Bayesian approach, Gelfand *et al.* (2005) utilize a multivariate dynamic spatial model to analyze precipitations and temperatures and Pollice and Jona Lasinio (2010) consider a hierarchical spatio-temporal model for three airborne pollutants.

In this paper we introduce a multivariate generalization of the univariate HDG model and obtain new formulas for the closed form expressions of maximum likelihood parameter estimates using the EM algorithm, handling multiple variables sampled at different monitoring networks and missing data. The code used in the case study is available to the statistical community as a new version of the open source software `D-STEM`, which was discussed in Finazzi and Fassò (2014). It can be downloaded from https://github.com/graspa-group/d-stem.

The remaining part of the paper is structured as follows. Section 2 discusses the multivariate generalization of the HDG model, in particular a discussion of similarities and novelties with respect to existing literature is given in Subsection 2.1. Section 3 gives the new updating formulas of the EM algorithm for the multivariate HDG model. An illustrative example is provided in Section 4, concerning air pollution in the Apulia region, Italy. Model selection is discussed in Subsection 4.1, while results are presented in Subsection 4.2, where a Saharan dust event is also considered. After the conclusions in Section 5, a detailed derivation of the closed form expressions for the EM algorithm is given in the Appendix. In addition, commented `D-STEM` code for multivariate HDG modeling can be found in the Supporting Information, together with parameter estimates for the various alternative HDG model specifications considered.

2. THE MULTIVARIATE HIDDEN DYNAMIC GEOSTATISTICAL MODEL

In this Section we introduce the notation and the general structure of the multivariate generalization of the HDG model. Let $\mathbf{y}(\mathbf{s},t)$ be the q-variate response variable at site $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2$ and discrete time $t = 1,\ldots,T$, where $\mathcal{D}$ is the geographic region of interest, a proper subset of $\mathbb{R}^2$. We say that a spatio-temporal stochastic process is a multivariate hidden dynamic geostatistical (HDG) model if it is defined by

$$\begin{aligned} \mathbf{y}(\mathbf{s},t) &= \mathbf{X}_{\beta}(\mathbf{s},t)\,\boldsymbol{\beta} + \mathbf{X}_{\mathbf{z}}(\mathbf{s},t)\,\mathbf{A}\mathbf{z}(\mathbf{s},t) + \boldsymbol{\varepsilon}(\mathbf{s},t) \\ \mathbf{z}(\mathbf{s},t) &= \mathbf{G}\mathbf{z}(\mathbf{s},t-1) + \boldsymbol{\eta}(\mathbf{s},t) \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects, $\mathbf{A}$ is a diagonal scale matrix while $\mathbf{X}_{\boldsymbol{\beta}}(\mathbf{s},t)$ and $\mathbf{X}_{\mathbf{z}}(\mathbf{s},t)$ are the fixed and random effects design matrix, respectively. Moreover, in model (1), $\boldsymbol{\varepsilon}(\mathbf{s},t)$ is the q-variate Gaussian measurement error independent in space and time, and $\mathbf{z}(\mathbf{s},t)$ is a p-variate latent random variable with Markovian dynamics ruled by matrix $\mathbf{G}$. The p-dimensional innovation $\boldsymbol{\eta}(\mathbf{s},t)$ is a sequence of unit variance Gaussian random fields independent in time, with matrix covariance function $\boldsymbol{\Gamma}$ given by

$$\boldsymbol{\Gamma}(\mathbf{s},\mathbf{s}') = \mathbf{V}\rho(\|\mathbf{s}-\mathbf{s}'\|;\theta) \tag{2}$$

where $\|\mathbf{s}-\mathbf{s}'\|$ is the distance between $\mathbf{s}$ and $\mathbf{s}' \in \mathcal{D}$, the correlation matrix $\mathbf{V}$ gives the cross-covariances of the elements of $\boldsymbol{\eta}(\mathbf{s},t)$ while $\rho(\|\mathbf{s}-\mathbf{s}'\|;\theta)$ is a valid spatial correlation function with range parameter θ.

The parameter set to be estimated is then given by

$$\Psi = \left\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2,\mathbf{g},\mathbf{V},\theta\right\}$$

where, with obvious notation, the aforementioned parameter set components are defined by $\mathrm{Var}(\boldsymbol{\varepsilon}(\mathbf{s},t)) = \mathrm{diag}(\boldsymbol{\sigma}^2) = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_q^2)$, $\mathbf{A} = \mathrm{diag}(\boldsymbol{\alpha}) = \mathrm{diag}(\alpha_1,\ldots,\alpha_q)$, $\mathbf{G} = \mathrm{diag}(\mathbf{g}) = \mathrm{diag}(g_1,\ldots,g_q)$ and, with abuse of notation, we use $\mathbf{V}$ for its strictly lower triangular submatrix. Note that the following restrictions hold: $\alpha_i > 0, \sigma_i^2 > 0, \theta > 0, |g| < 1$ and $\mathbf{V}$ is a strictly positive correlation matrix. Moreover,

note that the spatio-temporal parameters $\{\mathbf{g}, \mathbf{V}, \theta\}$ are not identifiable if $\alpha_i = 0$, hence, small estimated values are related to large variances of the estimates of, for example, θ, and small scale factors α have to be avoided. In D-STEM, these conditions are managed starting with initial values far away from the parameter space boundary and by monitoring these conditions at each iteration step.

2.1. Connections to literature

Going back to novelty aspects related to model (1), it is worth observing that Gelfand *et al.* (2005) considered the Bayesian estimation of a multivariate dynamic spatial model that is similar to our, but based on a fixed random walk dynamics rather than on the more flexible Markovian dynamics used here.

Model (1) is also similar to the fixed rank smoothing spatio-temporal random effects (FRS-STRE) model discussed in Katzfuss and Cressie (2011). Beside being univariate, FRS-STRE model induces spatial correlation through a special choice of term $\mathbf{X_z}$ in (1), which is given by appropriate basis functions. Moreover, they use a white noise process in space and time in place of our $\boldsymbol{\eta}(\mathbf{s}, t)$, that is, they assume a diagonal $\boldsymbol{\Gamma}$ a matrix function in (2). It follows that the FRS-STRE model is a special case of the HDG model, but it is particularly useful when massive datasets are considered and a large $\boldsymbol{\Gamma}$ matrix (if not sparse) may not be easy to handle.

Finally, while the multivariate spatio-temporal model proposed by some of these authors (Fassò and Finazzi, 2011; Finazzi and Fassò, 2014) is characterized by the sum of one or more independent components for the purely temporal dynamics and one or more independent components for the purely spatial correlation component, the present proposal extends the univariate HDG model and provides a first-order auto-regressive spatial component to jointly model both spatial and temporal dependencies. This requires new formulas for the EM algorithm, which are derived through the next sections.

3. ESTIMATION AND MAPPING WITH THE MULTIVARIATE HIDDEN DYNAMIC GEOSTATISTICAL MODEL

In this section, formulas for the EM algorithm and spatial interpolation of multivariate HDG model are presented, while some algebraic details are deferred to the Appendix. In the sequel, in order to simplify notation, we take $q = p$, which is consistent with model (12) of the case study.

In particular, Subsection 3.1 introduces the matrix notation of model (1) able to cover actual and missing observations at n spatial locations and T time points. Moreover, Subsection 3.2 presents the closed form expressions for EM iterations, which are obtained with more details in the Appendix, including the complete-data likelihood function. Finally, explicit formulas for spatial predictions and their variance–covariance matrices are reported in Subsection 3.3.

The EM algorithm is considered stable even when the number of parameters in Ψ is not small, as in the case study of the next section. In fact, the iterations of the EM algorithm have closed form expressions for a good part of the HDG model parameters, which improves numerical stability over Newton–Raphson and similar algorithms. Moreover, Monte Carlo exercises for the older D-STEM model gave assessment of convergence, stability (Fassò and Finazzi, 2011), and sensitivity to missing data (Fassò and Finazzi, 2010) of the EM algorithm in cases that have similar spatio-temporal complexity as the case study of the next section and not much different dimensionality of parameter space and latent factors. Finally, the Hessian matrix and parameter variances are important indirect quantities for assessing estimation stability and model identifiability of large models. Indeed, they are computed in the `D-STEM` software, following Shumway and Stoffer (2006), Section 6.3, Problem 6.12.

3.1. Matrix representation of the multivariate hidden dynamic geostatistical model

Suppose that each of the q variables $y_i(\mathbf{s}, t), i = 1, \ldots, q$, is observed at a set of spatial locations $\mathcal{S}_i = \{\mathbf{s}_{i,1}, \ldots \mathbf{s}_{i,n_i}\}, i = 1, \ldots, q$ and let $\mathbf{y}_i(\mathcal{S}_i, t) = \left(y_i(\mathbf{s}_{i,1}, t), \ldots, y_i(\mathbf{s}_{i,n_i}, t)\right)$. Then, for each time t, the $n \times 1$ observation vector is $\mathbf{y}_t = \left(\mathbf{y}_1(\mathcal{S}_1, t)', \ldots, \mathbf{y}_q(\mathcal{S}_q, t)'\right)'$, with $n = n_1 + \ldots + n_q$. In general, $\mathcal{S}_i \neq \mathcal{S}_j$, that is, each variable can be observed at a different set of spatial locations. Moreover, vector $\mathbf{y}_t$ may include missing values.

Considering n spatial locations, model (1) can be given in the following representation:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\mu}_t &= \mathbf{X}_t^{\boldsymbol{\beta}} \boldsymbol{\beta} + X_t^{\mathbf{z}} \tilde{\mathbf{A}} \mathbf{z}_t \\ \mathbf{z}_t &= \tilde{\mathbf{G}} \mathbf{z}_{t-1} + \boldsymbol{\eta}_t \end{aligned} \tag{3}$$

with $\mathbf{X}_t^{\boldsymbol{\beta}} = \text{blockdiag}\left(\mathbf{X}_{\boldsymbol{\beta},1}(\mathcal{S}_1, t), \ldots, \mathbf{X}_{\boldsymbol{\beta},q}(\mathcal{S}_q, t)\right)$, $\mathbf{X}_{\boldsymbol{\beta},i}(\mathcal{S}_i, t) = \text{stack}\left(\mathbf{x}_{\boldsymbol{\beta},i}(\mathbf{s}_{i,1}, t), \ldots, \mathbf{x}_{\boldsymbol{\beta},i}(\mathbf{s}_{i,n_i}, t)\right)$, and $\mathbf{x}_{\boldsymbol{\beta},i}(\mathbf{s}, t)$ is a $1 \times b_i$ vector, related to data in Equation(1) by $\mathbf{X}_{\boldsymbol{\beta}}(\mathbf{s}, t) = \text{blockdiag}\left(\mathbf{x}_{\boldsymbol{\beta},1}(\mathbf{s}, t), \ldots, \mathbf{x}_{\boldsymbol{\beta},q}(\mathbf{s}, t)\right)$. The operator blockdiag is the block diagonal matrix operator, and stack is the stacking operator. Similarly $\mathbf{X}_t^{\mathbf{z}} = \text{blockdiag}\left(\mathbf{X}_{\mathbf{z},1}(\mathcal{S}_1, t), \ldots, \mathbf{X}_{\mathbf{z},q}(\mathcal{S}_q, t)\right)$, where $\mathbf{X}_{\mathbf{z},i}(\mathcal{S}_i, t) = \text{diag}\left(x_{\mathbf{z},i}(\mathbf{s}_{i,1}, t), \ldots, x_{\mathbf{z},i}(\mathbf{s}_{i,n_i}, t)\right)$. Moreover, vectors $\boldsymbol{\mu}_t$, $\mathbf{z}_t$, $\boldsymbol{\eta}_t$, and $\boldsymbol{\varepsilon}_t$ are defined similarly to $\mathbf{y}_t$. Finally, $\tilde{\mathbf{A}} = \text{blockdiag}\left(\alpha_1 \mathbf{I}_{n_1}, \ldots, \alpha_q \mathbf{I}_{n_q}\right)$ and $\tilde{\mathbf{G}} = \text{blockdiag}\left(g_1 \mathbf{I}_{n_1}, \ldots, g_q \mathbf{I}_{n_q}\right)$, where $\mathbf{I}_{n_i}$ is the identity matrix of dimension n_i. As usual, in the sequel, we will assume that both the random and fixed effect designs are informative enough so that EM formulas are well defined and the Hessian matrix is non-singular.

3.2. Maximum likelihood estimation formulas

In order to define an estimation algorithm capable of handling heterotopic monitoring networks and missing data, the observation vector $\mathbf{y}_t$ is suitably partitioned as $\tilde{\mathbf{y}}_t = \left(\left(\mathbf{y}_t^{(1)}\right)', \left(\mathbf{y}_t^{(2)}\right)'\right)'$, where $\mathbf{y}_t^{(1)} = \mathbf{L}_t\mathbf{y}_t$ is the sub-vector of size $u_t \times 1$ of non-missing data at time t and $\mathbf{L}_t$ is the elimination matrix. Vector $\tilde{\mathbf{y}}_t$ is thus a permutation of $\mathbf{y}_t$ and $\mathbf{y}_t = \mathbf{D}_t\tilde{\mathbf{y}}_t$, with $\mathbf{D}_t$ the permutation matrix. In the sequel, given $\mathbf{b}$ a generic $n \times 1$ vector and $\mathbf{B}$ a generic $n \times n$ matrix, $\mathbf{b}^{(1)}$ and $\mathbf{B}^{(1)}$ will stand for $\mathbf{L}_t\mathbf{b}$ and $\mathbf{L}_t\mathbf{B}\mathbf{L}_t'$, respectively. On the other hand, if $\mathbf{B}$ is a $n \times m$ matrix, then $\mathbf{B}^{(1)} = \mathbf{L}_t\mathbf{B}$. Finally, $\mathbf{0}_n$ and $\mathbf{0}_{n\times m}$ will be used to define the $n \times 1$ vector and the $n \times m$ matrix of all zeros, respectively.

Starting from initial values $\Psi^{\langle 0\rangle}$, the updating formulas at the $(m+1)$-th step of the algorithm are the following:

$$\alpha_i^{\langle m+1\rangle} = \frac{\sum_{t=1}^{T} \operatorname{tr}\left[\mathbf{Q}_i^{(1)}\left(\mathbf{y}_t^{(1)} - \mathbf{X}_t^{\boldsymbol{\beta},(1)}\boldsymbol{\beta}^{\langle m\rangle}\right)\left(\mathbf{X}_t^{\mathbf{z},(1)}\mathbf{z}_t^{T,(1)}\right)'\left(\mathbf{Q}_i^{(1)}\right)'\right]}{\sum_{t=1}^{T} \operatorname{tr}\left[\mathbf{Q}_i^{(1)}\mathbf{X}_t^{\mathbf{z},(1)}\left(\mathbf{z}_t^T\left(\mathbf{z}_t^T\right)' + \mathbf{P}_t^T\right)^{(1)}\left(\mathbf{X}_t^{\mathbf{z},(1)}\right)'\left(\mathbf{Q}_i^{(1)}\right)'\right]}, \; i = 1, \ldots, q \tag{4}$$

$$\boldsymbol{\beta}^{\langle m+1\rangle} = \left[\sum_{t=1}^{T}\left(\mathbf{X}_t^{\boldsymbol{\beta},(1)}\right)'\mathbf{X}_t^{\boldsymbol{\beta},(1)}\right]^{-1}\left[\sum_{t=1}^{T}\left(\mathbf{X}_t^{\boldsymbol{\beta},(1)}\right)'\left(\mathbf{y}_t^{(1)} - \left(\mathbf{X}_t^{\mathbf{z}}\tilde{\mathbf{A}}^{\langle m\rangle}\mathbf{z}_t^T\right)^{(1)}\right)\right] \tag{5}$$

$$\left(\sigma_i^2\right)^{\langle m+1\rangle} = \frac{1}{n_i T}\operatorname{tr}\left\{\mathbf{Q}_i\left[\sum_{t=1}^{T}\mathbf{D}_t\begin{pmatrix}\boldsymbol{\Omega}_t^{(1)} & \mathbf{0}_{u_t\times(n-u_t)}\\ \mathbf{0}_{u_t\times(n-u_t)} & \mathbf{R}_{22}^{\langle m\rangle}\end{pmatrix}\mathbf{D}_t'\right]\mathbf{Q}_i'\right\}, \; i = 1, \ldots, q \tag{6}$$

$$g_i^{\langle m+1\rangle} = \frac{\operatorname{tr}\left[\mathbf{Q}_i\mathbf{S}_{10}\mathbf{Q}_i'\right]}{\operatorname{tr}\left[\mathbf{Q}_i\mathbf{S}_{00}\mathbf{Q}_i'\right]}, \; i = 1, \ldots, q \tag{7}$$

$$\boldsymbol{\mu}_0^{\langle m\rangle} = \mathbf{z}_0^T \tag{8}$$

where $\mathbf{z}_t^T = E_{\Psi^{\langle m\rangle}}\left(\mathbf{z}_t \mid \mathbf{Y}^{(1)}\right)$, $\mathbf{P}_t^T = \operatorname{Var}_{\Psi^{\langle m\rangle}}\left(\mathbf{z}_t \mid \mathbf{Y}^{(1)}\right)$ and $\mathbf{P}_{t,t-1}^T = \operatorname{Cov}_{\Psi^{\langle m\rangle}}\left(\mathbf{z}_t, \mathbf{z}_{t-1} \mid \mathbf{Y}^{(1)}\right)$ denote the output of the Kalman smoother while $\mathbf{S}_{10} = \sum_{t=1}^{T}\mathbf{z}_t^T\left(\mathbf{z}_{t-1}^T\right)' + \mathbf{P}_{t,t-1}^T$, $\mathbf{S}_{00} = \sum_{t=1}^{T}\mathbf{z}_{t-1}^T\left(\mathbf{z}_{t-1}^T\right)' + \mathbf{P}_{t-1}^T$, and $\mathbf{S}_{11} = \sum_{t=1}^{T}\mathbf{z}_t^T\left(\mathbf{z}_t^T\right)' + \mathbf{P}_t^T$ (used in (9) in the succeeding text) are the known EM second moments, for example, Shumway and Stoffer (2006). Moreover, matrices $\boldsymbol{\Omega}_t^{(1)}$ and $\mathbf{R}_{22}$ are given in the Appendix, while $\mathbf{Q}_i$ is the elimination matrix that restricts a vector to the elements related to the i-th variable. Also note that $\boldsymbol{\mu}_0$ is not considered as a model parameter, but it is needed to start the Kalman recursion. Finally, the geostatistical model parameters are updated through numerical optimization. In particular

$$\left\{\mathbf{V}^{\langle m+1\rangle}, \theta^{\langle m+1\rangle}\right\} = \arg\max_{\mathbf{V},\theta} T\log\left|\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\right| + \operatorname{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\left(\mathbf{S}_{11} - \mathbf{S}_{10}\tilde{\boldsymbol{G}}' - \tilde{\mathbf{G}}\mathbf{S}_{10}' + \tilde{\mathbf{G}}\mathbf{S}_{00}\tilde{\mathbf{G}}'\right)\right] \tag{9}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is straightforwardly defined, see the Appendix. In order to have a stable algorithm even when the matrices are ill-conditioned, which is not uncommon in large spatial datasets, D-STEM implements the numerical optimization in (9) using the MATLAB command `fminsearch` that is based on a direct search simplex method (Lagarias *et al.* (1998)).

The EM algorithm iterates until convergence, that is, until the observed data log-likelihood or the elements of the model parameters stop changing significantly. Notice that the structure of (4–8) is quite standard when the EM algorithm is adopted, but the closed form expressions provided allow to handle the multivariate setting, unbalanced and/or non-colocated networks and missing data.

3.3. Mapping with the hidden dynamic geostatistical model

Given the maximum likelihood estimates of the parameter set Ψ, predictions of the i-th response at new sites $\mathcal{S}_0$ and time $t = 1, \ldots, T$ are given by the plug-in approach as follows:

$$\hat{\mathbf{y}}_i(\mathcal{S}_0, t) = \mathbf{X}_{\boldsymbol{\beta},i}(\mathcal{S}_0, t)\hat{\boldsymbol{\beta}}_i + \mathbf{X}_{\mathbf{z},i}(\mathcal{S}_0, t)\hat{\mathbf{A}}_i(\mathcal{S}_0)\mathbf{z}_t^{T,i}(\mathcal{S}_0) \tag{10}$$

while the variance–covariance matrix of $\hat{\mathbf{y}}_i(\mathcal{S}_0, t)$ is given by

$$\boldsymbol{\Sigma}_{\hat{\mathbf{y}}_i}(\mathcal{S}_0, t) = \hat{\mathbf{A}}_i(\mathcal{S}_0)^2\mathbf{X}_{\mathbf{z},i}(\mathcal{S}_0, t)\mathbf{P}_{t,t-1}^{T,i}(\mathcal{S}_0)\mathbf{X}_{\mathbf{z},i}(\mathcal{S}_0, t)' \tag{11}$$

where $\mathbf{z}_t^{T,i}(\mathcal{S}_0)$ and $\mathbf{P}_{t,t-1}^{T,i}(\mathcal{S}_0)$ are the Kalman smoother outputs extended over sites $\mathcal{S}_0$ and for the i-th variable, while $\hat{\mathbf{A}}_i(\mathcal{S}_0) = \hat{\alpha}_i\mathbf{I}_{|\mathcal{S}_0|}$. If $\mathcal{S}_0$ overlays the entire region $\mathcal{D}$ as a fine regular grid, $\hat{\mathbf{y}}_i(\mathcal{S}_0, t)$ allows to draw a map and the ordered collection $\hat{\mathbf{Y}}_i(\mathcal{S}_0) = \{\hat{\mathbf{y}}_i(\mathcal{S}_0, 1), \ldots, \hat{\mathbf{y}}_i(\mathcal{S}_0, T)\}$ represents a dynamic map for the i-th variable.

Note that, considering a generic element $s \in \mathcal{S}_0$, if it is not in the estimation dataset, that is, $s \notin \mathcal{S}_i, i = 1, \ldots, q$, then Equation (10) gives an optimal estimate in the sense that

$$\hat{\mathbf{y}}_i(s,t) = E\left(\mathbf{y}_i(s,t)|\mathbf{Y}^{(1)}\right)$$

On the other side, if $s \in \mathcal{S}_i$ for one or more $i = 1, \ldots, q$, then Equation (10) gives smoothed values of the corresponding variables, filtering out the measurement error $\varepsilon(\mathbf{s}, t)$.

4. THE APULIA CASE STUDY

The methodology discussed in the previous sections is applied to air quality data of Apulia, which is a southeastern Italian region, with mostly coastal and highly inhabited territory. The region is characterized by some of the highest industry-related environmental risk areas in Italy (Martuzzi *et al.*, 2002).

We consider concentrations of two serious airborne pollutants over the study area, namely particulate matters (PM_{10}) and nitrogen dioxides (NO_2). Data are the daily concentrations (in $\mu g/m^3$) from non-traffic monitoring stations obtained by the local environmental protection agency (ARPA Puglia), which are 42 and 39 stations for NO_2 and PM_{10}, respectively, and have a spatial coverage depicted in Figure 1 (a)–(b).

Moreover, we consider the measurements of some meteorological variables possibly driving the pollutant diffusion and advection, namely, average temperature (TEMP, in °C), relative humidity (RH, in %), atmospheric pressure (AP, in hPa), wind speed toward East ($WIND_u$, in m/s), toward North ($WIND_v$, in m/s), and overall wind speed ($WIND_s$, in m/s). These data are provided by the Agrometeorological Service of the Apulia region (ASSOCODIPUGLIA) over a network of 58 stations, which is different from ARPA Puglia and is depicted in Figure 1(c). Precipitations are not considered, because of the lack of relevant rain in the period considered.

Finally, a set of time invariant covariates is also used, namely, population counts (*pop*), monitoring station coordinates (*lat* and *lon*, Universal Transverse Mercator coordinates in km) and land elevation (*elev*, in m). Population data are considered as a proxy of urban pollution emissions and are drawn from the LandScan™ ambient population count database, updated in the year 2008 (Bhaduri *et al.*, 2007), which provides global population data with approximately 1 km resolution (30"× 30"). Land elevation data come from the GTOPO30 global digital elevation model with a horizontal grid spacing of 30 arc-seconds (approximately 1 km).

The analysis is focused on $T = 92$ days of the summer 2012, from 1st July to 30th September. Considering all non-traffic stations available along years 2010–2014, Table 1 shows that summer 2012 is characterized by the highest average pollutant concentrations of the five years considered. Moreover, high temperatures and an event of Saharan dust have been recorded, with a large number of monitoring sites exceeding the daily PM_{10} limit value of 50 $\mu g/m^3$. In fact, back trajectory modeling based on Modis satellite images (Draxler and Rolph, 2015) and synoptic analysis indicated an advection of African dust particles over the Apulia region reaching its maximum on 29 September.

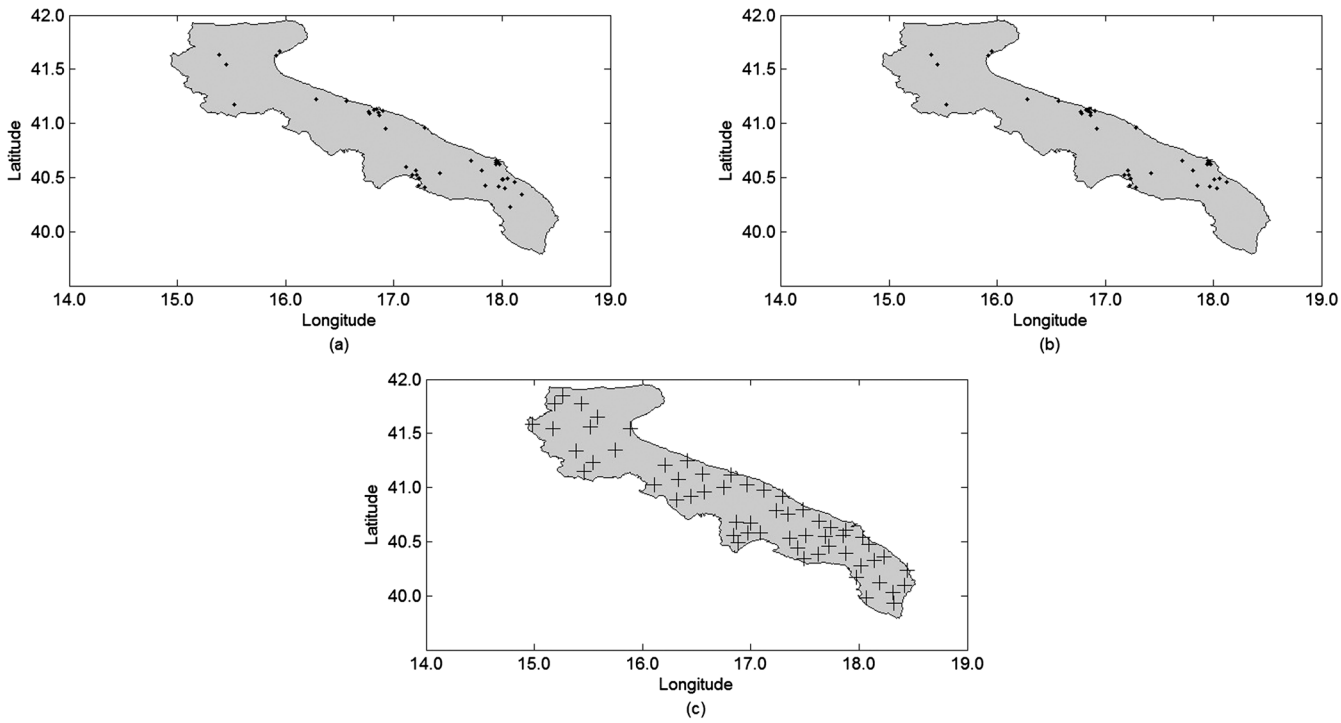

Figure 1. Spatial distribution of the background monitoring stations of the Apulia region air quality network: (a) NO_2 (42 sites); (b) PM_{10} (39 sites). Locations of the meteorological network sites: (c) meteorological stations (58 sites).

Table 1. Summary statistics of the pollutant concentrations of the Apulia air quality network (excluding traffic type sensors). Period: 1st July–30th September, 2010–2014

Year	Pollutant	Number of stations	Mean $(\mu g/m^3)$	Standard deviation $(\mu g/m^3)$	Missing (%)
2010	PM_{10}	27	24.10	9.51	7.35
	NO_2	31	36.22	28.84	8.48
2011	PM_{10}	29	25.88	10.96	9.49
	NO_2	36	38.44	22.03	14.50
2012	PM_{10}	39	27.09	12.45	17.34
	NO_2	42	38.42	29.36	15.65
2013	PM_{10}	35	21.84	8.41	20.27
	NO_2	38	27.93	18.16	16.07
2014	PM_{10}	40	19.09	9.30	27.08
	NO_2	43	24.15	14.84	23.95

Such events do not have a constant evolution in time and involved only part of the area under consideration, because of its geographical configuration. The missing data rate for airborne pollutants is quite relevant in summer 2012 as shown in Table 1. Instead, meteorological variables have at least 98.9% of valid data for each considered variable, and time-invariant covariates have no missing data.

A preliminary data analysis led to log-transform the response variables and to standardize both response variables and covariates. Beyond improving numerical stability, this simplifies the comparisons between the β parameter values across pollutants and meteorological variables. Moreover, it allows to read mean square error and measurement error variances approximately in a 0–1 scale. Using the transformed data, considered the heterotopic nature of the three monitoring networks, both air quality and meteorological variables are treated as components of the response vector $\mathbf{y}$ in the multivariate D-STEM model of Section 2, while the time invariant covariates are considered as fixed effects. In practice, the model used is of the form

$$y_i(\mathbf{s},t) = \boldsymbol{\beta}_i' \boldsymbol{x}_{\beta,i}(\mathbf{s}) + \alpha_i z_i(\boldsymbol{s},t) + \varepsilon_i(\mathbf{s},t) \tag{12}$$

for $i = 1,\ldots,q$, where $q = p$ may be 1, 2, or 8 as discussed in the subsection in the succeeding text. Moreover, the widely used exponential spatial correlation function is used, that is $\rho(\|\mathbf{s}-\mathbf{s}'\|;\theta) = \exp(-\|\mathbf{s}-\mathbf{s}'\|/\theta)$. Maximum likelihood parameter estimates and spatial predictions for these multivariate HDG models are obtained by the method of Section 3. In particular, the code for the calls to the `D-STEM` library are available in the Supporting Information. Model estimations and mapping were performed on an Intel(R) Core(TM) i3- 2370M CPU laptop with 2.40 GHz and 8 GB RAM. For the largest model estimated in Subsection 4.1, the EM algorithm required 42 iterations for convergence, and computing time was $\sim$ 31 min.

4.1. Model selection

In order to select parsimonious models with good performance, two different types of models are considered, namely, *Full* models, which contain all available time-invariant covariates (*elev*, *pop*, *lon*, and *lat*), and *Selected* models, which include only those covariates that are significant at 1% level in the *Full* model. Moreover, different dimensions are considered for each of these two types, namely, eight-variate models (PM_{10} and NO_2 and meteo), bi-variate models (PM_{10} and NO_2), and univariate models (PM_{10} or NO_2).

Model selection is based on predictive performance in cross-validation and performed using the leave one gauge out approach of Fassò *et al.* (2007). This is an iterative procedure that consists in removing one gauge at a time, estimating the model and predicting pollutant

Table 2. Cross-validation mean squared errors in log-standardized scale for two pollutants and eight model specifications

	Full		Selected	
CMSE	NO_2	PM_{10}	NO_2	PM_{10}
8-variate	0.664	0.540	0.668	0.539
2-variate	0.677	0.541	0.679	0.540
1-variate	0.688	0.540	0.689	0.539

CMSE, cross-validation mean squared errors; PM_{10}, particulate matters; NO_2, nitrogen dioxides.

concentrations at the removed gauge for the 92 time points. Cross-validation mean squared errors (CMSE), obtained by averaging the squared residuals for each day and gauge, are given in Table 2 for the two pollutants and the eight model specifications. Although the reported CMSE are somewhat larger than the corresponding measurement errors discussed in the succeeding text, they show that all considered models are quite close with respect to their forecasting capability. This is a known fact in environmental modeling, for example, it is discussed under the equifinality concept by Beven (2001). Hence, considering the slightly smaller sum of CMSEs for the eight-variate *Selected* model, the log-likelihoods of Table 3 and, last but not the least, the richer model interpretation capability, preference is given to the 8-variate *Selected* model. Moreover, the mentioned small variation of CMSE suggests that little is lost with respect to a finer grid model selection.

4.2. Results

Tables 4 and 5 report the maximum likelihood estimates of the parameters in $\boldsymbol{\Psi}$ for the eight-variate *Selected* HDG model, while the estimates for the other seven models of Table 2 are reported in the Supporting Information, Section 2. The estimated $\hat{\boldsymbol{\beta}}$-coefficients in Table 4 suitably describe the physics of air pollution. As expected, land elevation (*elev*) has a significant effect in reducing both PM_{10} and NO_2 concentrations. A positive relationship can be seen between the covariate *pop* and PM_{10}, showing higher concentrations of this pollutant in more populated areas. The signs of estimated coefficients related to meteorological variables are also in line with their physical behavior, for

Table 3. Log-likelihoods of the eight competing model specifications

Model	Full	Selected
8-variate	21,569.325	21,323.094
2-variate	−1, 182.420	−1, 126.678
1-variate NO_2	−853.257	−810.377
1-variate PM_{10}	−225.213	−223.311

PM_{10}, particulate matters; NO_2, nitrogen dioxides.

Table 4. Estimated parameters for the multivariate HDG model. Standard deviations in parentheses

	$\hat{\beta}_{elev}$	$\hat{\beta}_{pop}$	$\hat{\beta}_{lon}$	$\hat{g}_i$	$\hat{\alpha}_i$	$\hat{\sigma}_i^2$
NO_2	$-0.140_{(0.029)}$			$0.715_{(0.018)}$	$1.091_{(0.021)}$	$0.417_{(0.012)}$
PM_{10}	$-0.199_{(0.032)}$	$0.169_{(0.021)}$		$0.841_{(0.011)}$	$0.650_{(0.029)}$	$0.298_{(0.009)}$
TEMP	$-0.119_{(0.022)}$			$0.883_{(0.005)}$	$0.379_{(0.012)}$	$0.004_{(0.000)}$
RH	$-0.068_{(0.018)}$		$0.233_{(0.069)}$	$0.809_{(0.006)}$	$0.451_{(0.013)}$	$0.013_{(0.000)}$
AP	$-0.925_{(0.025)}$			$0.948_{(0.003)}$	$0.236_{(0.013)}$	$0.000_{(0.000)}$
$WIND_u$	$-0.109_{(0.020)}$			$0.609_{(0.016)}$	$0.813_{(0.017)}$	$0.132_{(0.004)}$
$WIND_v$	$-0.090_{(0.022)}$			$0.562_{(0.017)}$	$0.987_{(0.018)}$	$0.232_{(0.006)}$
$WIND_s$	$0.200_{(0.023)}$			$0.583_{(0.013)}$	$1.028_{(0.015)}$	$0.094_{(0.003)}$

HDG, hidden dynamic geostatistical; NO_2, nitrogen dioxides; PM_{10}, particulate matters; RH, relative humidity; AP, atmospheric pressure.

Table 5. Estimated $\hat{\mathbf{V}}$ correlation matrix of the multivariate HDG model. Standard deviations in parentheses

	NO_2	PM_{10}	TEMP	RH	AP	$WIND_u$	$WIND_v$	$WIND_s$
NO_2	1.000	$0.800_{(0.028)}$	$0.001_{(0.079)}$	$-0.139_{(0.067)}$	$-0.717_{(0.084)}$	$0.110_{(0.040)}$	$0.229_{(0.032)}$	$-0.189_{(0.029)}$
PM_{10}		$1.000_{(0.000)}$	$0.243_{(0.151)}$	$-0.110_{(0.133)}$	$-0.547_{(0.206)}$	$0.109_{(0.081)}$	$0.001_{(0.067)}$	$-0.276_{(0.058)}$
TEMP			$1.000_{(0.000)}$	$-0.615_{(0.082)}$	$-0.017_{(0.187)}$	$-0.075_{(0.075)}$	$-0.096_{(0.062)}$	$0.057_{(0.051)}$
RH				$1.000_{(0.000)}$	$0.009_{(0.172)}$	$0.051_{(0.066)}$	$0.006_{(0.055)}$	$-0.018_{(0.045)}$
AP					$1.000_{(0.000)}$	$0.163_{(0.116)}$	$-0.018_{(0.099)}$	$-0.166_{(0.077)}$
$WIND_u$						$1.000_{(0.000)}$	$0.140_{(0.033)}$	$0.079_{(0.029)}$
$WIND_v$							$1.000_{(0.000)}$	$-0.536_{(0.019)}$
$WIND_s$								$1.000_{(0.000)}$

HDG, hidden dynamic geostatistical; NO_2, nitrogen dioxides; PM_{10}, particulate matters; RH, relative humidity; AP, atmospheric pressure.

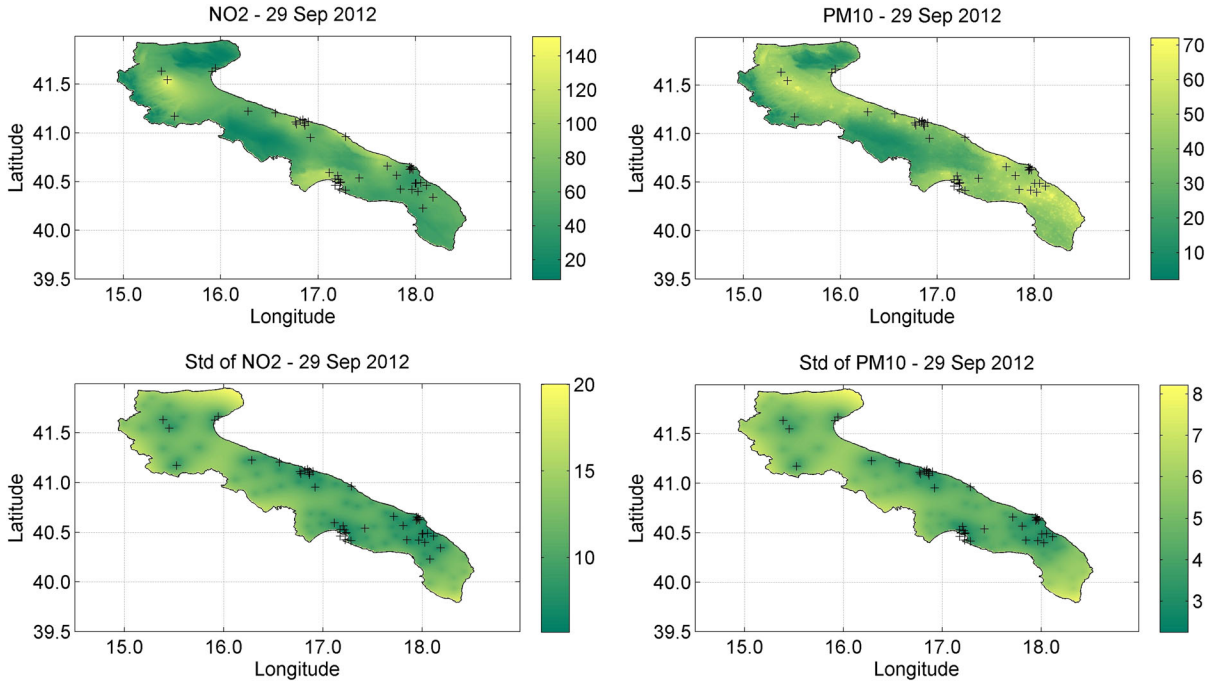

Figure 2. Estimated NO_2 concentration and standard deviation (top and bottom left) and estimated PM_{10} concentration and standard deviation (top and bottom right) for the dust day (29 September 2012) over the Apulia region. The '+' symbol represents the air quality network stations. The scale of all four maps is $\mu g/m^3$

example, the estimated $\hat{\beta}_{elev}$ is negative for the TEMP, RH, and AP responses, representing the tendency of these variables to decrease with growing altitude. Interestingly, longitude is significant only for RH capturing the effect of sea evaporation on ground humidity.

The analysis of the $\hat{g}_i$-values suggests that the eight response variables have different temporal dynamics. As expected, PM_{10} has higher persistence than NO_2 and wind speed the lowest persistence. The estimate of each α_i accounts for the importance of the corresponding spatio-temporal component $z_i(\mathbf{s}, t)$. Notice that all α_i's are far away from zero, which is important for identifiability of the spatio-temporal parameters.

The estimated measurement errors $\hat{\sigma}_i^2$ of Table 4 are quite different and interestingly show that meteorological variables are generally smoother than airborne pollutants. The exception to this, given by both directional wind speeds, $WIND_u$, and $WIND_v$, is related to the known instability of wind direction.

For the airborne pollutants, $\hat{\sigma}_i^2$ is relatively high, indicating that this model is intended to capture the underlying smooth spatial phenomenon rather than point measurements. In particular, $\hat{\sigma}_i^2$ is higher for NO_2 than PM_{10}. This is consistent with expectations, because gaseous nitrogen oxides are more volatile than particulate matters. Moreover, this is consistent with CMSE's results of Table 2.

The estimated parameter $\hat{\theta}$ represents the range of the spatial correlation common to all response variables, after adjusting for the other variables in the model. It amounts to 161 km (SD = 5.0) for this dataset, confirming a general smooth spatial behavior.

The estimated matrix $\hat{\mathbf{V}}$ is reported in Table 5 and represents the cross-correlations of the response variables after adjusting for all the variables in the model including spatial and temporal correlation. It is noticeable on the large positive correlation between NO_2 and PM_{10}, legitimizing the choice of the eight-variate model considering the two pollutants altogether. Considering only correlations significantly different from zero, it is worth to note the large negative correlations of both pollutant concentrations with pressure and, to a lesser extent, with overall wind speed. NO_2 is more affected by AP than PM_{10}, while particulate matter is more severely swept away by winds than the gaseous pollutant.

The mapping capability of the multivariate HDG model is exemplified focusing on the specific dust event of the 29 September discussed previously. Figure 2 shows the estimated daily concentrations of PM_{10} and NO_2, uncertainty included, which are computed as in Equations (10) and (11) for a fine regular grid $\mathcal{S}_0$ with high spatial resolution (1km × 1km approximatively) and are back-transformed into the original scale using corrected anti-log well-known formulas. It is seen that PM_{10} concentrations exceed the threshold of $50\,\mu g/m^3$ over the Salento peninsula (south of Apulia) and in the metropolitan areas of Taranto, Bari, and Foggia. Estimated concentrations of NO_2 are not as high, consistently with the unresponsiveness of this gas to advection of dust particles. As is common in Kriging, uncertainties related to both pollutant maps are larger in the unmonitored areas.

A further example is related to the monitoring station of Campi Salentina, which is located in the southern part of the study area, and is exposed to the considered dust event. In Figure 3, observed concentrations are compared with model-based smoothed values and cross-validation predictions both obtained by Equation (10). As expected, because of the outlying character of particulate matter in these days, there is some model underestimation of PM_{10}. Instead, because of positive cross correlation among pollutants, the model overestimates NO_2. This opposite behavior of the multivariate HDG model for the two pollutants can be considered as the basis to define a statistical method for the detection of dust events.

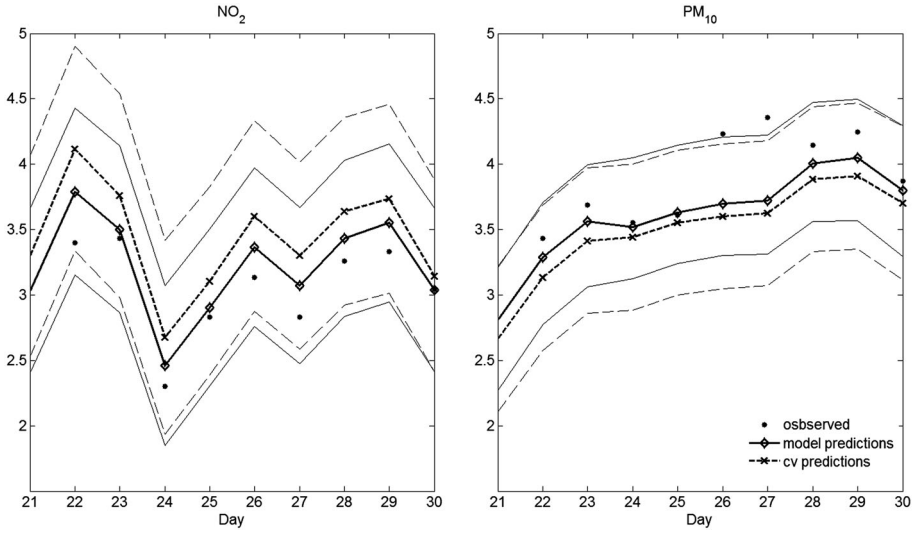

Figure 3. Log-transformed standardized concentrations of the two pollutants at the Campi Salentina monitoring station on 21–30 September 2012: observed, model predictions (95% bounds solid line) and cross-validation predictions (95% bounds dashed line)

5. CONCLUSION AND DISCUSSION

In this work, we obtain formulas for the maximum likelihood estimation of the multivariate HDG model, which is the extension of a spatio-temporal model widely used in environmental statistics. Model specification is based on a first-order auto-regressive spatial component to jointly model spatial and temporal dependencies. The use of the multivariate HDG model is illustrated for a case study concerning air quality over the Apulia region, Italy, considering NO_2 and PM_{10} concentrations and a set of six meteorological variables. The proposed approach allows to handle missing data, unbalanced and non-colocated monitoring networks in a natural way. For that purpose, we extended the `D-STEM` software based on the EM algorithm giving model parameter estimates, dynamic maps, and cross-validation results.

For the case study at hand, the model catches the smooth part of air quality dynamics, as the remaining (measurement) error term is still important. Nonetheless, the interpretation is quite interesting and consistent with previous environmental science. The results on the dust event considered are intriguing and suggest that the possibility to detect dust events on a statistical basis is worth of further research.

Acknowledgements

The authors thank ASSOCODIPUGLIA - Associazione Regionale dei Consorzi di Difesa della Puglia for providing the meteorological data. The analysis utilized the LandScan™ high resolution global population dataset copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under contract DE-AC05-00OR22725 with the US Department of Energy. For this work, F. Finazzi was funded through the FIRB2012 project “Statistical modeling of environmental phenomena: pollution, meteorology, health and their interactions” (RBFR12URQJ).

REFERENCES

Amodio M, Andriani E, Angiuli L, Assennato G, De Gennaro G, Gilio AD, Giua R, Intini M, Menegotto M, Nocioni A, Palmisani J, Perrone MR, Placentino CM, Tutino M. 2011. Chemical characterization of PM in the Apulia region: local and long-range transport contributions to particulate matter. *Boreal Environmental Research* **16**: 251–261.

Beven KJ. 2001. *Rainfall-runoff modelling*. Wiley: New York.

Bhaduri B, Bright E, Coleman P, Urban M. 2007. Landscan USA: a high resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal* **69**: 103–117.

Calder CA. 2008. A dynamic process convolution approach to modeling ambient particulate matter concentrations. *Environmetrics* **19**(1): 39–48.

Cameletti M, Ignaccolo R, Bande S. 2011. Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* **22**(8): 985–996.

Cameletti M, Lindgren F, Simpson D, Rue H. 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis* **97**(2): 109–131.

Contini D, Cesari D, Donateo A, Chirizzi D, Belosi F. 2014. Characterization of pm10 and pm2.5 and their metals content in different typologies of sites in south-eastern italy. *Atmosphere* **5**: 435–453.

De Iaco S, Maggio S, Palma M, Posa D. 2012. Towards an automatic procedure for modeling multivariate space-time data. *Computers & Geosciences* **41**: 1–11.

Draxler RR, Rolph GD. 2015. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory), NOAA Air Resources Laboratory, Silver Spring, MD. Available at: http://ready.arl.noaa.gov/HYSPLIT.php.

Fassò A, Finazzi F. 2010. Air quality mapping using the dynamic coregionalization model. *In Proceedings of 45th Scientific Meeting of the Italian Statistical Society,* Padua, 1–8.

Fassò A, Cameletti M, Nicolis O. 2007. Air quality monitoring using heterogeneous networks. *Environmetrics* **18**(3): 245–264.

Fassò A, Finazzi F. 2011. Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics* **22**(6): 735–748.
Fassò A. 2013. Statistical assessment of air quality interventions. *Stochastic Environmental Research and Risk Assessment* **27**(7): 1651–1660.
Finazzi F, Scott EM, Fassò A. 2013. A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. *Journal of The Royal Statistical Society: Series C* **62**(2): 287–308.
Finazzi F, Fassò A. 2014. D-STEM: a software for the analysis and mapping of environmental space-time variables. *Journal of Statistical Software* **62**(6): 1–29.
Gelfand AE, Banerjee S, Gamerman D. 2005. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* **16**(5): 465–479.
Huang HC, Cressie N. 1996. Spatio-temporal prediction of snow equivalent using the Kalman filter. *Computational Statistics and Data Analysis* **22**: 159–175.
Huang HC, Martinez F, Mateu J, Montes F. 2007. Model comparison and selection for stationary space-time models. *Computational Statistics and Data Analysis* **51**: 4577–4596.
Katzfuss M, Cressie N. 2011. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis* **4**(1): 430–446.
Katzfuss M, Cressie N. 2012. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics* **23**(1): 94–107.
Lagarias JC, Reeds JA, Wright MH, Wright PE. 1998. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization* **9**: 112–147.
Martuzzi M, Mitis F, Biggeri A, Terracini B, Bertollini R. 2002. Environment and health status of the population in areas with high risk of environmental crisis in italy. *Epidemilogia e Prevenzione* **26**(6): 1–53.
McMillan N, Holland DM, Morara M, Feng J. 2010. Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics* **21**: 48–65.
Pirani M, Gulliver J, Fuller GW, Blangiardo M. 2013. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *Journal of Exposure Science & Environmental Epidemiology* **27**: 1–9. DOI: 10.1038/jes.2013.85.
Pollice A, Jona Lasinio G. 2010. A multivariate approach to the analysis of air quality in a high environmental risk area. *Environmetrics* **21**(7–8): 741–754.
Querol X, Alastuey A, Rodríguez S, Viana MM, Artínano B, Salvador P, Mantilla E, Santos S, Patier R, De la Rosa J, De la Campa AS, Menéndez M, Gil JJ. 2004. Levels of PM in rural, urban and industrial sites in Spain. *The Science of Total Environment* **334**: 359–376.
Querol X, Alastuey A, Viana MM, Rodríguez S, Artínano B, Salvador P, Do Santos SG, Patier R, Ruiz C, De la Rosa J, De la Campa AS, Menendez M, Gil JJ. 2004. Speciation and origin of PM10 and PM2.5 in Spain. *Aerosol Science* **35**: 1151–1172.
Sahu SK, Gelfand AE, Holland D. 2006. Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural Biological and Environmental Statistics* **11**(1): 61–86.
Smith RL, Kolenikov S, Cox LH. 2003. Spatio-temporal modeling of PM2.5 data with missing values. *Journal of Geophysical Research* **108**(D24). 9004, DOI: 10.1029/2002JD002914.
Shaddick G, Yan H, Vienneau D. 2013. A Bayesian hierarchical model for assessing the impact of human activity on nitrogen dioxide concentrations in Europe. *Environmental and Ecological Statistics* **20**: 553–570. DOI: 10.1007/s10651-012-0234-z.
Shumway R, Stoffer D. 2006. *Time Series Analysis and its Applications, with R Examples*. Springer: New York.

APPENDIX

In this Appendix, we give the detailed derivation of the closed form expressions for the maximum likelihood estimation of model parameters by the expectation maximization (EM) algorithm. The complete-data likelihood function is defined in the first part, and the EM algorithm is subsequently applied in the second part of the Appendix. Notation related to missing data introduced in Section 3.2 is used here extensively.

A.1. Complete-data likelihood function

In order to define the complete-data log-likelihood function, notice that the following Gaussian distributions are implied by the model structure:

$$\begin{aligned}\mathbf{y}_t \mid \mathbf{z}_t &\sim N_n\left(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_\varepsilon\right)\\ \mathbf{z}_t \mid \mathbf{z}_{t-1} &\sim N_n\left(\tilde{\mathbf{G}}\mathbf{z}_{t-1}, \boldsymbol{\Sigma}_\eta\right)\\ \mathbf{z}_0 &\sim N_n\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right)\end{aligned}$$

with variance–covariance matrices given by $\boldsymbol{\Sigma}_\varepsilon = \text{blockdiag}\left(\sigma_1^2\mathbf{I}_{n_1}, \ldots, \sigma_q^2\mathbf{I}_{n_q}\right)$ and $\boldsymbol{\Sigma}_\eta = \left(v_{i,j}\rho\left(\mathbf{H}_{ij}\right)\right)_{i,j=1,\ldots,q}$, where $\mathbf{H}_{ij} = d\left(\mathcal{S}_i, \mathcal{S}_j\right)$ is the distance matrix between the spatial locations in $\mathcal{S}_i$ and in $\mathcal{S}_j$, and the variance–covariance matrix $\boldsymbol{\Sigma}_0$ is known. With this notation, the complete-data log-likelihood function for observations $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$ and $\mathbf{Z} = \{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_T\}$ is given by

$$\begin{aligned}-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right) = T\log\left|\boldsymbol{\Sigma}_\varepsilon\right| + \sum_{t=1}^{T}\mathbf{e}_t'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{e}_t + \log\left|\boldsymbol{\Sigma}_0\right| + \left(\mathbf{z}_0 - \boldsymbol{\mu}_0\right)'\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{z}_0 - \boldsymbol{\mu}_0\right)\\ + T\log\left|\boldsymbol{\Sigma}_\eta\right| + \sum_{t=1}^{T}\left(\mathbf{z}_t - \tilde{\mathbf{G}}\mathbf{z}_{t-1}\right)'\boldsymbol{\Sigma}_\eta^{-1}\left(\mathbf{z}_t - \tilde{\mathbf{G}}\mathbf{z}_{t-1}\right)\end{aligned}$$

where $\mathbf{e}_t = \mathbf{y}_t - \boldsymbol{\mu}_t$. Because of the hierarchical structure of HDG model, $l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right)$ may be written as $l\left(\Psi\right) = l\left(\Psi_1\right) + l\left(\Psi_0\right) + l\left(\Psi_2\right)$, where $\Psi_1 = \left\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \mathbf{g}\right\}$, $\Psi_0 = \left\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right\}$, and $\Psi_2 = \left(\mathbf{V}, \theta\right)$, reducing the optimization dimensionality in the maximization step of the EM algorithm. Because of missing data, the observation vector $\mathbf{y}_t$ is partitioned as $\mathbf{y}_t^{(l)} = \boldsymbol{\mu}_t^{(l)} + \boldsymbol{\varepsilon}_t^{(l)}$, $l = 1, 2$. The variance–covariance matrix of the permuted errors is conformably partitioned, namely,

$$\text{Var}\begin{pmatrix}\boldsymbol{\varepsilon}_t^{(1)}\\ \boldsymbol{\varepsilon}_t^{(2)}\end{pmatrix} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{12}' & \mathbf{R}_{22}\end{pmatrix}$$

Because $\boldsymbol{\Sigma}_\varepsilon$ is diagonal, it follows that $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ are diagonal matrices while $\mathbf{R}_{12} = \mathbf{0}_{(n-u_t)\times u_t}$.

B. Expectation maximization algorithm

The EM algorithm is considered here to obtain the maximum likelihood estimate of the model parameter set Ψ. The algorithm is based on the iteration of an expectation step and a maximization step. The expectation step amounts at computing $Q\left(\Psi, \Psi^{\langle m\rangle}\right) = E_{\Psi^{\langle m\rangle}}\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right) \mid \mathbf{Y}^{(1)}\right]$, where $\mathbf{Y}^{(1)} = \left\{\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_T^{(1)}\right\}$, and the maximization step, namely, $\Psi^{\langle m+1\rangle} = \arg\max_{\Psi} Q\left(\Psi, \Psi^{\langle m\rangle}\right)$, gives the updating formulas for the model parameter estimates. The conditional expectation of the complete-data log-likelihood is given by

$$\begin{aligned} Q\left(\Psi, \Psi^{\langle m\rangle}\right) &= E_{\Psi^{\langle m\rangle}}\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right) \mid \mathbf{Y}^{(1)}\right] \\ &= E_{\Psi^{\langle m\rangle}}\left[E_{\Psi^{\langle m\rangle}}\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right) \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right] \mid \mathbf{Y}^{(1)}\right] \end{aligned}$$

In what follows, let $E\left(\cdot \mid \cdot\right) \equiv E_{\Psi^{\langle m\rangle}}\left(\cdot \mid \cdot\right)$ and $\mathrm{Var}\left(\cdot \mid \cdot\right) \equiv \mathrm{Var}_{\Psi^{\langle m\rangle}}\left(\cdot \mid \cdot\right)$. Moreover, $\boldsymbol{\mu}_0 \equiv \boldsymbol{\mu}_0^{\langle m\rangle}$, $\boldsymbol{\beta} \equiv \boldsymbol{\beta}^{\langle m\rangle}$, $\tilde{\mathbf{A}} \equiv \tilde{\mathbf{A}}^{\langle m\rangle}$, $\tilde{\mathbf{G}} \equiv \tilde{\mathbf{G}}^{\langle m\rangle}$, $\boldsymbol{\Sigma}_\eta \equiv \boldsymbol{\Sigma}_\eta^{\langle m\rangle}$, $\boldsymbol{\Sigma}_\varepsilon \equiv \boldsymbol{\Sigma}_\varepsilon^{\langle m\rangle}$, and $\mathbf{R}_{22} \equiv \mathbf{R}_{22}^{\langle m\rangle}$, that is, vectors and matrices are evaluated using the value of the model parameters at the m-th iteration of the EM algorithm. Considering the inner conditional expectation, the following result holds:

$$\begin{aligned} E\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}\right) \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right] &= T\log\left|\boldsymbol{\Sigma}_\varepsilon\right| + \mathrm{tr}\left[\boldsymbol{\Sigma}_\varepsilon^{-1}\sum_{t=1}^{T}\left(E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)' + \mathrm{Var}\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)\right)\right] \\ &\quad + \log\left|\boldsymbol{\Sigma}_0\right| + \mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{z}_0 - \boldsymbol{\mu}_0\right)\left(\mathbf{z}_0 - \boldsymbol{\mu}_0\right)'\right] \\ &\quad + T\log\left|\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\right| + \mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\sum_{t=1}^{T}\left(\mathbf{z}_t - \tilde{\mathbf{G}}\mathbf{z}_{t-1}\right)\left(\mathbf{z}_t - \tilde{\mathbf{G}}\mathbf{z}_{t-1}\right)'\right] \end{aligned} \tag{A.1}$$

where, recalling that $\mathbf{R}_{12} = \mathbf{0}_{(n-u_t)\times u_t}$, we have

$$E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right) = \mathbf{D}_t\begin{pmatrix}\mathbf{e}_t^{(1)} \\ \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{e}_t^{(1)}\end{pmatrix} = \mathbf{D}_t\begin{pmatrix}\mathbf{e}_t^{(1)} \\ \mathbf{0}_{(n-u_t)\times 1}\end{pmatrix}$$

and

$$\mathrm{Var}\left[\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right] = \mathbf{D}_t\begin{pmatrix}\mathbf{0}_{u_t\times u_t} & \mathbf{0}_{u_t\times(n-u_t)} \\ \mathbf{0}_{(n-u_t)\times u_t} & \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\end{pmatrix}\mathbf{D}_t' = \mathbf{D}_t\begin{pmatrix}\mathbf{0}_{u_t\times u_t} & \mathbf{0}_{u_t\times(n-u_t)} \\ \mathbf{0}_{(n-u_t)\times u_t} & \mathbf{R}_{22}\end{pmatrix}\mathbf{D}_t'$$

Applying the outer conditional expectation to the right hand side of (A.1), it follows that

$$\begin{aligned} Q\left(\Psi, \Psi^{\langle m\rangle}\right) &= T\log\left|\boldsymbol{\Sigma}_\varepsilon\right| + \mathrm{tr}\left(\boldsymbol{\Sigma}_\varepsilon^{-1}\sum_{t=1}^{T}\boldsymbol{\Omega}_t\right) \\ &\quad + \log\left|\boldsymbol{\Sigma}_0\right| + \mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\left\{\left[E\left(\mathbf{z}_0 \mid \mathbf{Y}^{(1)}\right) - \boldsymbol{\mu}_0\right]\left[E\left(\mathbf{z}_0 \mid \mathbf{Y}^{(1)}\right) - \boldsymbol{\mu}_0\right]' + \mathrm{Var}\left(\mathbf{z}_0 \mid \mathbf{Y}^{(1)}\right)\right\}\right] \\ &\quad + T\log\left|\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\right| + \mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\left(\mathbf{S}_{11} - \mathbf{S}_{10}\tilde{\mathbf{G}}' - \tilde{\mathbf{G}}\mathbf{S}_{10}' + \tilde{\mathbf{G}}\mathbf{S}_{00}\tilde{\mathbf{G}}'\right)\right]. \end{aligned} \tag{A.2}$$

In Equation (A.2), matrix $\boldsymbol{\Omega}_t$ is derived as follows:

$$\begin{aligned} \boldsymbol{\Omega}_t &= E\left[E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)' + \mathrm{Var}\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right) \mid \mathbf{Y}^{(1)}\right] \\ &= E\left[E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)E\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right)' \mid \mathbf{Y}^{(1)}\right] + \mathrm{Var}\left(\mathbf{e}_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}\right) \\ &= \mathbf{D}_t\begin{pmatrix}\boldsymbol{\Omega}_t^{(1)} & \boldsymbol{\Omega}_t^{(1)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21} \\ \mathbf{R}_{12}\mathbf{R}_{11}^{-1}\left(\boldsymbol{\Omega}_t^{(1)}\right)' & \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\boldsymbol{\Omega}_t^{(1)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21}\end{pmatrix}\mathbf{D}_t' + \mathbf{D}_t\begin{pmatrix}\mathbf{0}_{u_t\times u_t} & \mathbf{0}_{u_t\times(n-u_t)} \\ \mathbf{0}_{(n-u_t)\times u_t} & \mathbf{R}_{22}\end{pmatrix}\mathbf{D}_t' \\ &= \mathbf{D}_t\begin{pmatrix}\boldsymbol{\Omega}_t^{(1)} & \mathbf{0}_{u_t\times(n-u_t)} \\ \mathbf{0}_{u_t\times(n-u_t)} & \mathbf{R}_{22}\end{pmatrix}\mathbf{D}_t' \end{aligned}$$

In the definition of $\boldsymbol{\Omega}_t$ and in the updating formulas in Section 3.2, matrix $\boldsymbol{\Omega}_t^{(1)}$ is given by

$$\boldsymbol{\Omega}_t^{(1)} = E\left(\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right)E\left(\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right)' + \mathrm{Var}\left(\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right) \tag{A.3}$$

where

$$E\left(\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right) = E\left(\mathbf{y}_t^{(1)} - \boldsymbol{\mu}_t^{(1)} \mid \mathbf{Y}^{(1)}\right) = \mathbf{y}_t^{(1)} - \mathbf{X}_t^{\boldsymbol{\beta},(1)}\boldsymbol{\beta} - \mathbf{X}_t^{\mathbf{z},(1)}\tilde{\mathbf{A}}^{(1)}\mathbf{z}_t^{T,(1)} \tag{A.4}$$

and

$$\mathrm{Var}\left[\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right] = \mathrm{Var}\left[\mathbf{X}_t^{\mathbf{z},(1)}\tilde{\mathbf{A}}^{(1)}\mathbf{z}_t^{(1)} \mid \mathbf{Y}^{(1)}\right] = \left(\tilde{\mathbf{A}}^{(1)}\right)^2 \mathbf{X}_t^{\mathbf{z},(1)}\mathbf{P}_t^{T,(1)}\left(\mathbf{X}_t^{\mathbf{z},(1)}\right)' \tag{A.5}$$

Notice that, because of the hierarchical structure of HDG model, $Q\left(\Psi, \Psi^{\langle m\rangle}\right)$ in (A.2) is composed of three summands related to different parameters, say

$$Q\left(\Psi, \Psi^{\langle m\rangle}\right) = Q_1\left(\Psi_1, \Psi^{\langle m\rangle}\right) + Q_0\left(\Psi_0, \Psi^{\langle m\rangle}\right) + Q_2\left(\Psi_2, \Psi^{\langle m\rangle}\right)$$

This simplifies the maximization step as the closed form updating formulas for $\Psi_1 = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{g}\}$ can be derived by solving

$$\frac{\partial}{\partial \Psi_1} Q_1\left(\Psi_1, \Psi^{\langle m\rangle}\right) = \frac{\partial}{\partial \Psi_1} T \log|\boldsymbol{\Sigma}_\varepsilon| + \mathrm{tr}\left(\boldsymbol{\Sigma}_\varepsilon^{-1}\sum_{t=1}^{T}\boldsymbol{\Omega}_t\right) = \mathbf{0}$$

and they are given in Section 3.2.

