

Linear Smoothing of Noisy Spatial Temporal Series

¹Valter Di Giacinto, ²Ian Dryden, ³Luigi Ippoliti and ⁴Luca Romagnoli

¹Bank of Italy - 67100, L'Aquila, Italy

²School of Mathematical Sciences, University of Nottingham, NG7 2RD, UK

³DQMTE, University 'G. d'Annunzio', Viale Pindaro 42, 65127 Pescara, Italy

⁴Istituto Nazionale di Statistica - U.R. Molise, 86100 Campobasso, Italy

Abstract: The main objective of the study is the development of a linear filter to extract the signal from a spatio-temporal series affected by measurement error. We assume that the evolution of the unobservable signal can be modelled by a space time autoregressive process. In its vectorial form, the model admits a state space representation allowing the direct application of the Kalman filter machinery to predict the unobservable state vector on the basis of the sample information. Having introduced the model, referred to as a STARG+Noise model, the study discusses Maximum Likelihood (ML) parameter estimation assuming knowledge of the variance of the noise process. Consistent method of moments estimators of the autoregressive coefficients and noise variance are also derived, primarily to be used as inputs in the ML estimation procedure. Finally, we consider some simulation studies and an investigation involving sulphur dioxide level monitoring.

Keywords: Gaussian markov random field, image analysis, maximum likelihood estimation, measurement error, Kalman filter, STARMA model, STARG model, state space model

INTRODUCTION

The study is concerned with parameter estimation and smoothing of a spatio-temporal series corrupted by noise. In particular, we assume that $y(\mathbf{s}, t)$ is a spatio-temporal process observed over time t and general location \mathbf{s} within the geographical domain of interest \mathbf{D} at a finite number of points, i.e. we consider $y(\mathbf{s}_i, t)$ at (\mathbf{s}_i, t) for $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$. We assume the observed series is generated by the process

$$y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + x(\mathbf{s}, t) + e(\mathbf{s}, t) \quad (1)$$

where $\mu(\mathbf{s}, t)$ is a deterministic spatial trend, $x(\mathbf{s}, t)$ is a zero mean, L_2 -continuous, second order Gaussian stationary process; $e(\mathbf{s}, t)$ is a white noise measurement error with second moments

$$E[e(s_i, t)e(s_j, t-h)] = \begin{cases} \sigma_e^2, & \text{if } s_i = s_j \text{ and } h = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$E[x(\mathbf{s}_i, t) e(\mathbf{s}_j, t-h)] = 0, \quad \forall \mathbf{s}_i, \mathbf{s}_j, t, h \quad (3)$$

In (1), $x(\mathbf{s}, t)$ represents the variable of interest and constitutes the *state* or the *signal process*. In what follows, we will in general assume that $t \in \mathbf{Z}^+$, the set of positive integers and that $\mathbf{s}=(r, c) \in \mathbf{Z}^2$. Such setting is particularly well suited for the analysis of agricultural field trials or digital image data or, in general, for spatial data that are collected over regular lattices. However it can be extended to deal with point data collected with an irregular pattern, provided a suitable lattice is superimposed on the given area or reasonable methods of spatial weighting are considered^[1].

In this study the state process is assumed to be represented by a STARG $(p, \lambda_0, \lambda_1, \dots, \lambda_p)$ (Space Time Autoregressive Generalised) model^[2,3], stationary around a possibly time-varying deterministic spatial trend. Under this assumption, expression (1) for the state process will be referred to as a STARG+Noise model. Given the structure of the signal and considering assumptions (2)-(3), we also consider the estimation and the identifiability of model parameters.

Under normality assumptions for the signal and noise processes, we discuss maximum likelihood (ML) estimators under general boundary conditions. Following Dryden *et al.*^[4], we also propose the *Space-Time Adjusted Maximum Likelihood Estimator* (ST-AMLE), which is an approximation to the ML estimators, provided the noise variance is known or can be consistently estimated. The ST-AMLE has the appeal of being robust to the specification of the noise process distribution. The remaining portion regards the state space formulation of the model and provides some computational insights to obtain a *block Kalman filter*. Consistent moment based estimators of the autoregressive coefficients and noise variance are also derived, which are primarily used as starting values for the ML estimation procedures. Finally, we carry out a simulation study to assess the performance of the proposed estimators and a demonstration of the methodology using data on sulphur dioxide levels in Milan.

THE STARG+NOISE PROCESS

Specification of the model: In space-time data analysis, one class of models that has been found useful is the Space Time Autoregressive (STAR) family. The characteristics and properties of such models are well documented^[5], but the lack of simultaneous spatial interdependence structure was identified as a weakness of these models. The deficiency is particularly serious in situations where data have been aggregated across both a region and a period of time. A typical example concerns environmental data in which there is quite likely to be correlation between neighboring regions within the same time of observation. Failure to handle this spatial structure at zero time lag will result in a deficiency in the model. To address this problem, we assume that the zero-mean unobserved state process $x(r, c, t)$, $(r, c, t) \in \mathbf{Z}^2 \times \mathbf{Z}^+$ evolves in space and time according to the STARG($p, \lambda_0, \lambda_1, \dots, \lambda_p$) model which involves the introduction of lagged spatial variables at time lag zero^[2,3]

$$x(r, c, t) = \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \sum_{\|\kappa\|=\Delta_k} \frac{\phi_{hl_k m_k}}{N_k(r, c)} x(r-l_k, c-m_k, t-h) + u(r, c, t) \quad (4)$$

where p is the temporal order of the model; λ_h is the spatial order of the h -th autoregressive component; $\kappa = (l_k, m_k)$ is a spatial displacement operator; Δ_k is the distance defining the k -th order set of neighbours, given a distance measure on \mathbf{Z}^2 , $\|\cdot\|$; $\phi_{hl_k m_k}$ is the autoregressive parameter at temporal lag h and spatial lag (l_k, m_k) ; $N_k(r, c)$ is a scale factor equal to the number of k -th order neighbours of spatial unit (r, c) ; $u(r, c, t)$ is a homoskedastic space-time white-noise process with variance σ_u^2 , independent of the x 's.

To ensure identifiability the following restrictions are imposed on the autoregressive coefficients

$$\begin{aligned} \phi_{000} &\equiv 0 \\ \phi_{hl_k m_k} &\equiv \phi_{h, -l_k, -m_k} \end{aligned}$$

Note that when the further condition $\phi_{0l_k m_k} \equiv 0$ is considered, expression (4) represents a STAR model^[5]. Furthermore, when $\lambda_0 = \lambda_1 = \dots = \lambda_p = \lambda$ say, the model will be referred to as a complete STARG(p, λ). We shall write Λ for the total number of autoregressive parameters.

To deal with possibly anisotropic processes the autoregressive coefficients ϕ are allowed to vary not only with lag in time and space but also with direction in space. However, the isotropic specification can be immediately derived from (4) by imposing the further restrictions that $\phi_{hl_k m_k} = \phi_{hk}$ for each (l_k, m_k) such that $\|\kappa\| = \Delta_k, \forall h, k$. As the isotropic specification is nested in the anisotropic one, expressions will be mostly given

with reference to the general anisotropic case. Introducing the displacement operators the model can be given the following compact formulation

$$\phi(\mathbf{B})x_{rct} = u_{rct} \quad (5)$$

where $\phi(\mathbf{B})$ is a polynomial in the three-dimensional displacement operator $\mathbf{B} = [B_r B_c B_t]'$ with expression

$$\phi(\mathbf{B}) = 1 - \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \sum_{\|\kappa\|=\Delta_k} \frac{\phi_{hl_k m_k}}{N_k(r, c)} B_r^{l_k} B_c^{m_k} B_t^h$$

where the individual displacement operators are defined by the following general expressions

$$B_r^l x_{rct} = x_{r-l, ct}; B_c^m x_{rct} = x_{r, c-m, t};$$

$$B_t^h x_{rct} = x_{rc, t-h}$$

The process is stable and hence asymptotically stationary, if the characteristic polynomial $\phi(\mathbf{B})$ has all the roots outside the unit circle. Stable processes can be given the following convergent moving average representation

$$x_{rct} = \frac{1}{\phi(\mathbf{B})} u_{rct} = \psi(\mathbf{B})u_{rct} = \sum_{i,j,h} \psi_{ijh} B_r^i B_c^j B_t^h u_{rct} \quad (6)$$

from which the process autocovariance generating function (AGF) is readily seen to be

$$\begin{aligned} AGF_x &= \sum_{i,j,h} g_{ij}^{(h)} B_r^i B_c^j B_t^h \\ &= \sigma_u^2 \psi(\mathbf{B})\psi(\mathbf{B}^{-1}) = \frac{\sigma_u^2}{\phi(\mathbf{B})\phi(\mathbf{B}^{-1})} \end{aligned}$$

where $g_{ij}^{(h)} = E[x_{rct}x_{r-i, c-j, t-h}]$.

The spectral density of the process is directly linked to its AGF and has expression

$$f_x(\omega) = \frac{\sigma_u^2}{(2\pi)^3 \phi(z)\phi(z^*)} \quad (7)$$

where $\omega = [\omega_r, \omega_c, \omega_t]'$ and $\mathbf{z} = [e^{-i\omega_r}, e^{-i\omega_c}, e^{-i\omega_t}]'$ and \mathbf{z}^* is the complex conjugate of \mathbf{z} .

Considering a finite rectangular portion of the process (4) with R rows and C columns, sorting the lattice cells by lexicographic order and stacking the observations for each time period, the model can be put in the following vector form

$$\mathbf{x}_t = \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \sum_{\|\kappa\|=\Delta_k} \phi_{hl_k m_k} \mathbf{W}_{l_k m_k} \mathbf{x}_{t-h} + \mathbf{u}_t \quad (8)$$

where $\mathbf{x}_t = [x_{1t}, x_{2t}, \dots, x_{i-1t}, x_{it}, x_{i+1t}, \dots, x_{n-1t}, x_{nt}]'$, $x_{it} = x(r, c, t)$ with $i = (r-1)C+c$ ($r = 1, 2, \dots, R; c = 1, 2, \dots, C$), $n = RC$ and \mathbf{u}_t defined accordingly. The $n \times n$ matrices $\mathbf{W}_{l_k m_k}$ have non zero entries $w_{l_k m_k}(i, j) = N_k^{-1}(i)$ if $j = i+l_k C+m_k, 1 \leq i, j \leq n$ and are defined so that $\sum \mathbf{W}_{l_k m_k} = \mathbf{W}_k$. The summation extends over all κ such that $\|\kappa\| = \Delta_k$ and \mathbf{W}_k is the usual k -th order spatial

weights matrix with equal scaled weights^[5]. Consequently, in the isotropic case, model expression simplifies to

$$\mathbf{x}_t = \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \phi_{hk} \mathbf{W}_k \mathbf{x}_{t-h} + \mathbf{u}_t \quad (9)$$

as in Di Giacinto^[2]. For example, the 1st order isotropic STARG(1,1,1) model has three autoregressive parameters: 1st order temporal ϕ_{10} , 1st order spatial ϕ_{01} , and 1st order spatio-temporal ϕ_{11} parameters.

Finally, to complete the definition the *STARG+Noise model* is given by (1), with the noise defined in (2), (3) and $x(r, c, t)$ follows the STARG model of (4).

Identifiability: A general issue in errors-in-variables models refers to the identifiability of the structural parameters based on the likelihood function of the observable process. Now, following a common approach in the literature^[6,7] we will only deal with zero mean purely stochastic Gaussian processes, for which the identification problem involves the assessment of whether the autoregressive coefficients and the shock and error variance parameters σ_u^2 and σ_e^2 can be uniquely recovered from the covariance function or, equivalently, from the spectral density function of the observed process.

In time series analysis, when the unobservable signal has an autoregressive structure the errors-in-variables model is known to be identifiable^[7]. Based on the approach of Anderson and Deistler for rational spectral density functions, the following proposition proves that an analogous result holds also in the spatio-temporal case.

Proposition 1: Let $\delta(\phi(\mathbf{z}))$ indicate the degree of $\phi(\mathbf{z})$ and $\Phi \subset \mathbf{R}^\Lambda$ indicate the admissible parameter space for the vector ϕ of the autoregressive coefficients. The model parameters σ_e^2 , σ_u^2 and $\phi \in \Phi$ are identifiable if $\delta(\phi(\mathbf{z})) > 0$.

Proof: Given assumptions (2)-(4), the spectral density of the observable process, apart from a constant, has expression

$$f_y(\omega) = f_x(\omega) + f_e(\omega) = \sigma_u^2 \phi^1(\mathbf{z}) \phi^1(\mathbf{z}^*) + \sigma_e^2 \quad (10)$$

where $f_y(\omega)$ and $f_x(\omega)$ have the same poles. Since the coefficients of the polynomial $\phi(\mathbf{z})$ can be uniquely recovered from the poles of $f_y(\omega)$, they are identified. Consequently, we can write

$$f_y(\omega) \phi(\mathbf{z}) \phi(\mathbf{z}^*) = \sigma_u^2 + \sigma_e^2 \phi(\mathbf{z}) \phi(\mathbf{z}^*) \quad (11)$$

and equating coefficients corresponding to non zero powers of \mathbf{z} , which exist under the assumption that $\delta > 0$, gives σ_e^2 . Finally, once σ_e^2 is identified, σ_u^2 is given by the usual factorization of $f_y(\omega) - \sigma_e^2$.

Separability conditions: In modelling space-time processes one often makes use of separable covariance structures. Letting $\xi = \text{vec}([\xi_1, \xi_2, \dots, \xi_T])$, its space-time covariance matrix $\Gamma = E(\xi\xi^T)$, has a separable structure when

$$\Gamma = \Sigma_T \otimes \Sigma_S \quad (12)$$

where Σ_S is the covariance matrix of a purely spatial process and Σ_T is the covariance matrix of a purely temporal process. As can be immediately verified, when the process is assumed to evolve over time according to a p -th order autoregressive model, assuming a separable covariance is equivalent to postulating the following linear model for the observed spatial time series

$$\alpha(B)\xi_t = \varepsilon_t$$

where ε_t is a zero mean multivariate white noise process with autocovariance function $E(\varepsilon_t \varepsilon_{t+h}^T) = \Sigma_S$ if $h = 0$ and $E(\varepsilon_t \varepsilon_{t+h}^T) = \mathbf{0}$ elsewhere and $\alpha(B)$ is the scalar autoregressive polynomial $\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p$, where $\alpha_1, \alpha_2, \dots, \alpha_p$ are scalar parameters and B is the usual backward shift operator.

In the case of the STARG process, the VAR (Vector AutoRegressive) form has the following expression

$$\mathbf{A}(B)\mathbf{x}_t = \mathbf{u}_t \quad (13)$$

where $\mathbf{A}(B)$ is the matrix polynomial $\mathbf{A}(B) = \mathbf{A}_0 - \mathbf{A}_1 B - \mathbf{A}_2 B^2 - \dots - \mathbf{A}_p B^p$, $E(\mathbf{u}_t \mathbf{u}_t^T) = \sigma_u^2 \mathbf{I}_n$ (\mathbf{I}_n the n -dimensional identity matrix) and

$$\mathbf{A}_0 = \mathbf{I}_n - \sum_{k=0}^{\lambda_0} \sum_{\|\mathbf{k}\|=\Delta_k} \phi_{0l_k m_k} \mathbf{W}_{l_k m_k} \quad \text{for } h = 0$$

$$\mathbf{A}_h = \sum_{k=0}^{\lambda_h} \sum_{\|\mathbf{k}\|=\Delta_k} \phi_{hl_k m_k} \mathbf{W}_{l_k m_k} \quad \text{for } h = 1, 2, \dots, p$$

Pre-multiplying by \mathbf{A}_0^{-1} (the invertibility of \mathbf{A}_0 can always be achieved provided adequate constraints are imposed on model coefficients), the model can be put in the equivalent form

$\tilde{\mathbf{A}}(B) = \tilde{\mathbf{u}}_t$
with $\tilde{\mathbf{A}}(B) = \mathbf{I} - \tilde{\mathbf{A}}_1 B - \tilde{\mathbf{A}}_2 B^2 - \dots - \tilde{\mathbf{A}}_p B^p$, $\tilde{\mathbf{u}}_t = \mathbf{A}_0^{-1} \mathbf{u}_t$ and $\tilde{\mathbf{A}}_h = \mathbf{A}_0^{-1} \mathbf{A}_h$, $h = 1, 2, \dots, p$. Of course, the modelling of temporal interaction by means of a matrix polynomial allows for much more general dependence structures than those that can be dealt with using a separable covariance assumption. However, separable structures can be obtained from the STARG specification by imposing appropriate parameter constraints, that can be derived under the assumption that the matrix polynomial $\tilde{\mathbf{A}}(B)$ reduces to a scalar polynomial $\alpha(B)$, implying that the following equality holds

$$\tilde{\mathbf{A}}(B) = \alpha(B)\mathbf{I} \tag{14}$$

Equating coefficients on both sides of equation (13) leads to the following set of matrix equalities $\tilde{\mathbf{A}}_h = \alpha_h \mathbf{I}$, $h = 1, 2, \dots, p$, from which constraints on model coefficients can be derived. To illustrate the issue let us consider the case of an isotropic STARG(1,1,1) with parameters: ϕ_{10} , ϕ_{01} and ϕ_{11} . In this case we have $\tilde{\mathbf{A}}(B) = \mathbf{I} - \tilde{\mathbf{A}}_1 B$ and $\mathbf{A}_1 = \mathbf{A}_0^{-1} \mathbf{A}_1$, where $\mathbf{A}_0 = \mathbf{I} - \phi_{01} \mathbf{W}_1$ and $\mathbf{A}_1 = \phi_{10} \mathbf{I} + \phi_{11} \mathbf{W}_1$.

To generate a separable covariance the following relation must hold

$$\tilde{\mathbf{A}}_1 = \mathbf{A}_0^{-1} \mathbf{A}_1 = \alpha_1 \mathbf{I}$$

which implies $\mathbf{A}_1 = \alpha_1 \mathbf{A}_0$, i.e. matrix \mathbf{A}_1 must be proportional to \mathbf{A}_0 . In this case, imposing proportionality amounts to setting

$$\mathbf{I} - \phi_{01} \mathbf{W}_1 = \phi_{10} \left(\mathbf{I} + \frac{\phi_{11}}{\phi_{10}} \mathbf{W}_1 \right)$$

and consequently, the isotropic STARG(1,1,1) model reduces to a separable model when the three autoregressive coefficients lie on the surface in \mathbb{R}^3 with expression

$$\phi_{01} = -\frac{\phi_{11}}{\phi_{10}} \tag{15}$$

When the previous equality holds, the covariance matrix of the process has precisely the form given by expression (12), with $\Sigma_s = E(\tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i') = \sigma_u^2 (\mathbf{A}_0' \mathbf{A}_0)^{-1}$ and

$$\Sigma_T = \begin{bmatrix} (1-\phi_{10}^2)^{-1} & \phi_{10} & \phi_{10}^2 & \dots & \phi_{10}^{T-1} \\ \phi_{10} & (1-\phi_{10}^2)^{-1} & \phi_{10} & \dots & \phi_{10}^{T-2} \\ \phi_{10}^2 & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \phi_{10} & (1-\phi_{10}^2)^{-1} & \phi_{10} \\ \phi_{10}^{T-1} & \phi_{10}^{T-2} & \dots & \phi_{10} & (1-\phi_{10}^2)^{-1} \end{bmatrix}$$

so that it represents the composition of a univariate first order autoregression in time and a simultaneous first order autoregression in space, equivalent, for lattice data, to a third order Gaussian Markov random field with specific parameter constraints. Note that the separable STARG model is a parsimonious model.

Under these constraints the STARG process reduces to a sequence of time-correlated Gaussian Markov random fields and the choice between the separable and non-separable model specifications becomes essentially an empirical matter that can be subject to hypothesis testing based on the available data. In particular, given the fact that the separable specification is fully nested in the general STARG formulation, it can easily be tested against the unrestricted model by means of the usual likelihood ratio or Wald test criteria.

MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood estimation with toroidal boundary condition: Setting the initial values of the process equal to their unconditional zero mean and assuming that T contiguous observations in time are

available for each location, the STARG($p, \lambda_0, \lambda_1, \dots, \lambda_p$) model can be written as

$$\mathcal{A} \mathbf{x} = \mathbf{u} \tag{16}$$

where $\mathbf{x} = \text{vec}([\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T])$, $\mathbf{u} = \text{vec}([\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T])$,

$$\mathcal{A} = (\mathbf{I} \otimes \mathbf{A}_0) - \sum_{h=1}^p (\mathbf{C}_h \otimes \mathbf{A}_h)$$

with \mathbf{C}_h a $T \times T$ matrix having ones along the h -th lower diagonal and zero elsewhere.

With analogous notation, the STARG+Noise model expression becomes

$$\mathbf{y} = \mathbf{u} + \mathbf{x} + \mathbf{e}$$

with $\mathbf{y} = \text{vec}([\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T])$, $\mathbf{u} = \text{vec}([\mu_1, \mu_2, \dots, \mu_T])$

and $\mathbf{e} = \text{vec}([\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T])$.

The mean component is assumed to be represented by a parametric spatial trend $f(\mathbf{s}, \beta_i)$ of general functional form and possibly time-varying parameters β_i . Assuming a polynomial approximation for f we can set

$$\mu_i = \mathbf{D} \beta_i$$

where β_i is a d -dimensional vector of coefficients and \mathbf{D} is an $n \times d$ suitable design matrix. As a consequence we have

$$\mathbf{u} = \mathcal{D} \beta \tag{17}$$

where $\mathcal{D} = (\mathbf{I}_T \otimes \mathbf{D})$ and $\beta = \text{vec}([\beta_1, \beta_2, \dots, \beta_T])$. When the trend coefficients are fixed over time, i.e. when $\beta_1 = \beta_2 = \dots = \beta_T = \beta$, \mathcal{D} in (17) simplifies to $(\mathbf{1}_T \otimes \mathbf{D})$, with $\mathbf{1}_T$ a T -dimensional vector of ones.

Assuming a Gaussian distribution for both state and noise processes, the log-likelihood function of the STARG+Noise model parameters has expression

$$l(\mathbf{y} | \phi, \sigma_u^2, \sigma_e^2; \mathbf{Y}_0) = -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{\Omega}|) - \frac{1}{2} (\mathbf{y} - \mathbf{u})' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{u}) \tag{18}$$

where $\mathbf{Y}_0 = [\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{p+1}]$, ϕ is the vector of the autoregressive coefficients and

$$\begin{aligned} \mathbf{\Omega} &= E[(\mathbf{y} - \mathbf{u})(\mathbf{y} - \mathbf{u})'] = E(\mathbf{x} \mathbf{x}') + E(\mathbf{e} \mathbf{e}') \\ &= \sigma_u^2 (\mathcal{A} \mathcal{A}')^{-1} + \sigma_e^2 \mathbf{I} \end{aligned}$$

If *toroidal boundary conditions* are assumed, the computational burden is low – of order of $O(nT \log(nT))$ steps – through use of the 3D discrete Fourier transform^[4,8]. Accordingly, \mathbf{C}_h and $\mathbf{W}_{(\cdot)}$ are slightly modified to take into account the required edge corrections^[8]. By the matrix spectral decomposition theorem, we can write

$$\begin{aligned} \mathbf{\Omega}^{-1} &= (\sigma_u^2 \mathbf{P} \mathbf{Q}^{-1} \mathbf{P}' + \sigma_e^2 \mathbf{I})^{-1} \text{ (remembering that } \mathbf{P} \mathbf{P}' = \mathbf{I}) \\ &= \mathbf{P} (\sigma_u^2 \mathbf{Q}^{-1} + \sigma_e^2 \mathbf{I})^{-1} \mathbf{P}' = \mathbf{P} \mathbf{Q} \mathbf{P}' \end{aligned}$$

where $\mathbf{Q} = \text{diag}(q_1, \dots, q_n)$, $q_i = (\sigma_u^2 q_i^{-1} + \sigma_e^2)$ the i -th diagonal element of $(\sigma_u^2 \mathbf{Q}^{-1} + \sigma_e^2 \mathbf{I})$ and q_i is the i -th eigenvalue of $\mathcal{A} \mathcal{A}'$. The log-likelihood will be therefore

$$l(\underline{y}; \phi, \sigma_u^2, \sigma_e^2 | \mathbf{Y}_0) = -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{nT} \log(q_i) - \frac{1}{2} (\underline{y} - \underline{\mu})' \mathbf{P} \mathbf{Q} \mathbf{P}' (\underline{y} - \underline{\mu}) \quad (19)$$

Hence, we can use the 3D discrete Fourier transform to evaluate the log-likelihood and the ML estimates of the parameters can be found by numerical maximization.

EXACT MLE FOR GENERAL BOUNDARY CONDITIONS

In most cases toroidal boundary conditions are used as a computationally convenient approximation of more realistic boundary conditions. In the following, we will consider estimation using Dirichlet or other boundary conditions. Notice that the same results hold for any other boundary conditions that are different from the toroidal one. For a review of the most commonly employed boundary conditions see Moura and Balram^[20].

The log-likelihood function of the observable process (see eq. 18), including the determinant and inverse of a non sparse $nT \times nT$ matrix, can become computationally intractable even for moderate sample dimensions. A direct calculation would need $O((nT)^3)$ steps. However, in the case of Gaussian measurement noise, it is possible to reduce the computational effort to $O((nT)^2)$ steps^[4], by noting that in general, for a noisy model, the log-likelihood can be written as follows:

$$l_Y(y | \phi, \sigma_u^2, \sigma_e^2) = \log f(x, y | \phi, \sigma_u^2, \sigma_e^2) - \log f(x | y, \phi, \sigma_u^2, \sigma_e^2) \quad (20)$$

where $f(x, y | \cdot)$ denotes the joint p.d.f. of X and Y and $f(x | y, \cdot)$ denotes the conditional p.d.f. of X given Y. Since Equation (20) holds for all x , we can evaluate the right-hand side setting x equal to its constant mean μ . With reference to our STARG+Noise model, the first term of the right-hand side is evaluated in $O((nT)^2)$ steps, since $f(\underline{x} = \underline{0}, \underline{y} | \cdot) = f(\underline{x} = \underline{0} | \cdot) f(\underline{y} | \underline{x} = \underline{0}, \cdot)$ which only requires the calculation of the determinant of the $n \times n$ matrix \mathbf{A}_0 .

Furthermore, using the usual results for the conditional multivariate normal distributions, we have that $f(\underline{x} | \underline{y}, \cdot)$ can be evaluated in $O((nT)^2)$ steps. In fact, by putting $\underline{z} = [\underline{x}' \ \underline{y}']'$ and assuming that all the requested inverses exist, it is possible to derive the elements of the partitioned matrix

$$\mathbf{\Omega}_z^{-1} = \begin{bmatrix} \Theta_{xx} & \Theta_{xy} \\ \Theta_{yx} & \Theta_{yy} \end{bmatrix}$$

Hence, the conditional density $f(\underline{x} | \underline{y}, \cdot)$ is also a Gaussian Markov Random Field (GMRF), with $\Theta_{xx} = \Theta_{x|y} = \mathbf{\Omega}_{x|y}^{-1} = (\sigma_u^{-2} \mathcal{A} \mathcal{A}) + \sigma_e^{-2} \mathbf{I}$ and conditional mean $\underline{\mu}_{x|y}$ obtained as solution of the large but sparse and positive definite linear system $\Theta_{xx} \underline{\mu}_{x|y} = -\Theta_{xy} \underline{y}$, with $\Theta_{xy} = \sigma_e^{-2} \mathbf{I}$.

We use the notation EMLE to denote the Exact Maximum Likelihood Estimator.

The adjusted space-time ML estimator: An alternative ML estimator is the Space-time Adjusted Maximum Likelihood Estimator (ST-AMLE). It is obtained as a direct extension of the AMLE derived in the spatial context by Dryden *et al.*^[4]. The main feature of this estimator is that it can be applied under very general conditions on the noise, provided we are able to estimate only the first two measurement noise moments.

From a theoretical standpoint, ST-AMLE is based upon the possibility of expressing the likelihood function of the unobservable state process through the observed variable \underline{Y} . Using (1)-(3), we get

$$l_{\underline{x}}(\underline{y} - \underline{e}; \phi, \sigma_u^2 | \mathbf{Y}_0) = -\frac{nT}{2} \log(2\pi \sigma_u^2) + T \log(|\mathbf{A}_0|) - \frac{1}{2\sigma_u^2} (\underline{y} - \underline{e} - \underline{\mu})' \mathcal{A} \mathcal{A} (\underline{y} - \underline{e} - \underline{\mu}) \\ = -\frac{nT}{2} \log(2\pi \sigma_u^2) + T \log(|\mathbf{A}_0|) - \frac{1}{2\sigma_u^2} [(\underline{y} - \underline{\mu})' \mathcal{A} \mathcal{A} (\underline{y} - \underline{\mu}) + 2 \underline{e}' \mathcal{A} \mathcal{A} (\underline{y} - \underline{\mu}) - \underline{e}' \mathcal{A} \mathcal{A} \underline{e}]$$

Now, since $E[\underline{e}' \mathcal{A} \mathcal{A} \underline{e}] = E[\underline{e}' \mathcal{A} \mathcal{A} \underline{y}] = \sigma_e^2 \text{tr}(\mathcal{A} \mathcal{A})$, it follows that

$$l_{\underline{x}}(\underline{y} - \underline{e}; \phi, \sigma_u^2 | \mathbf{Y}_0) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_u^2) + T \log(|\mathbf{A}_0|) - \quad (21)$$

$$\frac{1}{2\sigma_u^2} [(\underline{y} - \underline{\mu})' \mathcal{A} \mathcal{A} (\underline{y} - \underline{\mu}) - \sigma_e^2 \text{tr}(\mathcal{A} \mathcal{A})] + O_p((nT)^{1/2}) \\ = l_{\underline{x}}^c(\underline{y}; \phi, \sigma_u^2 | \mathbf{Y}_0)$$

which, apart from an adjustment term in the sum of squares component, is equivalent to the likelihood function of a STARG process in the observable variable Y . The first order derivative conditions for the maximum of the *adjusted* log-likelihood are

$$\frac{\partial l_{\underline{x}}^c}{\partial \beta} = \frac{1}{\sigma_u^2} \mathcal{D}' \mathcal{A} \mathcal{A} (\underline{y} - \underline{\mu}) = 0$$

$$\frac{\partial l_{\mathbf{x}}^c}{\partial \phi_{hl_k m_k}} = -T \text{tr}(\mathbf{A}_0^{-1} \mathbf{W}_{l_k m_k}) \delta_{0h} + \frac{1}{\sigma_u^2} \mathbf{y}'_{-hl_k m_k} \mathcal{A} \mathbf{y} - \frac{1}{\sigma_u^2} \sigma_e^2 \text{tr}[(\mathbf{C}_h \otimes \mathbf{W}_{l_k m_k}) \mathcal{A}] = 0$$

$$\frac{\partial l_{\mathbf{x}}^c}{\partial \sigma_u^2} = -\frac{nT}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} [\mathbf{y}' \mathcal{A} \mathcal{A} \mathbf{y} - \sigma_e^2 \text{tr}(\mathcal{A} \mathcal{A})] = 0$$

where $\mathbf{y}_{-hl_k m_k} = (\mathbf{C}_h \otimes \mathbf{W}_{l_k m_k}) \mathbf{y}$; $h = 0, 1, \dots, p$; $k = 0, 1, \dots, \lambda_h$.

If the noise variance is assumed to be known *a priori*, the adjusted likelihood can be directly maximized to obtain the ST-AMLE estimators of remaining parameters. If the noise variance is unknown, a consistent estimator can be employed in its place, leaving the asymptotic properties of the estimators unaffected.

From the first order conditions closed form solutions for the maximum can be obtained for the trend coefficients and innovation variance and these have expressions

$$\hat{\underline{\beta}} = (\mathcal{D}' \mathcal{A}' \mathcal{A} \mathcal{D})^{-1} \mathcal{D}' \mathcal{A}' \mathcal{A} \mathbf{y} \quad (22)$$

$$\hat{\sigma}_u^2 = \frac{1}{nT} [(\mathbf{y} - \underline{\mu})' \mathcal{A}' \mathcal{A} (\mathbf{y} - \underline{\mu}) - \sigma_e^2 \text{tr}(\mathcal{A} \mathcal{A})] \quad (23)$$

Substituting expressions (22) and (23) for $\hat{\underline{\beta}}$ and $\hat{\sigma}_u^2$ in expression (21) we have

$$l_{\mathbf{x}}^c = \text{const} + T \log(|\mathbf{A}_0|) - \frac{nT}{2} \log[(\mathbf{y} - \hat{\underline{\mu}})' \mathcal{A}' \mathcal{A} (\mathbf{y} - \hat{\underline{\mu}}) - \sigma_e^2 \text{tr}(\mathcal{A} \mathcal{A})] \quad (24)$$

where $\hat{\underline{\mu}} = \mathcal{D} \hat{\underline{\beta}}$.

Since $\hat{\underline{\mu}}$ is a non linear function of the AR coefficients ϕ , maximization of (24) is not an easy task, but it can be greatly facilitated by the adoption of the following stepwise optimization procedure:

- * a preliminary estimate $\hat{\underline{\beta}}_0$ of $\hat{\underline{\beta}}$ is computed setting $\mathcal{A} = \mathbf{I}$ in (22), i.e. taking the OLS estimates of the regression of \mathbf{y} on \mathcal{D} ; given the deterministic nature of \mathcal{D} the OLS estimators are consistent and thus provide a valid and easily computed starting point;
- * $\hat{\underline{\mu}}_0 = \mathcal{D} \hat{\underline{\beta}}_0$ is substituted for $\hat{\underline{\mu}}$ in (24) and this is maximized through an iterative search algorithm (e.g. a Newton-Raphson iterative procedure) to give estimates $\hat{\phi}_0$ of $\hat{\phi}$;

- * based on the estimates $\hat{\phi}_0$ the matrix $\hat{\mathcal{A}}_0$ is derived and the GLS estimates $\hat{\underline{\beta}}_1$ are computed from (22);
- * steps 2 and 3 are iterated until convergence is achieved.

Upon convergence, hypothesis tests on model coefficients $\underline{\beta}$ and ϕ can be based on the asymptotic covariance matrices of the estimators

$$\text{VAR}[\hat{\underline{\beta}}] = \sigma_u^2 (\mathcal{D}' \mathcal{A}' \mathcal{A} \mathcal{D})^{-1}$$

$$\text{VAR}[\hat{\phi}] = E \left[\left(-\frac{\partial^2 l_{\mathbf{x}}^c}{\partial \phi \partial \phi'} \right)^{-1} \right]$$

Filtering and smoothing in a state-space framework:

From expressions (1) and (13) it is straightforward to obtain the state space representation of the STARG+Noise model^[9] through the following equations

$$\mathbf{X}_t = \Phi \mathbf{X}_{t-1} + \mathbf{U}_t \quad \text{state equation} \quad (25)$$

$$\mathbf{y}_t = \mathbf{H} \mathbf{X}_t + \mathbf{e}_t \quad \text{measurement equation} \quad (26)$$

where \mathbf{X}_t is the *state vector*, Φ is the *transition matrix*, \mathbf{H} is the *measurement matrix* and \mathbf{U}_t is the *model noise*; their structure is as follows:

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{I} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{U}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{H} = [\mathbf{A}_0^{-1} \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \quad \mathbf{0}].$$

In the framework of state-space models, estimation, filtering and smoothing algorithms are naturally performed by means of the well-known *Kalman filter*^[10]. Following Shumway and Stoffer^[11], Stoffer^[12] also combined the EM algorithm with Kalman filter to derive a recursive procedure for estimating the parameters of a STARMAX model as well dealing with missing data. The implementation of such algorithms is not a difficult task on its own; however, for huge spatial temporal datasets, the filter dimensionality as well as matrix inversions may suggest the adoption of some “tricks” which can reduce the computational burden.

Thus, we will briefly deal with some relevant computational simplifications.

1. Inversion of \mathbf{A}_0 : The task of inverting \mathbf{A}_0 is not always a big problem, but it may require a strong

computational effort for large lattice sizes (say, e.g., 256x256); in these cases we suggest adopting an approximate inversion for \mathbf{A}_0 using the corresponding convergent series. Let us define

$$\mathbf{S} = \sum_{k=0}^{\lambda_0} \sum_{\|\kappa\|=\Delta_k} \phi_{0l_k m_k} \mathbf{W}_{l_k m_k}$$

i.e. $\mathbf{A}_0 = (\mathbf{I} - \mathbf{S})$. Now, we know that

$$\sum_{i=0}^{\infty} \mathbf{S}^i = (\mathbf{I} - \mathbf{S})^{-1} = \mathbf{A}_0^{-1} \Leftrightarrow \lim_{j \rightarrow \infty} \mathbf{S}^j \rightarrow \mathbf{0} .$$

It is often very difficult to find the conditions under which the generic term of the series converges to $\mathbf{0}$, but a sufficient condition which ensures the possibility to adopt the convergent series approximation is^[13,14]

$$\sum_{k=0}^{\lambda_0} \sum_{\|\kappa\|=\Delta_k} |\phi_{0l_k m_k}| < 1 \Rightarrow \lim_{j \rightarrow \infty} \mathbf{S}^j \rightarrow \mathbf{0} .$$

In this case, the following holds

$$\sum_{i=0}^a \mathbf{S}^i \approx \sum_{i=0}^{\infty} \mathbf{S}^i = \mathbf{A}_0^{-1} ,$$

where a may be chosen to be a low integer ($10 \leq a \leq 15$) with a very low loss of precision, since the convergence of \mathbf{S}^i is very fast.

Implementation of a block Kalman filter: For the sake of simplicity, we will treat the problem for the STARG(1,1,1) model, for which the transition matrix is represented by \mathbf{A}_1 ; provided p and $\lambda_0, \lambda_1, \dots, \lambda_p$ are small, the extension to higher temporal order models is straightforward at a low additional computational cost.

If we are prepared to accept a small loss of precision, the following block Kalman filter produces good results in a computationally efficient way. A point worth noting is that the Markovian structure of \mathbf{A}_0 and \mathbf{A}_1 is very helpful for our problem. In fact, given the space-time $n \times T$ data matrix, where the sites \mathbf{s}_i ; $i = 1, 2, \dots, n$ are stacked in lexicographic order, we exploit the Markovian structure of the model to select, for each observation time t , the same block of v sites in \mathbf{y}_t ; $t = 1, 2, \dots, T$ with v a subset of consecutive rows of the lattice such that n/v is a suitably large integer. We then embed our block into a *guard area* given by one upper and one lower row of the lattice, as shown in Fig. 1.

When implemented this way, the block Kalman filter has a dimension $(v+2 \times C)$ instead of $n = R \times C$, where as aforementioned R and C are, respectively, the numbers of lattice rows and columns. Accordingly, we extract the corresponding $(v+2 \times C) \times (v+2 \times C)$ blocks from the matrices \mathbf{A}_0^{-1} and \mathbf{A}_1 , which are involved in the recursion.

Notice that, owing to the structure of \mathbf{A}_0 , its inverse has entries $\mathbf{A}_{0,(i,j)}^{-1} \rightarrow 0$ for (i,j) outside the neighbourhood of site \mathbf{s}_i . For a (32×32) lattice, a spatial

parameter $\phi_{01} = 0.5$ and a block of v sites represented by two rows of the lattice, Fig. 2 shows the complete structure of \mathbf{A}_0^{-1} and its residual part which is



Fig. 1: The block v and the lower and upper guard area for a (8×8) lattice

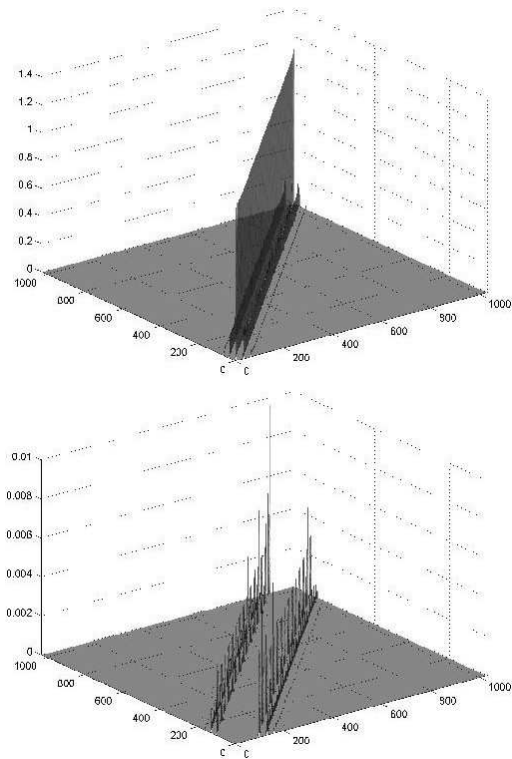


Fig. 2: (Above) Structure of the inverse of the \mathbf{A}_0 matrix, for a (32×32) lattice and a spatial parameter $\phi_{01} = 0.5$. (Below) The entries of \mathbf{A}_0^{-1} which are not involved in the reduced Kalman filter for the same model

not involved in the reduced Kalman filter procedure. As is evident, only a small amount of very small entries are left out and hence

$$f(\mathbf{y}_t(v)|\text{block Kalman } f.) \approx f(\mathbf{y}_t(v)|\text{ordinary Kalman } f.)$$

Kalman filtering to evaluate the likelihood function: The Kalman filter provides a computationally convenient approach for evaluating the likelihood

function for state space models^[10]. If the initial state \mathbf{X}_0 and the innovations \mathbf{U}_t and \mathbf{e}_t are independent multivariate Gaussian, then the distribution of \mathbf{y}_t conditional on \mathbf{X}_t and \mathfrak{S}_{t-1} is

$$\mathbf{y}_t | \mathbf{X}_t, \mathfrak{S}_{t-1} \sim MVN(\mathbf{H} \hat{\mathbf{X}}_{t|t-1}; \mathbf{H} \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}' + \mathbf{R}), t=1, \dots, T \quad (27)$$

where $\mathfrak{S}_{t-1} \equiv (\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_0)$, $\boldsymbol{\Sigma}_{t|t-1}$ is the MSE of the state vector forecast $\hat{\mathbf{X}}_{t|t-1}$ obtained via the Kalman filter and \mathbf{R} is the measurement noise covariance matrix. From (27), following the block Kalman filter strategy described above, it is straightforward to construct the sample log-likelihood

$$\sum_{i=1}^{n/v} \sum_{t=1}^T \log f_i(\mathbf{y}_t | \mathbf{X}_t, \mathfrak{S}_{t-1}) \quad (28)$$

where $f_i(\cdot)$ is the Gaussian density of $\mathbf{y}_t | \mathbf{X}_t, \mathfrak{S}_{t-1}$ evaluated at the i -th block consisting of v sites.

Expression (28) represents an appealing procedure to compute the likelihood, however, the presence of the double summation leads to slow numerical maximization so its use is not recommended in practice for data sets with large spatial dimension.

Generalized moments estimation: When *a priori* information on the magnitude of the measurement error is not available, a consistent estimator of the noise variance is needed to implement the adjusted maximum likelihood estimation procedure. Moreover, given the need for an iterative search procedure to obtain the ST-AMLE of the autoregressive coefficients, the availability of consistent preliminary estimators of these parameters can reduce the computation time substantially, especially when the spatial sample is large. Here, the issue of consistent estimators of AR coefficients and noise variance is considered based on relating the moments of the observable process to the parameters of the underlying process. For the sake of brevity the exposition focuses on the isotropic case, but analogous results hold for the anisotropic model.

The isotropic case: In the case of the isotropic model on \mathbf{Z}^2 and by analogy, in the case of an irregular spatial configuration, the preliminary estimation can be based on the spatio-temporal autocovariance function of the observed process y_{it} , $i = 1, \dots, n$; $t = 1, \dots, T$. Using the definition of Pfeifer and Deutsch^[5] the autocovariance function has expression

$$g_{skv}^y = E(L_s y_{it} L_k y_{it-v})$$

where L_k is the spatial lag operator of order k , defined by the expression

$$L_k y_{it} = \sum_{j=1}^n w_k(i, j) y_{jt}$$

Setting $g_k^{yu} = E(u_{it} L_k y_{it})$, the following set of non linear equations relating the $\Lambda+2$ isotropic model

parameters to the space-time covariances g^y and g^{yu} , can be obtained by $E(L_s y_{it-v} y_{it})$ for $v = 0, 1, \dots, V$; $s = 0, 1, \dots, S_v$

$$g_{s0v}^y = \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \phi_{hk} (g_{skv-h}^y - \sigma_e^2 \delta_{0s} \delta_{0k} \delta_{0v-h}) + g_s^{yu} \delta_{0v} + \sigma_e^2 \delta_{0s} \delta_{0v} \quad (29)$$

where δ_{ij} is the Kronecker's delta.

The system, derived from the autocovariance function of the underlying STARG process $g_{skv}^x = E(L_s x_{it} L_k x_{it-v})$, has $M = \sum_{v=0}^V S_v$ equations in $\Lambda+2$ variables and is expressed in terms of observable covariances g_{skh}^y , where, given assumptions (2)-(3),

$g_{skh}^y = g_{skh}^x + \sigma_e^2$, if $s = k = h = 0$ and $g_{skh}^y = g_{skh}^x$ elsewhere. From the general covariance system, when $v > 0$ and $s \neq 0$ the following sub-system can be extrapolated, consisting of $M_I = \sum_{v=1}^V (S_v - 1)$ linear equations with expression

$$\mathbf{g}_1 = \mathbf{\Gamma}_1 \boldsymbol{\phi} \quad (30)$$

where

$$\mathbf{g}_1 = [g_{101}^y, \dots, g_{S_1 01}^y, g_{102}^y, \dots, g_{S_2 02}^y, \dots, g_{10V}^y, \dots, g_{S_V 0V}^y]'$$

$\mathbf{\Gamma}_1 =$

$$\begin{bmatrix} g_{111}^y & \dots & g_{1\lambda_1}^y & g_{110}^y & \dots & g_{1\lambda_0}^y & \dots & g_{111-p}^y & \dots & g_{1\lambda_{1-p}}^y \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ g_{S_1 11}^y & \dots & g_{S_1 \lambda_1}^y & g_{S_1 10}^y & \dots & g_{S_1 \lambda_0}^y & \dots & g_{S_1 11-p}^y & \dots & g_{S_1 \lambda_{1-p}}^y \\ g_{112}^y & \dots & g_{1\lambda_2}^y & g_{111}^y & \dots & g_{1\lambda_1}^y & \dots & g_{112-p}^y & \dots & g_{1\lambda_{2-p}}^y \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ g_{S_1 12}^y & \dots & g_{S_1 \lambda_2}^y & g_{S_1 11}^y & \dots & g_{S_1 \lambda_1}^y & \dots & g_{S_1 12-p}^y & \dots & g_{S_1 \lambda_{2-p}}^y \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ g_{11V}^y & \dots & g_{1\lambda_V}^y & g_{11V-1}^y & \dots & g_{1\lambda_{V-1}}^y & \dots & g_{11V-p}^y & \dots & g_{1\lambda_{V-p}}^y \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ g_{S_V 1V}^y & \dots & g_{S_V \lambda_V}^y & g_{S_V 1V-1}^y & \dots & g_{S_V \lambda_{V-1}}^y & \dots & g_{S_V 1V-p}^y & \dots & g_{S_V \lambda_{V-p}}^y \end{bmatrix}$$

This is a set of Yule-Walker type equations that do not involve the parameters σ_u^2 and σ_e^2 and provides a set of moment restrictions that the autocovariance function of a STARG process must satisfy.

When exactly Λ moment restrictions are imposed, the system (30) can be directly solved to give the $\boldsymbol{\phi}$ coefficients in terms of the observable autocovariances, obtaining a Yule-Walker type estimator, with expression

$$\tilde{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}_1^{-1} \hat{\mathbf{g}}_1 \quad (31)$$

where $\hat{\mathbf{\Gamma}}_1^{-1}$ and $\hat{\mathbf{g}}_1$ are obtained from $\mathbf{\Gamma}_1$ and \mathbf{g}_1 by replacing theoretical covariances g_{skh}^y with the corresponding sample estimators

$$\hat{g}_{skh}^y = \frac{1}{nT} \sum_{i=1}^n \sum_{t=h+1}^T L_s y_{it} L_k y_{it-h} .$$

For stationary and ergodic processes the autocovariances g_{skh}^y can be estimated consistently from the corresponding sample statistics and consequently the Yule-Walker estimators $\tilde{\phi}$ will be consistent.

If $M_1 > \Lambda$ covariance restrictions are imposed, a generalized method of moments estimator can be derived by minimizing an appropriate distance function $f: \mathbf{R}^{M_1} \rightarrow \mathbf{R}$, i.e. by setting

$$\tilde{\phi} = \arg \min_{\phi} f(\hat{\mathbf{g}}_1 - \hat{\Gamma}_1 \phi) .$$

Following Klajisen and Prucha^[15], we choose the following simple quadratic form for the function $f(\cdot)$

$$\tilde{\phi} = \arg \min_{\phi} (\hat{\mathbf{g}}_1 - \hat{\Gamma}_1 \phi)' (\hat{\mathbf{g}}_1 - \hat{\Gamma}_1 \phi)$$

from which the Generalized Moment Estimator is simply given by the least squares solution of the overidentified system $\hat{\mathbf{g}}_1 = \hat{\Gamma}_1 \phi$, i.e.

$$\tilde{\phi} = (\hat{\Gamma}_1' \hat{\Gamma}_1)^{-1} \hat{\Gamma}_1' \hat{\mathbf{g}}_1 . \tag{32}$$

The generalized moment estimators, exploiting more sample information than the standard Yule-Walker estimators, are expected to be more efficient, although the asymptotic properties of the two estimators and their respective performance in finite samples remain to be investigated.

Setting $v = 1, 2, \dots, V$ and $s = 0$ a second subsystem can be extracted from (30). This system is composed of V equations expressing the noise variance σ_e^2

$$g_{00v}^y = \phi_{v0} (g_{000}^y - \sigma_e^2) + \sum_{(h,k) \neq (v,0)} \phi_{hk} g_{0kv-h}^y$$

in terms of observable covariances and unknown autoregressive coefficients. Replacing the latter with

consistent estimators $\bar{\phi}$, like $\tilde{\phi}$ or $\hat{\phi}$ above, a consistent method of moments estimator can be derived. When only the v -th equation is used the following expression for the estimator gives

$$\tilde{\sigma}_{e,v}^2 = \hat{g}_{000}^y - \bar{\phi}_{10}^{-1} \left(\hat{g}_{00v}^y - \sum_{(h,k) \neq (v,0)} \bar{\phi}_{hk} \hat{g}_{0kv-h}^y \right) = \varphi_v. \tag{33}$$

When a system with $V > 1$ equations is considered, in analogy with the previous case, we have the overidentified system

$$\varphi = \mathbf{t}_v \sigma_e^2$$

where $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_V]'$ and \mathbf{t}_v is a V -dimensional vector with unit elements. Also in this case a

generalized moment estimator can be obtained from the least squares solution of the system, yielding

$$\tilde{\sigma}_e^2 = V^{-1} \sum_{v=1}^V \tilde{\sigma}_{e,v}^2 \tag{34}$$

i.e. the simple arithmetic mean of the estimators $\tilde{\sigma}_{e,v}^2$ for $v = 1, \dots, V$. Although consistent, the method of moments estimators of σ_e^2 has the disadvantage that it is not constrained to yield positive results and this is more likely to happen the smaller is σ_e^2 . The generalized moment estimator, being an average of the moment estimators $\tilde{\sigma}_{e,v}^2$ will be less variable in general and consequently it should reduce the frequency of the occurrence of negative results in finite samples.

The anisotropic case: Defining $\gamma_{ijv}^y = E[y_{ret} y_{r-i,c-j,t-v}]$,

$\gamma_{ij}^{yu} = E[u_{ret} y_{r-i,c-j,t-v}]$ the following set of equations relating the parameters of the anisotropic model to space-time autocovariances can be derived by taking expectations for $i = -I, -I+1, \dots, -1, 0, 1, \dots, I-1, I; j = 0, 1, \dots, J; v = 0, 1, \dots, V$

$$\gamma_{ijv}^y = \sum_{h=0}^p \sum_{k=0}^{\lambda_h} \sum_{\|k\|=\Delta_k} \phi_{hk m_k} (\gamma_{i-l_k, j-m_k, v-h}^y - \sigma_e^2 \delta_{0i-l_k} \delta_{0j-m_k} \delta_{0v-h}) + \gamma_{ij}^{yu} \delta_{0v} + \sigma_e^2 \delta_{0i-l_k} \delta_{0j-m_k} \delta_{0v}$$

The system has the same structure of (29) and the derivation of method of moments estimators of autoregressive coefficients ϕ and σ_e^2 from the appropriate sub-systems follows applying the same line reasoning given for the isotropic case.

SIMULATION STUDIES

The isotropic case: In order to test the performance of the estimators we have selected some example situations. In particular, for 8×8 and 16×16 regular lattices and $T = 30$, we have first simulated 100 samples from an isotropic STARG(1,1,1)+Noise model (9) with parameters: $\phi_{01} = 0.5; \phi_{10} = -0.35; \phi_{11} = 0.45$ and $\sigma_u^2 = 1$. For all the simulations, the noise variance parameter σ_e^2 has been fixed to produce a Log Signal-to-Noise-Ratio (SNR) equal to 5 dB, where SNR is the log of the ratio of the signal variance and the noise variance. In each simulation the measurement noise variance σ_e^2 was fixed as being equal to its true value. Furthermore, in all cases the optimization procedure was performed by setting the starting parameter values at output of the Yule-Walker type procedure outlined before. The means and the standard errors (in brackets) of the estimated parameters are shown in Table 1. As can be seen, the standard deviations decrease as the lattice size becomes larger.

Table 1: The means (and standard errors) of the parameter estimates from 100 simulations of an isotropic STARG(1,1,1)+Noise model. The true parameters are $\phi_{01} = 0.5$; $\phi_{10} = -0.35$; $\phi_{11} = 0.45$ and $\sigma_u^2 = 1$

Lattice Size	8 × 8				16 × 16			
	ϕ_{01}	ϕ_{10}	ϕ_{11}	σ_u^2	ϕ_{01}	ϕ_{10}	ϕ_{11}	σ_u^2
EMLE	0.495 (0.028)	-0.353 (0.032)	0.461 (0.043)	0.999 (0.029)	0.494 (0.017)	-0.343 (0.015)	0.444 (0.024)	1.004 (0.011)
ST-AMLE	0.501 (0.035)	-0.357 (0.036)	0.464 (0.048)	0.955 (0.070)	0.506 (0.018)	-0.351 (0.016)	0.456 (0.026)	0.954 (0.026)
TOROIDAL	0.490 (0.030)	-0.305 (0.030)	0.404 (0.053)	1.033 (0.031)	0.493 (0.021)	-0.326 (0.015)	0.430 (0.024)	1.018 (0.013)
KALMAN	0.447 (0.028)	-0.316 (0.028)	0.400 (0.042)	1.026 (0.019)	0.477 (0.018)	-0.325 (0.014)	0.429 (0.021)	1.016 (0.015)

Table 2: The means (and standard errors) of the parameter estimates from 100 simulations of a homogeneous STARG(1,1,1)+Noise model. For three different lattice sizes, the field was simulated with $T = 20$ temporal observations. The true parameters are $\phi_{01}^v = 0.3$; $\phi_{01}^h = -0.15$; $\phi_{10} = -0.3$; $\phi_{11}^v = 0.4$; $\phi_{11}^h = 0.2$ and $\sigma_u^2 = 1$

<i>EMLE</i>	T = 20		
	<i>Lattice Size</i>		
<i>Parameters</i>	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.248 (0.069)	0.268 (0.043)	0.282 (0.018)
$\phi_{01}^h = -0.15$	-0.127 (0.062)	-0.145 (0.040)	-0.148 (0.019)
$\phi_{10} = -0.3$	-0.297 (0.040)	-0.297 (0.015)	-0.297 (0.008)
$\phi_{11}^v = 0.4$	0.300 (0.105)	0.357 (0.056)	0.384 (0.029)
$\phi_{11}^h = 0.2$	0.172 (0.124)	0.186 (0.049)	0.198 (0.025)
$\sigma_u^2 = 1$	1.003 (0.028)	1.009 (0.014)	1.006 (0.008)
<i>ST-AMLE</i>	T = 20		
<i>Parameters</i>	<i>Lattice Size</i>		
	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.256 (0.073)	0.275 (0.044)	0.288 (0.020)
$\phi_{01}^h = -0.15$	-0.126 (0.064)	-0.148 (0.040)	-0.150 (0.020)
$\phi_{10} = -0.3$	-0.301 (0.043)	-0.304 (0.017)	-0.304 (0.009)
$\phi_{11}^v = 0.4$	0.312 (0.109)	0.367 (0.059)	0.395 (0.031)
$\phi_{11}^h = 0.2$	0.179 (0.123)	0.186 (0.055)	0.198 (0.028)
$\sigma_u^2 = 1$	0.978 (0.030)	0.989 (0.015)	0.993 (0.009)

Table 3: The means (and standard errors) of the parameter estimates from 100 simulations of a homogeneous STARG(1,1,1)+Noise model. For three different lattice sizes, the field was simulated with $T = 50$ temporal observations. The true parameters are $\phi_{01}^v = 0.3$; $\phi_{01}^h = -0.15$; $\phi_{10} = -0.3$; $\phi_{11}^v = 0.4$; $\phi_{11}^h = 0.2$ and $\sigma_u^2 = 1$

<i>EMLE</i>	T = 50		
	<i>Lattice Size</i>		
<i>Parameters</i>	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.250 (0.048)	0.281 (0.023)	0.290 (0.014)
$\phi_{01}^h = -0.15$	-0.133 (0.043)	-0.138 (0.025)	-0.145 (0.009)
$\phi_{10} = -0.3$	-0.300 (0.029)	-0.301 (0.015)	-0.300 (0.006)
$\phi_{11}^v = 0.4$	0.340 (0.073)	0.376 (0.031)	0.387 (0.013)
$\phi_{11}^h = 0.2$	0.169 (0.078)	0.186 (0.040)	0.195 (0.008)
$\sigma_u^2 = 1$	1.006 (0.018)	1.002 (0.010)	1.002 (0.004)
<i>ST-AMLE</i>	T = 50		
<i>Parameters</i>	<i>Lattice Size</i>		
	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.257 (0.054)	0.285 (0.026)	0.297 (0.014)
$\phi_{01}^h = -0.15$	-0.137 (0.043)	-0.144 (0.024)	-0.148 (0.011)
$\phi_{10} = -0.3$	-0.303 (0.030)	-0.305 (0.015)	-0.302 (0.006)
$\phi_{11}^v = 0.4$	0.344 (0.073)	0.379 (0.033)	0.392 (0.015)
$\phi_{11}^h = 0.2$	0.172 (0.076)	0.187 (0.040)	0.197 (0.011)
$\sigma_u^2 = 1$	0.993 (0.020)	0.995 (0.012)	0.995 (0.007)

Table 4: The means (and standard errors) of the parameter estimates from 100 simulations of a homogeneous STARG(1,1,1)+Noise model. For three different lattice sizes, the field was simulated with $T = 100$ temporal observations. The true parameters are $\phi_{01}^v = 0.3$; $\phi_{01}^h = -0.15$; $\phi_{10} = -0.3$; $\phi_{11}^v = 0.4$; $\phi_{11}^h = 0.2$ and $\sigma_u^2 = 1$

<i>EMLE</i>	T = 100		
	<i>Lattice Size</i>		
<i>Parameters</i>	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.259 (0.029)	0.283 (0.024)	0.292 (0.005)
$\phi_{01}^h = -0.15$	-0.134 (0.025)	-0.138 (0.013)	-0.144 (0.004)
$\phi_{10} = -0.3$	-0.305 (0.014)	-0.298 (0.005)	-0.300 (0.004)
$\phi_{11}^v = 0.4$	0.345 (0.044)	0.373 (0.019)	0.390 (0.009)
$\phi_{11}^h = 0.2$	0.168 (0.037)	0.189 (0.014)	0.194 (0.007)
$\sigma_u^2 = 1$	0.998 (0.012)	1.002 (0.004)	1.001 (0.002)
<i>ST-AMLE</i>	T = 100		
	<i>Lattice Size</i>		
<i>Parameters</i>	8×8	16×16	32×32
$\phi_{01}^v = 0.3$	0.259 (0.030)	0.283 (0.025)	0.296 (0.008)
$\phi_{01}^h = -0.15$	-0.139 (0.030)	-0.146 (0.016)	-0.148 (0.006)
$\phi_{10} = -0.3$	-0.302 (0.017)	-0.300 (0.006)	-0.302 (0.003)
$\phi_{11}^v = 0.4$	0.343 (0.050)	0.379 (0.025)	0.390 (0.008)
$\phi_{11}^h = 0.2$	0.179 (0.031)	0.189 (0.018)	0.195 (0.008)
$\sigma_u^2 = 1$	0.990 (0.014)	0.994 (0.007)	0.995 (0.005)

Furthermore, as expected, *EMLE* and *ST-AMLE* are by far the best estimators, while a slight bias is observed for the Kalman and the toroidal cases. This might be explained by the suboptimal block-algorithm strategy and the mis-specification of the boundary conditions.

The anisotropic case: To complete the study of a STARG(1,1,1)+Noise model, we have considered a more complete simulation of fields which are homogeneous in space. In particular, for different lattice sizes and observation time T , we have generated 100 samples from model (8) with temporal parameter $\phi_{10} = -0.3$; spatial parameters $\phi_{01}^v = 0.3$; $\phi_{01}^h = -0.15$; and

spatio-temporal parameters $\phi_{11}^v = 0.4$; $\phi_{11}^h = 0.2$ (where v and h denote vertical and horizontal autocorrelations) and $\sigma_u^2 = 1$. Also in this case, the noise variance parameter σ_e^2 has been fixed to produce a Log Signal-to-Noise-Ratio equal to 5 dB and was considered fixed at its true level. However, to avoid the boundary and the block filter approximations, in this simulation we have only compared the performance of *EMLE* and *ST-AMLE* estimators which are valid for general boundary conditions. In particular, considering three different lattice sizes, Table 2-5 show simulation results, respectively, for $T = 20, 50, 100$. As can be seen, also in this case both *EMLE* and *ST-AMLE* perform very similarly, although, according to our Matlab code, *ST-AMLE* is much faster than *EMLE*.

Simulation based on sulphur dioxide levels in the Milan district: The model and inferential methods outlined previously find a natural field of application in an environmental pollution context. In particular, we have chosen pollutant data from the Milan district in a test bed simulation study for our analysis and statistical modelling. The data set consists of 365 daily averages (from January to December 2001) of Sulphur Dioxide (SO₂) levels at 24 monitoring stations. The data were provided by the Environmental Agency (ARPA) of the Lombardy Region. The coordinate system of the monitoring stations is referred to the Italian national grid system (Gauss-Boaga), which is based on the Universal Transverse Mercator (UTM) projection. To demonstrate the performance of our spatiotemporal model for large data sets, given the irregularly spaced data, we have created a “synthetic” data set representing a temporal sequence of (32×32) regular grids of interest. In particular, using an inverse distance procedure, we have first predicted the process at each spatial location on the grid for each data time. Then, for each spatial location, we have smoothed the data over time using a cubic smoothing spline with smoothing parameter equal to 0.5 (see the matlab command “csaps” in the Matlab spline toolbox). To test the method’s ability to recover the underlying signal (state-variable) \mathbf{x}_t , we have added independent Gaussian measurement noise with variance $\sigma_e^2 = 0.0158$ (SNR equal to 5) to give the noisy data \mathbf{y}_t . This procedure generates a large final data set of 373760 observations arranged in a data matrix of $n = 1024$ rows and $T = 365$ columns. Our aim is to determine how well we can predict the unobservable variable of interest \mathbf{x}_t given the noisy data \mathbf{y}_t . However, note that for temporal prediction purposes, we have taken out the last week ($t = 359, \dots, 365$) from the exploratory analysis.

The study of the temporal pattern of the data set highlights the highest values of SO₂ in the autumn and winter months. Furthermore, for each of the 358 spatial series, locations in the city of Milan show a higher daily average with respect to the other sites.

Table 5: Summary statistics for real and predicted series

Series	Obs. Time	Mean	St. Dev.	Min	Max	MSE	Corr. Coef.
Real – 25 Dec 2001	359	0.000	0.245	-1.142	0.938	-	-
Predicted	359	0.000	0.165	-0.614	0.658	0.009	0.962
Real – 26 Dec 2001	360	0.000	0.229	-0.803	0.903	-	-
Predicted	360	0.000	0.157	-0.528	0.609	0.008	0.964
Real – 27 Dec 2001	361	0.000	0.213	-0.684	0.873	-	-
Predicted	361	0.001	0.150	-0.462	0.569	0.007	0.961
Real – 28 Dec 2001	362	0.000	0.200	-0.675	0.824	-	-
Predicted	362	0.001	0.144	-0.411	0.537	0.006	0.950
Real – 29 Dec 2001	363	0.000	0.189	-0.644	0.758	-	-
Predicted	363	0.001	0.138	-0.371	0.509	0.006	0.938
Real – 30 Dec 2001	364	0.000	0.187	-0.622	0.734	-	-
Predicted	364	-0.001	0.133	-0.339	0.484	0.006	0.933
Real – 31 Dec 2001	365	0.000	0.204	-0.859	0.753	-	-
Predicted	365	-0.001	0.129	-0.313	0.462	0.012	0.878

This is not surprising and may be attributed to a variety of factors as emissions from vehicles, manufacturing and heating systems. Because concentration data are always positive, it is convenient to operate on a logarithmic scale to remove the effect of heteroskedasticity and thereby stabilizing the variance. The exploratory data analysis indicates a spatial trend parametrized as a six-parameter quadratic surface is appropriate and this has been subtracted from the original data to create a zero mean data set.

Variogram based analysis confirms that isotropy is reasonable for the process and the omnidirectional variogram also allowed us to asses the presence of a nugget effect which, as a result of its temporal mean, provided an estimate of the measurement error equal to $\hat{\sigma}_e^2 = 0.0137$. The generalized moment estimator provided an estimate of $\hat{\sigma}_e^2$ equal to 0.0134. As can be seen, although they provide very similar results, both procedures underestimate the real value of the noise variance confirming that its estimation is a very difficult task. The estimation of the measurement noise variance can be carried out through several methods (Olsen^[16], provides a wide range of methods on this topic), but in this study we have chosen the variogram approach following Huang and Cressie^[17] who also propose to estimate the noise as a nugget effect. Given $\hat{\sigma}_e^2$, for a STARG(1,1,1) *ST-AMLE* provided the following parameter estimates (standard errors in brackets) $\hat{\phi}_{01} = 0.489 (0.003)$, $\hat{\phi}_{10} = 0.680 (0.002)$, $\hat{\phi}_{11} = 0.174 (0.003)$, which highlights a strong spatial interaction structure. Finally, we used the state-space formulation and the block Kalman filter to perform the smoothing of the observed series as well as the predictions of the state variable. To show the prediction ability of the filter, Table 5 presents some summary statistics for the zero mean real signal \mathbf{x}_t and the predicted state variable $\hat{\mathbf{x}}_t$.

As can be seen, predictions are able to track the true signal although, as is typical for the Kalman filter,

there is some difficulty in predicting the extreme values of the series.

CONCLUSION

In this study we have advocated the use of a STARG+Noise model for describing the dynamic of noisy random fields in climate and environmental systems. We have shown that the model can be fruitfully used when the available spatio-temporal data is rich both in time and spatial dimensions and when the purpose of the analysis is to provide time-forward predictions at the spatial locations where historical data is available. To deal with huge data sets, we have exploited the Markovian structure of the model to implement a *reduced* Kalman filter to be used for smoothing and prediction aims. Although the state-space formulation also allows for estimating the model parameters, we have suggested an “off-line” inference which can be implemented efficiently. This is critical in problems related to the monitoring of pollutants, where fast algorithms are necessary to issue warnings in a timely manner. For example, as also remarked by Stroud *et al.*^[18], when large data sets are considered the EM algorithms are computationally inefficient because they rely on the repeated inversion of large matrices.

Conversely, it is relatively straightforward to compute the *ST-AMLE* estimator when the noise variance is known. The estimation method is useful when only the moments of the noise are specified or when the exact maximum likelihood estimator is difficult to compute (e.g., for certain non-Gaussian noise distributions). As remarked in Dryden *et al.*^[4], simultaneous estimation of σ_e^2 is very difficult and there are usually confounding problems. There are several methods which can be carried out to estimate the noise variance. For example, following a geostatistical approach, one could estimate σ_e^2 as the time average of the nugget effects obtained from the sample semi-variogram of the spatial series. Considering the *median absolute deviations* (MAD) of

wavelet coefficients, a robust estimator was also provided by Donoho *et al.*^[19]. In this study we also proposed a consistent Yule-Walker type estimator comparable to the ones mentioned above. As well as the STARG model presented in the study extensions are possible. For example, it can be applied to multivariate data (where multiple observations are observed at each point in space and time) or also tailored to allow spatial predictions at unobserved spatial locations.

REFERENCES

1. Stoffer, D.S., 1986. Estimation and identification of space-time ARMAX models in the presence of missing data. *J. Am. Stat. Assoc.*, 81: 395, 762-772.
2. Di Giacinto, V., 1994. Su una generalizzazione dei modelli spazio-temporali autoregressivi media mobile (STARMAG). *Atti XXXVII Riunione SIS, San Remo*, 2: 35-42.
3. Terzi, S., 1995. Maximum likelihood estimation of a generalized STAR(p, λ_p) model. *J. Italian Stat. Soc.*, 4: 377-393.
4. Dryden, I., L. Ippoliti and L. Romagnoli, 2002. Adjusted maximum pseudo-likelihood estimation for noisy gaussian markov random fields. *J. Comput. and Graph. Stat.*, 11: 370-388.
5. Pfeifer, P.E. and S.J. Deutsch, 1980. A three-stage iterative procedure for space-time modeling. *Technometrics*, 22: 35-47.
6. Anderson, B.D.O. and M. Deistler, 1984. Identifiability in dynamic errors-in-variables models. *J. Time Series Analysis*, 5: 1-13.
7. Maravall, A., 1989. Identification in Dynamic Shock-Error models. Berlin, Springer-Verlag.
8. Besag, J., 1977. Errors-in-variables estimation for Gaussian lattice systems. *J. Royal Stat. Soc. B*, 39: 73-78.
9. Ippoliti, L., E. Redfern and L. Romagnoli, 1998. Kalman filter on generalised STARMA model. Technical Report, Department of Statistics, Leeds University.
10. Lütkepohl, H., 1993. Introduction to Multiple Time Series Analysis. 2nd. Edn. Berlin, Springer.
11. Shumway, R.H. and D.S. Stoffer, 1982. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis*, 3: 253-264.
12. Stoffer, D.S., 1985. Maximum likelihood fitting of STARMAX models to incomplete Space-Time series data. *Time Series Analysis: theory and practice 6*, Eds. O.D. Anderson, J.K. Ord and E.A. Robinson, Elsevier Science Publishers B.V. (North-Holland).
13. Harville, D.A., 1997. Matrix Algebra from a Statistician's Perspective. Springer-Verlag, New York.
14. Terzi, S. and G. Gnagnano, 1997. Identifiability conditions for generalised STARMA models. *J. Italian Stat. Soc.*, 6: 257-272.
15. Kelejian, H.H. and I.R. Prucha, 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *Intl. Econ. Rev.*, 40: 509-533.
16. Olsen, S.I., 1993. Estimation of noise in images: An evaluation. *Graphical Models and Image Processing*, 55: 319-323.
17. Huang, H.C. and N. Cressie, 2000. Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, 42: 262-276.
18. Stroud, J.R., P. Müller and B. Sansó, 2001. Dynamic models for spatiotemporal data. *J. Royal Stat. Soc. B*, 63: 673-689.
19. Donoho, D.L., I.M. Johnstone, G. Kerkycharian and D. Picard, 1995. Wavelet shrinkage: Asymptopia? *J. Royal Stat. Soc. B*, 57: 301-369.
20. Moura, J. M. F., and Balram, N. (1992), Recursive Structure of Non-causal Gauss-Markov Random Fields, *IEEE Transactions on Information Theory*, 38, 334-354.