CRRAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

Research Report

Author (s): **Nataly A. Jimenez Monroy, Valderio A. Reisen, Tata Subba Rao & G.Terdik**

Title of the Report: **Modeling and Forecasting PM_{10} concentrations using the Space-Time ARFIMA Model**

Research Report No.: **RR2015-11**

Date: **October 26, 2015**

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

Modeling and Forecasting PM_{10} concentrations using the Space-Time ARFIMA Model

Nátaly A. Jiménez Monroy[a,b], Valdério A. Reisen[a,b], Tata Subba Rao[c,d], G. Terdik[e]

[a] *PPGEA – Universidade Federal do Espírito Santo, Vitória, Brazil.*
[b] *DEST – Universidade Federal do Espírito Santo, Vitória, Brazil.*
[c] *School of Mathematics, University of Manchester, UK.*
[d] *CRRAO AIMSCS, University of Hyderabad Campus, India.*
[e] *University of Debrecen, Hungary.*

Abstract

This paper proposes the Space-Time ARFIMA model (STARFIMA) as an extension of the STARMA class models in order to account for time series with long-memory behavior, a phenomenon that is quite common in the atmospheric pollutant variables. The model is introduced and the semiparametric estimation procedure given in Shimotsu (2007) is suggested to estimate the fractional parameters of the STARFIMA processes. Empirical results from Monte Carlo simulations show the importance of considering not only the spatial dependence between the processes, but also the long memory characteristics of the time series involved. The proposed methodology is applied to PM_{10} daily average concentrations. The comparison of the results obtained using STARFIMA and STARMA models reinforces the usefulness of considering the long-memory characteristics to this particular data set in order to improve the forecasting ability.

Keywords: atmospheric pollution, ARFIMA, forecasting, long-memory, space-time models, STARMA, particulate matter.

---

1. Introduction

The space-time models have shown their usefulness in situations where the data are observed simultaneously in time and space scales. This is the case of the air quality monitoring networks, where the concentration of various pollutants are measured over several spatial locations (monitoring stations) along time (usually at each minute or hour). See, for example

Email address: nataly.monroy@ufes.br (Nátaly A. Jiménez Monroy)

Rouhani et al. (1992), De-Iaco et al. (2003), Huerta et al. (2004), Yu and Chang (2006) and Zeri et al. (2011), among others.

In particular, the class of STARMA (space-time autoregressive moving average) models has been used successfully in several research areas as meteorology (Glasbey and Allcroft (2008)), oceanography (Stoffer (1986), LaValle et al. (2001)), ecology (Reynolds and Madden (1988), Reynolds et al. (1988), Epperson (1994), Epperson (2000)), spatial econometrics (Terzi (1995), Pace et al. (1998), Giacinto (2006)), hydrology (Deutsch and Ramos (1986)), transportation research (Garrido (2000), Kamarianakis and Prastacos (2005)) and imaging (Soni et al. (2004), Crespo et al. (2007)). Nevertheless, its application to atmospheric pollution studies is rare (Antunes and Subba Rao (2006), Glasbey and Allcroft (2008), Monroy et al. (2015)).

In time series modeling is fundamental to analyze the stochastic dependence structure of the series. The class of dependence between the observations determines the model underlying the process. In general, the dependence (or memory) classes are characterized in three forms: short, intermediate and long.

The ARFIMA(p, d, q) (Fractionally Integrated Autoregressive Moving Average) class models, suggested by Granger and Joyeux (1980) and Hosking (1981), has been broadly used due to its capability for capturing the three memory classes previously described in univariate time series. The parameter d assumes real values and characterizes the memory of the process as follows: short ($d = 0$), intermediate ($d < 0$) and long-memory ($d > 0$). The ARMA(p, q) model is a particular case of the ARFIMA(p, d, q) that has the short memory property. The use of long-memory model in pollution variables is the motivation of Reisen et al. (2014) and some references therein.

The aim of this work is to propose the STARFIMA model, as an extension of the STARMA class models, taking into account the long memory of the processes under analysis, a phenomenon which is usually observed in the dispersion dynamics of some atmospheric pollutants. The paper also suggests a two-step procedure to estimate the model. The model and the estimation procedure are the motivations of Section 2. In Section 3 a simulation study is presented in order to show the performance of the model estimates for small sample sizes and other considerations. Section 4 shows an application of the proposed model for forecasting PM_{10} concentrations at the Greater Vitória Region (GVR), Brazil. In addition, the comparison of the fitting and forecasting ability of the proposed model with respect to the STARMA approach is studied. Some final remarks and recommendations are presented in Section 5.

2. The space-time ARFIMA model

Let $\boldsymbol{Z}_t = (Z_{1,t}, Z_{2,t}, \ldots, Z_{N,t})'$ be a vector of observations at N fixed spatial locations on time t. The space-time autoregressive fractionally integrated moving average (STARFIMA) model, denoted as STARFIMA$(p_{\lambda_1,\lambda_2,\ldots,\lambda_p}; \boldsymbol{d}; q_{m_1,m_2,\ldots,m_q})$, is defined as

$$\boldsymbol{\Phi}_{p,\lambda}(B)\boldsymbol{Z}_t = \boldsymbol{\Theta}_{q,m}(B)\mathcal{D}(B)^{-1}\boldsymbol{\varepsilon}_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

where $\boldsymbol{\Phi}_{p,\lambda}(x) = \mathbb{I}_N - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}\boldsymbol{W}_l x^k$, $x \in \mathbb{C}$, represents the autoregressive polynomial with temporal order p and spatial order λ_k, $\boldsymbol{\Theta}_{q,m}(x) = \mathbb{I}_N - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}\boldsymbol{W}_l x^k$, $x \in \mathbb{C}$, represents the moving averaged polynomial with temporal order q and spatial order m_k, $\mathbb{I}_N$ is the $N \times N$ identity matrix and $\boldsymbol{W}_l$ is a nonzero $N \times N$ matrix of weights for the spatial order l with diagonal entries 0 and off-diagonal entries related to the distances between the sites. By definition, $\boldsymbol{W}_0 = \mathbb{I}_N$. Each row of $\boldsymbol{W}_l$ adds up to 1. $\boldsymbol{d} = (d_1, \ldots, d_N)'$ is the difference vector, $\mathcal{D}(B)$ is the $N \times N$ fractional difference operator matrix such that $\mathcal{D}(B) = \text{diag}\left\{(1-B)^{d_1}, (1-B)^{d_2}, \ldots, (1-B)^{d_N}\right\}$, where

$$(1-B)^{-d_i} = \sum_{k=0}^{\infty}\frac{\Gamma(d_i + k)}{\Gamma(d_i)\Gamma(k+1)}B^k, \quad i = 1, 2, \ldots, N, \tag{2}$$

with $d_i \in \mathbb{R}$, B is the backward shift operator and $\Gamma(\cdot)$ represents the Gamma function. The N-dimensional vectors $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, ..., \varepsilon_{N,t}]'; t = 1, 2, \ldots, n$ are weakly stationary processes, such that $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbb{E}[\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}] = 0$ and

$$\mathbb{E}\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_{t+s}|\mathcal{F}_{t-1}\right] = \begin{cases}\boldsymbol{\Sigma}_\varepsilon, & \text{for } s = 0;\\ 0, & \text{otherwise.}\end{cases}$$

$\mathcal{F}_{t-1}$ represents the past information available at time t.

A special class of the STARFIMA model defined in Eq. 1 is the space-time autoregressive moving averaged (STARMA), obtained when $\boldsymbol{d} = \boldsymbol{0}$. It was proposed by Cliff and Ord (1975) and broadly studied by Pfeifer and Deutsch (1980a,b,c), Stoffer (1986), Antunes and Subba Rao (2006) , Monroy et al. (2015) among others.

The representation given in Eq. 1 is akin to that used in multivariate ARFIMA models, also known as VARFIMA. Both models consider the intrinsic relationships between the processes under study and have $N \times N$ coefficient matrices. The fundamental difference between them is the fact that in STARFIMA models, the spatial dependencies are imposed a priori by the model builder using a weighting matrix. Therefore, the coefficient matrices are simpler since they are products of scalars and known weighting matrices. However, as pointed out by Antunes and Subba Rao (2006), the parameters of the STARFIMA model cannot be obtained from the parameters of a VARFIMA model. Therefore, the STARFIMA is not a special case of the multivariate ARFIMA models except for the particular case when both models have the same orders.

2.1. The spatial weighting matrix

The definition of the weighting matrix $\boldsymbol{W}_l$ is non-trivial and can be rather arbitrary. There are several suggestions to define the weights of $\boldsymbol{W}_l$, all of them depend on the regularity of the grid. For example, when the grid is regularly spaced, uniform weights can be used, i.e., $w_{ij}^{(k)} = \frac{1}{n_{ki}}$ if the sites i and j are $k-$th order neighbors, and zero otherwise. The value n_{ki} represents the number of k-th order neighbors at the $i-$th site (Besag, 1974).

One widely used approach is based on the definition of the weights as the inverse of the Euclidean distances between sites (Cliff and Ord, 1981). It is specially useful when the sampled sites are not on a regular grid. In this case, defining weighting matrices of higher spatial order is not an easy task. As pointed out by Gao and Subba Rao (2011), to avoid these difficulties, all sites may be considered as the first order neighbors of each other site. That is, it can be assumed the spatial orders $\lambda_k = 1$ and $m_k = 1$ for all k. In such a case, there are only two weighting matrices: $\boldsymbol{W}_0 = \mathbb{I}_N$ and $\boldsymbol{W}_1 = \boldsymbol{W}$. Other ways to define the weighting matrix can be found in Bennet (1979), Anselin and Smirnov (1996) , Garrido (2000) among others.

Since in most of practical applications the sites are not scattered on a regular grid, here we consider the weighting matrices based on irregularly spaced sites, i.e. we only consider weighting matrices up to first order neighborhood. In this case, the STARFIMA process in Eq. 1 is simplified to the STARFIMA($p_1; \boldsymbol{d}; q_1$) given by

$$\boldsymbol{\Phi}_{p,1}(B)\boldsymbol{Z}_t = \boldsymbol{\Theta}_{q,1}(B)[\mathcal{D}(B)]^{-1}\boldsymbol{\varepsilon}_t, \quad t = 1, 2, \ldots, n, \tag{3}$$

where $\boldsymbol{\Phi}_{p,1}(z) = \mathbb{I}_N - \sum_{k=1}^{p}(\phi_{k0}\mathbb{I}_N + \phi_{k1}\boldsymbol{W})z^k$, $z \in \mathbb{C}$ and $\boldsymbol{\Theta}_{q,1}(z) = \mathbb{I}_N - \sum_{k=1}^{q}(\theta_{k0}\mathbb{I}_N + \theta_{k1}\boldsymbol{W})z^k$, $z \in \mathbb{C}$.

2.2. Properties of the STARFIMA($p_1; \boldsymbol{d}; q_1$) process

The values of ϕ_{k0}, ϕ_{k1} and $\boldsymbol{W}$ must keep stationary and causal conditions in order to assure the existence of a unique solution of the difference equations representing the process. The space-time ARFIMA process is said to be *causal* (or stable) if there is an equivalent infinite moving average representation. Additionally, the process is *invertible* if it can be expressed as an infinite order autoregressive process. The conditions for stationarity and invertibility of the STARFIMA($p_1; \boldsymbol{d}; q_1$) process are given by the Theorem 1. The Theorem 2 defines the functions for analyzing the space-time dependence structure of the process in time and frequency domains, respectively. The proofs of the theorems are given in the Appendix.

Theorem 1. *Let $\boldsymbol{Z}_t$ the STARFIMA process defined in Eq. 3 with $d_i \in (-1, 0.5)$, $i = 1, 2, \ldots, N$. Then,*

a. *if* $\det\left\{\mathbb{I}_N - \sum_{k=1}^{p}(\phi_{k0}\mathbb{I}_N + \phi_{k1}\boldsymbol{W})z^k\right\} \neq 0$, *for* $|z| \leq 1$ *with* $z \in \mathbb{C}$, *there is a unique stationary condition solution of 3 given by*

$$\boldsymbol{Z}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j} \tag{4}$$

where $\boldsymbol{\Psi}(z) = (\boldsymbol{\Phi}_{p,1}(z)\mathcal{D}(z))^{-1}\boldsymbol{\Theta}_{q,1}(z)$.

b. *if* $\det\left\{\mathbb{I}_N - \sum_{k=1}^{p}(\phi_{k0}\mathbb{I}_N + \phi_{k1}\boldsymbol{W})z^k\right\} \neq 0$, *for* $|z| \leq 1$ *with* $z \in \mathbb{C}$, *the process is said to be causal.*

c. *if* $\det\left\{\mathbb{I}_N - \sum_{k=1}^{q}(\theta_{k0}\mathbb{I}_N + \theta_{k1}\boldsymbol{W})z^k\right\} \neq 0$, *for* $|z| \leq 1$ *with* $z \in \mathbb{C}$, *the process is said to be invertible.*

Theorem 2. *Let $\boldsymbol{Z}_t$ a casual and invertible STARFIMA process with representation in Eq. 3 and $d_i \in (-1, 0.5)$, $i = 1, 2, \ldots, N$.*

a. *The space-time covariance function is*

$$\gamma_{lk}(s) = tr\left[\frac{\boldsymbol{W}_k'\boldsymbol{W}_l\boldsymbol{\Gamma}(s)}{N}\right], \quad k,l = 0,1, \tag{5}$$

where $tr[A]$ *is the trace of the square matrix* A*. The function* $\gamma_{lk}(s)$ *represents the covariance between the* l *and* k *order neighbors at the time lag* s *and the* $\boldsymbol{\Gamma}(s)$ *matrix is such that*

$$\boldsymbol{\Gamma}(s) \sim diag\{s^{d_1-0.5}, s^{d_2-0.5}, \ldots, s^{d_N-0.5}\}\boldsymbol{A}\ diag\{s^{d_1-0.5}, s^{d_2-0.5}, \ldots, s^{d_N-0.5}\}, \quad s \to \infty,$$

where the (i,j)*th element of the* $N \times N$ *matrix* $\boldsymbol{A}$ *is*

$$\frac{\Gamma(1-d_i-d_j)}{\Gamma(d_j)\Gamma(1-d_j)}\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\pi}_j,$$

$\Gamma(\cdot)$ *the gamma function,* $\boldsymbol{\pi}_j$ *the* j*th row of* $\boldsymbol{\Phi}_{p,1}(1)^{-1}\boldsymbol{\Theta}_{q,1}(1)$ *and the symbol "*$\sim$*" means that the ratio of left- and right-hand sides tends to* 1*.*

b. *The spectral matrix density function* $\boldsymbol{f}(\omega)$ *at* ω *frequency, is given by*

$$\boldsymbol{f}(\omega) = \mathcal{D}\left(e^{i\omega}\right)^{-1}\boldsymbol{f}_{ST}(\omega)\left[\mathcal{D}\left(e^{i\omega}\right)^{-1}\right]^*, \tag{6}$$

with $\boldsymbol{f}_{ST}(\omega) = \frac{1}{2\pi}\boldsymbol{\Phi}_{p,1}\left(e^{i\omega}\right)^{-1}\boldsymbol{\Sigma}_\varepsilon\left[\boldsymbol{\Phi}_{p,1}\left(e^{i\omega}\right)^{-1}\right]^*$ *and* $\boldsymbol{M}^*$ *represents the conjugate transpose of the complex matrix* $\boldsymbol{M}$*. The matrix function* $\boldsymbol{f}_{ST}(\omega)$ *represents the space-time spectral density of the process.*

It can be seen that the dependence structure of the process is influenced by the memory parameter. Furthermore, as $s \to \infty$, the autocovariances die out as a hyperbolic rate.

Note that, as $\omega \to 0^+$ we have

$$f_{ST}(\omega) = \frac{1}{2\pi}\boldsymbol{\Phi}_{p,1}(1)^{-1}\Sigma\left[\boldsymbol{\Phi}_{p,1}(1)^{-1}\right]^* \sim G, \tag{7}$$

where G is a symmetric and positive definite matrix. Hence, the espectral density defined in Eq. 6 is such that $f(\omega) \sim \Lambda(\omega;d)G\Lambda^*(\omega;d)$ where $\Lambda(\omega;d) = \mathcal{D}\left(1-\omega e^{i\frac{(\omega-\pi)}{2}}\right)^{-1}$ and the symbol "$\sim$" means that the ratio of left- and right-hand sides tends to 1. In this case, to estimate the vector of parameters $\boldsymbol{d} = (d_1, d_2, \ldots, d_N)'$ we may apply the existing results for vector ARFIMA models.

2.3. Parameter estimation

The procedure of parameter estimation is carried out in two steps. In the first step, we consider the semiparametric estimation of the vector $\boldsymbol{d} = (d_1, d_2, \ldots, d_N)'$ in a neighborhood of the origin, based on the local Whittle estimator suggested by Kunsch (1987) and widely studied in a series of papers by Robinson(1995a; 1995b; 2008). Having estimated the memory parameters, the data must be filtered in order to obtain the vector that will be analyzed.

In the second step, we estimate the vector of parameters of the STARMA model for the filtered series from the first step by using the methodology developed by Pfeifer and Deutsch (1980*a*).

2.3.1. Memory estimates

Let $\boldsymbol{I}(\omega_j)$ be the periodogram matrix function of $\boldsymbol{Z}_t$ evaluated at Fourier frequencies $\omega_j = \frac{2\pi j}{n}$ and given by

$$\boldsymbol{I}(\omega_j) = \frac{1}{2\pi n} \left(\sum_{t=1}^{n} \boldsymbol{Z}_t e^{it\omega_j} \right) \left(\sum_{t=1}^{n} \boldsymbol{Z}_t e^{it\omega_j} \right)^{*}, \tag{8}$$

where $j = 1, 2, \ldots, \lfloor \frac{n}{2} \rfloor$ and $\lfloor \cdot \rfloor$ denotes the integer part. The periodogram function is an estimator of the espectral density function of the process $\boldsymbol{Z}_t$ and it can be rapidly computed by fast Fourier transform, even when n is quite large.

The local approximation of the Gaussian log-likelihood function at the origin is given by

$$\mathcal{Q}(G, \boldsymbol{d}) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log\det\left\{ \Lambda(\omega_j; \boldsymbol{d}) G \Lambda^*(\omega_j; \boldsymbol{d}) \right\} + \mathrm{tr}\left[\Lambda(\omega_j; \boldsymbol{d}) G \Lambda^*(\omega_j; \boldsymbol{d}) \boldsymbol{I}(\omega_j)^{-1} \right] \right\},$$

where $\boldsymbol{I}(\omega_j)$ is defined in Eq. 8, $m \in [1, n/2]$ is a bandwidth number which satisfies at least $\frac{1}{m} + \frac{m}{n} \to 0$ as $n \to \infty$ (e.g., $m = o(n)$ and tends to infinity as $n \to \infty$, but at a slower rate than n) and "tr" denotes the trace of a matrix, G and Λ are defined in 7. The local Whittle estimator of the vector $\boldsymbol{d}$ is defined as

$$\widehat{\boldsymbol{d}} = \arg\min_{\boldsymbol{d}} \mathcal{R}(\boldsymbol{d}), \tag{9}$$

where $\mathcal{R}(\boldsymbol{d}) = \mathcal{Q}(\widehat{G}; \boldsymbol{d}) = \log\det \widehat{G}(\boldsymbol{d}) - \frac{2}{m} \sum_{i=1}^{N} \sum_{j=1}^{m} d_i \log \omega_j$, and

$$\widehat{G} = \frac{1}{m} \sum_{j=1}^{m} \mathbf{Re}\left\{ \Lambda(\omega_j; \boldsymbol{d})^{-1} \boldsymbol{I}(\omega_j) \Lambda^*(\omega_j; \boldsymbol{d})^{-1} \right\}$$

and **Re** denotes the real part of a complex number.

Lobato (1999) derived the semi-parametric two-step estimator in a multivariate long memory model, by extending the work by Robinson (1995a) on the univariate local Whittle (LW) estimator, initially proposed by Kunsch (1987). Shimotsu (2007) shows that the estimator of Lobato (1999) is consistent since the spectral density representation is more precise, and the limiting distribution is more evolved. Therefore, it follows that the estimator of Shimotsu (2007) has a smaller limiting distribution than the two-step estimator of Lobato (1999). Under some regularity conditions, Shimotsu (2007) established the asymptotic normality of the Gaussian semi-parametric estimator of multivariate stationary fractionally integrated processes in Eq. 9, i.e.,

$$m^{1/2}(\widehat{\boldsymbol{d}} - \boldsymbol{d}_0) \xrightarrow{\mathcal{D}} N(0, \Omega^{-1}), \;\; \Omega = 2\left[G^0 \odot (G^0)^{-1} + \mathbb{I}_N + \frac{\pi^2}{4}(G^0 \odot (G^0)^{-1} - \mathbb{I}_N)\right],$$

$\widehat{G}(\widehat{\boldsymbol{d}}) \xrightarrow{p} G^0$, where $\odot$ denotes the Hadamard product and the true parameter values are denoted by $\boldsymbol{d}_0$ and G^0. Nielsen (2011) extend the results, presented by Shimotsu (2007), to cover non-stationary values of $\boldsymbol{d}$ by using the notion of the extended discrete Fourier transform. The author established the central limit theorem under the same argument as in the stationary case $|d_i| < \frac{1}{2}$, $i = 1, \ldots, N$, derived by Robinson (1995a), for the univariate case, and Shimotsu (2007), for the multivariate case, for $d_i \in \left(-\frac{1}{2}, \infty\right)$, $i = 1, \ldots, N$.

3. Empirical Results

We conducted a simulation study aiming to explore the behavior of the proposed estimation methodology for different values of the parameters and weighting matrices.

We assume a STARFIMA$(1_1, \boldsymbol{d}, 0)$ process with four variables. The considered weighting matrix is based on the real data matrix obtained for the monitoring stations analyzed in Section 4. It is given by:

$$\boldsymbol{W}^{(1)} = \begin{pmatrix} 0.00 & 0.40 & 0.25 & 0.35 \\ 0.40 & 0.00 & 0.30 & 0.30 \\ 0.30 & 0.55 & 0.00 & 0.15 \\ 0.08 & 0.20 & 0.78 & 0.00 \end{pmatrix}.$$

The data were generated assuming combinations of the parameters $\phi_{10} = 0.1, 0.12$; $\phi_{11} = 0.1, 0.51$ and $\boldsymbol{d} = \{(0,0,0,0), (0.0, 0.1, 0.1, 0.2), (0.1, 0.1, 0.3, 0.3), (0.45, 0.45, 0.45, 0.45)\}$, in order to reflect different assumptions about them. These values of the parameters jointly with the specifications of the matrix $\boldsymbol{W}$ are such that the causality condition is satisfied. The combinations $(\phi_{01}, \phi_{11}) = (0.1, 0.1)$, $(0.12, 0.1)$ lead to the maximal absolute eigenvalue of the matrix $(\phi_{10}\mathbb{I}_N + \phi_{11}\boldsymbol{W})$[1] equal to 0.58, whilst the combinations $(\phi_{01}, \phi_{11}) = (0.1, 0.51)$, $(0.12, 0.51)$, lead to the maximal absolute eigenvalue 0.99. Sample sizes were set to $n = \{300, 1000\}$ and bandwidth $m = \lfloor n^{\alpha} \rfloor$, were $\alpha \in \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The mean and MSE were computed using 1000 replications. Due to space issues, we present the results for $m = n^{0.5}$ since this value is widely used in the literature of long-memory models (see, for example, Geweke and Porter-Hudak (1983),Reisen (1994) among others) . The remaining results are available upon request.

Here we concentrate on the performance of the memory parameter estimates, since the behavior of the parameter estimates from the second step of the estimation procedure are highly influenced by the estimates of $\boldsymbol{d}$. Studies on the performance of the parameter estimates for the STARMA processes (second step) have been conducted by Subba Rao and Antunes (2003), Giacomini and Granger (2004) and Borovkova et al. (2008) among others.

Table 1 shows the estimates of the memory parameter when there is no long-range dependence ($\boldsymbol{d} = \mathbf{0}$), i.e., the classic STARMA case. It can be observed that the estimates are close to the real value when the maximal eigenvalues of the matrix $(\phi_{01} + \phi_{11}\boldsymbol{W})$ are within the unit circle, even for the smaller sample size. Nevertheless, when the eigenvalues are close to 1, the bias increases significantly for small sample sizes. In this case, even a small raise of the ϕ_{01} parameter causes an increase of the bias. The MSE stays stable for all combinations of the parameters.

When there is long-range dependence and the processes are stationary (Tables 2 and 3), the simulation results show that, as n increases, the bias of the $\boldsymbol{d}$ estimates tends to decrease. For those models which the maximal eigenvalues are close to 1, the bias is large even for larger sample sizes. As in the case of the STARMA process, a small increase of the ϕ_{10} parameter leads to a significant increasing of the bias at a slower rate if the sample size is greater. The MSE remains stable for all the cases.

Table 4 displays the performance of the estimates when the memory parameter is close to the non stationary region. In this case, the bias is significantly large even for the larger sample sizes. The performance of the estimates get poorer when the maximal eigenvalues are close to 1.

---

[1] This condition is analogous to the causality condition in Theorem 1.

Table 1 Memory parameter values and estimates for the STARMA$(1_1, 0)$ process ($\boldsymbol{d} = \mathbf{0}$).

n	300				1000			
ϕ_{01}	0.10		0.12		0.10		0.12	
ϕ_{11}	0.10	0.51	0.10	0.51	0.10	0.51	0.10	0.51
Mean	0.0309	0.1133	0.0314	0.1267	-0.0162	0.0115	-0.0161	0.0161
MSE	0.0394	0.0484	0.0393	0.0495	0.0245	0.0208	0.0245	0.0208
Mean	-0.0084	0.1176	-0.0075	0.1310	-0.0176	0.0188	-0.0172	0.0245
MSE	0.0342	0.0325	0.0341	0.0321	0.0234	0.0179	0.0233	0.0174
Mean	0.0250	0.1220	0.0257	0.1347	0.0028	0.0363	0.0031	0.0431
MSE	0.0251	0.0314	0.0251	0.0323	0.0196	0.0179	0.0196	0.0172
Mean	-0.0134	0.0928	-0.0128	0.1046	0.0029	0.0387	0.0034	0.0427
MSE	0.0407	0.0577	0.0408	0.0585	0.0197	0.0202	0.0195	0.0207

4. Application: daily average PM_{10} in GVR

In this section, we apply the developed methodology to daily average PM_{10} concentrations (μg/m^3). We compare the fitting and forecasting ability of the proposed STARFIMA model with the performance of the STARMA model with no consideration about the memory properties of the PM_{10} time series.

The raw series consists of observations from June 15, 2008 to December 31, 2009, obtained from six monitoring stations of the Automatic Air Quality Monitoring Network (AAQMN) in the Greater Vitória Region, Brazil. Thus, we have $N = 6$ sites and $n = 560$ observations in time. Figures 1 and 2 show the locations of the sites and the time series obtained from each one of them, respectively.

We estimated the missing values using the Gibbs sampling for multiple imputations of the incomplete multivariate data suggested by Aerts et al. (2002). The first 546 observations were used for modeling purposes and the last 14, corresponding to the last two weeks of the full period, were used for forecasting purposes.

Since the region has a small number of stations distributed irregularly over a relatively small area, we consider the weighting matrix $\boldsymbol{W}$ as suggested by Gao and Subba Rao (2011). Then we

Table 2 Memory parameter values and estimates for the STARFIMA$(1_1, \boldsymbol{d}, 0)$ process

$\boldsymbol{d}$	n	300				1000			
	ϕ_{01}	0.10		0.12		0.10		0.12	
	ϕ_{11}	0.10	0.51	0.10	0.51	0.10	0.51	0.10	0.51
0.0	**Mean**	0.0295	0.1174	0.0300	0.1327	-0.0169	0.0120	-0.0168	0.0183
	MSE	0.0396	0.0406	0.0395	0.0383	0.0248	0.0186	0.0248	0.0181
0.1	**Mean**	0.0950	0.1986	0.0959	0.2277	0.0837	0.1246	0.0842	0.1297
	MSE	0.0337	0.0269	0.0336	0.0315	0.0228	0.0182	0.0227	0.0179
0.1	**Mean**	0.1266	0.1981	0.1274	0.2237	0.1028	0.1446	0.1031	0.1495
	MSE	0.0242	0.0377	0.0242	0.0380	0.0191	0.0162	0.0191	0.0163
0.2	**Mean**	0.1851	0.2880	0.1855	0.3026	0.2045	0.2477	0.2048	0.2516
	MSE	0.0420	0.0392	0.0423	0.0430	0.0175	0.0172	0.0174	0.0169

obtain

$$\boldsymbol{W} = \begin{bmatrix} 0.0000 & 0.4879 & 0.2292 & 0.1066 & 0.0872 & 0.0891 \\ 0.3887 & 0.0000 & 0.3355 & 0.1076 & 0.0818 & 0.0864 \\ 0.2031 & 0.3732 & 0.0000 & 0.1762 & 0.1183 & 0.1292 \\ 0.0850 & 0.1077 & 0.1586 & 0.0000 & 0.2212 & 0.4275 \\ 0.0989 & 0.1164 & 0.1513 & 0.3145 & 0.0000 & 0.3189 \\ 0.0768 & 0.0934 & 0.1256 & 0.4618 & 0.2424 & 0.0000 \end{bmatrix}.$$

The analysis of the periodograms of the series from each station (Figure 3) reveals that there are some significant periods at each site. Following Antunes and Subba Rao (2006), we subtracted the cyclical component in each time series individually. Denoting by $\boldsymbol{Y}_t$ the original time series, the transformed series can be written as $\boldsymbol{Z}_t = \boldsymbol{Y}_t - \boldsymbol{X}_t$, where $\boldsymbol{X}_t = [X_{1,t}, \ldots, X_{6,t}]'$ is a periodic function that can be represented as harmonic series, that is

$$X_{i,t} = \sum_{k=1}^{s} \left[\xi_{i,k} \cos\left(\frac{2\pi k t}{p_k}\right) + \xi^{\dagger}_{i,k} \sin\left(\frac{2\pi k t}{p_k}\right) \right], \; i \qquad = 1, \ldots, 6, \; t = 1, \ldots, n$$

where $\xi_{i,k}$ and $\xi^{\dagger}_{i,k}$ are unknown parameters which have to be estimated by least squares and p_k represents the periods of the time series.

Once the transformed series $Z(t)$ were obtained, we proceed to differentiate them by using the approach presented in Section 2.3.1. These filtered series are the time series to be used for

Table 3 Memory parameter values and estimates for the STARFIMA($1_1, \boldsymbol{d}, 0$) process

$\boldsymbol{d}$	n	300				1000			
	ϕ_{01}	0.10		0.12		0.10		0.12	
	ϕ_{11}	0.10	0.51	0.10	0.51	0.10	0.51	0.10	0.51
0.1	**Mean**	0.1348	0.2249	0.1355	0.2383	0.0859	0.1307	0.0860	0.1376
	MSE	0.0377	0.0429	0.0376	0.0428	0.0238	0.0215	0.0238	0.0208
0.1	**Mean**	0.0976	0.2146	0.0985	0.2295	0.0869	0.1299	0.0874	0.1344
	MSE	0.0341	0.0347	0.0341	0.0336	0.0226	0.0196	0.0225	0.0197
0.3	**Mean**	0.3280	0.4048	0.3285	0.4169	0.3046	0.3514	0.3050	0.3561
	MSE	0.0242	0.0363	0.0244	0.0374	0.0183	0.0159	0.0183	0.0159
0.3	**Mean**	0.2895	0.4064	0.2901	0.4187	0.3048	0.3415	0.3050	0.3467
	MSE	0.0445	0.0468	0.0445	0.0474	0.0176	0.0184	0.0176	0.0183

modeling. The estimates of the memory parameters were obtained using different bandwidth values $m = \lfloor n^\alpha \rfloor, \alpha \in \{0.4, 0.5, 0.6\}$. The estimates showed to be stable across the bandwidth values, inspired on the results showed by the simulation procedures, we decided to chose the estimates for $\alpha = 0.5$. Here we only present the results for this bandwidth, however the results for the other m values are available upon request. Thus, the estimates are $\widehat{\boldsymbol{d}} = (0.47, 0.40, 0.31, 0.38, 0.35, 0.49)$. From the estimates, it can be observed that the series in all the monitoring stations have long memory behavior and are stationary.

The temporal order is chosen by analyzing the space-time autocorrelation (STACF) and partial autocorrelation (STPACF) functions (Figures 4a and 4b). The cutting-off in the STFAC and STPACF after the second time lag suggest that a suitable model is a STARFIMA with maximum order 2 for the AR and MA components. There are some significant partial correlations at the first spatial lag, which indicates that this spatial order in the autoregressive component should be included.

The model with the best performance for the filtered series is the STARFIMA($2_{10}, \widehat{\boldsymbol{d}}, 0$) with estimates of the parameters given by[2]: $\phi_{10} = 0.1060$ (0.01978), $\phi_{20} = 0.1101$ (0.02697) and $\phi_{11} =$

[2]The standard deviations are shown in parentheses.

Table 4 Memory parameter values and estimates for the STARFIMA($1_1, \boldsymbol{d}, 0$) process

$\boldsymbol{d}$	n	300				1000			
	ϕ_{01}	0.10		0.12		0.10		0.12	
	ϕ_{11}	0.10	0.51	0.10	0.51	0.10	0.51	0.10	0.51
0.45	**Mean**	0.4895	0.5893	0.4903	0.6021	0.4550	0.4749	0.4551	0.4809
	MSE	0.0421	0.0437	0.0420	0.0460	0.0246	0.0212	0.0246	0.0211
0.45	**Mean**	0.4697	0.5764	0.4706	0.5898	0.4463	0.4744	0.4464	0.4785
	MSE	0.0309	0.0382	0.0308	0.0376	0.0233	0.0165	0.0233	0.0162
0.45	**Mean**	0.4849	0.5535	0.4855	0.5681	0.4568	0.4959	0.4570	0.4999
	MSE	0.0227	0.0412	0.0228	0.0400	0.0185	0.0164	0.0186	0.0163
0.45	**Mean**	0.4515	0.5476	0.4524	0.5585	0.4635	0.4945	0.4638	0.4992
	MSE	0.0527	0.0510	0.0524	0.0526	0.0170	0.0184	0.0171	0.0185

-0.0980 (0.01981). The STACF of the residuals, displayed in Figure 5, shows very small autocorrelation values, suggesting that the assumption of uncorrelated errors is satisfied by the fitted model.

According to the model, the influence of the PM_{10} over the region is around 1-2 days. The concentrations of the pollutant are highly influenced by the concentrations observed in the site and its neighbors the day before ($\phi_{10} = 0.1060$ and $\phi_{11} = -0.0980$).

STARMA Modeling

Considering the STARMA modeling methodology, the model with the best performance is the STARMA($2_{10}, 0$) with estimated parameters $\phi_{10} = -0.3372$ (0.0198), $\phi_{20} = -0.1029$ (0.0269) and $\phi_{11} = -0.0987$ (0.0198). The STACF of the residuals (not shown here, but available upon request) indicate that the model is adequate for the data.

Performance comparison

Figure 6 displays the predicted values of the observed time series by using the two fitted models. Figure 6b shows the superior in-sample performance of the STARFIMA model. It can be considered as a more suitable method for estimating missing data than the STARMA model (Figure 6a) because it can predict the larger values with more accuracy.

Figure 1 Map of the studied AAQMN monitoring stations in the Greater Vitória Region.

Regarding to the forecasting ability, we obtained one-step-ahead forecasts for a 14-days period using the Minimum Mean Square Error (MMSE) criterion. Figure 7 displays the forecasts and their 95% prediction intervals. The forecasts obtained using the STARMA model follow well the behavior of the time series (Figure 7a), nevertheless, the model cannot capture the variability with good reliability. In this sense, the results showed in Figure 7b show that the performance of the STARFIMA model is superior for all the sites.

Aiming to quantify the forecasting ability for each monitoring station, we calculated the root mean squared error (RMSE) for both models. As observed in Table 5, taking into account the memory characteristics in the model led to an improvement of the accuracy of, at least, 38%. For example, the RMSE of Vila Velha Centro obtained using the STARMA model is 1.39 times the RMSE obtained using the STARFIMA methodology. Similarly, the RMSE for Enseada do Suá station using the STARMA model is 1.78 times the RMSE obtained with the STARFIMA model, which means an approximately 78% improving of the forecasting performance.

5. Final Remarks

This study presents the space-time ARFIMA model as a suitable alternative for modeling air pollution data. The developed methodology is applied to daily average PM_{10} concentrations in order to describe the dynamics of the pollutant at the Greater Vitória Region, as well as to forecast

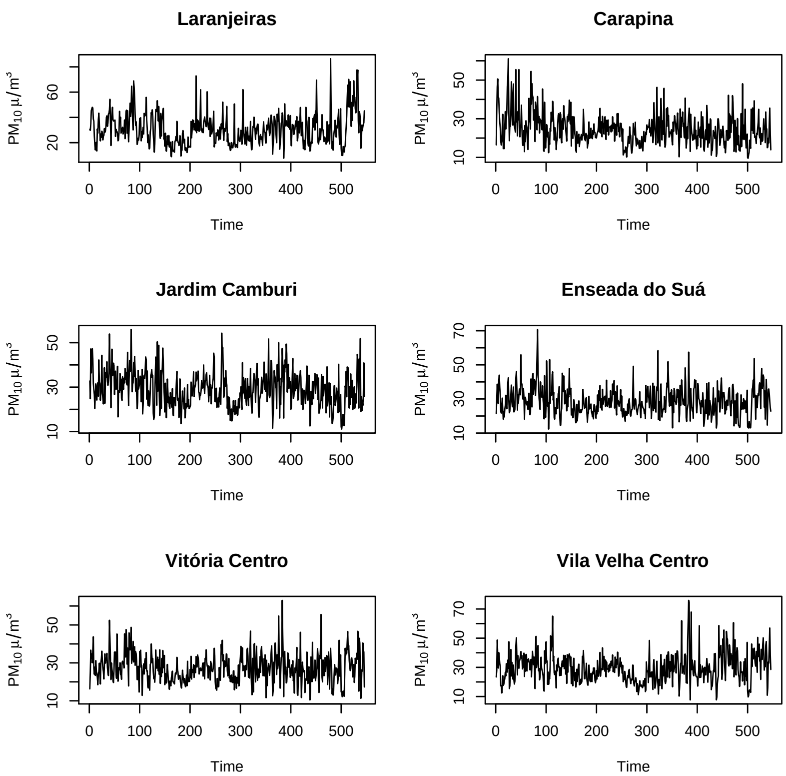


Figure 2 Time series obtained for each monitoring station.

future concentrations.

According to the fitted model, the persistence of the PM_{10} in the region is about two days and its concentration levels are highly influenced by the levels observed at the closest sites the day before. The residual analysis indicated a good fit for in-sample observations, so that it can be used for imputation of missing values. Regarding the out-of-sample performance, the model showed to be a very good tool for predicting future values.

Acknowledgements

This work was performed under the CAPES financial support and it is a part of the PhD thesis of the first author under supervision of the second author at the PPGEA–UFES-Brazil.
Professor T. Subba Rao (Adjunct Professor, CRRAO AIMSCS) thanks the University of Manchester, UK. He also wishes to thank the Department of Science and Technology, Government of India, for their financial support through their research grant to the institute No. SR/S4/516/07 which supported his visit to the Institute. Subbarao's s visit to the CRRAO AIMSCS, India is partially

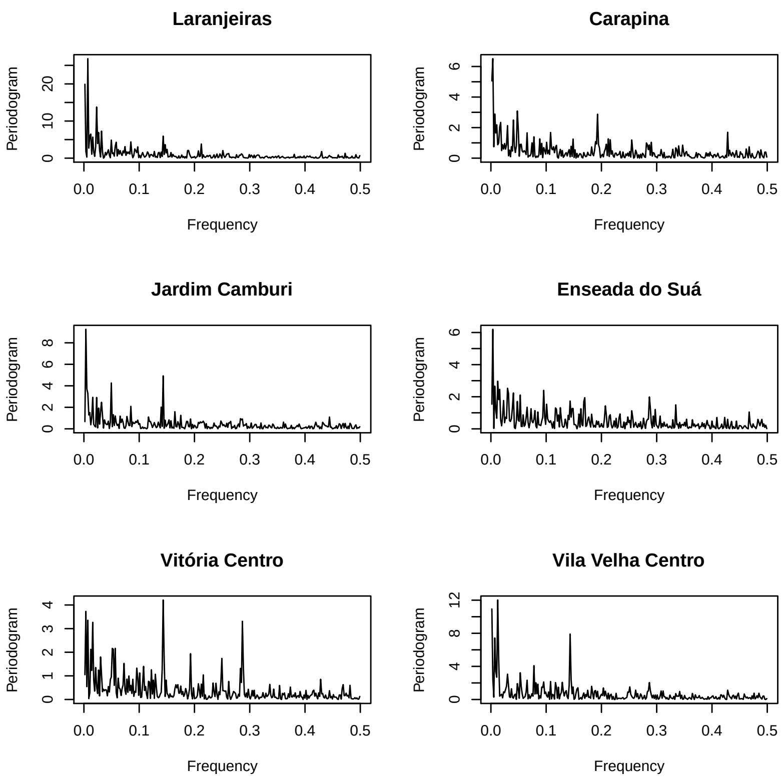


Figure 3 Periodograms for the time series at each monitoring station.

supported by the DST-CMS Gol Project No. SR/S4/MS:516/07 dated 21.04.2008

Prof. Valdério Reisen thanks FAPES and CNPq for the financial support.

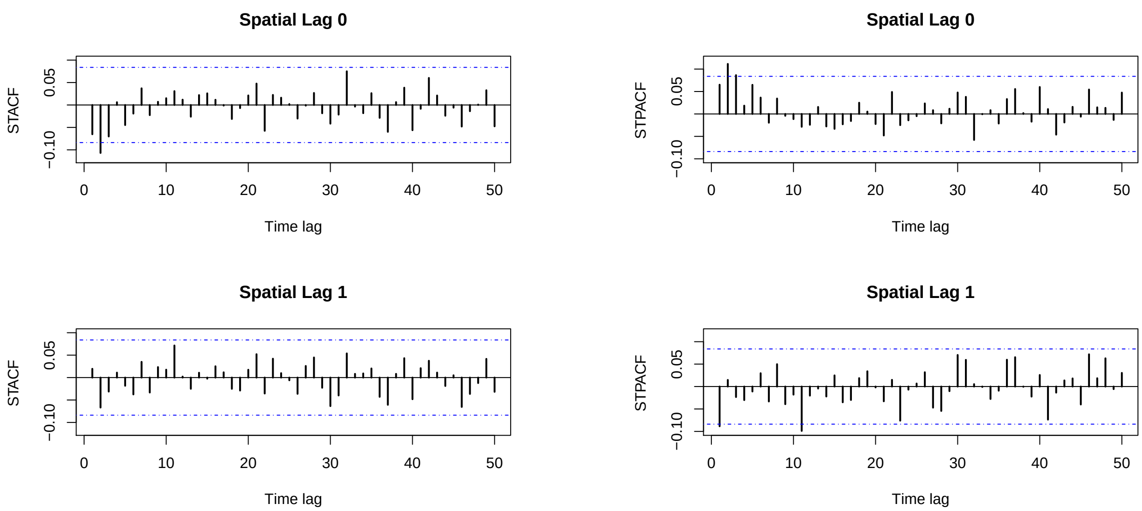
The authors would like to thank the Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA) of Espírito Santo State for providing the data.

Aerts, M., Claeskens, G., Hens, N. and Molenberghs, G. . (2002), 'Local multiple imputation', *Biometrika* **89**, 375–388.

Anselin, L. and Smirnov, O. (1996), 'Efficient algorithms for constructing proper higher order spatial lag operators', *Journal of Regional Science* **36**(1), 67–89.

Antunes, A. and Subba Rao, T. (2006), 'On hypotheses testing for the selection of spatio-temporal models', *Journal of Time Series Analysis* **27**(5), 767–791.

Bennet, R. (1979), *Spatial time series, analysis-forecasting-control*, Holden-Day, Inc. San Francisco, CA.



(a) STACF **(b)** STPACF

Figure 4 Space-time Autocorrelation (STACF) and Partial Autocorrelation (STPACF) Functions for the differenced PM_{10} daily average.

Besag, J. S. (1974), 'Spatial interaction and the statistical analysis of lattice system', *Journal of the Royal Statistical Society, Series B* **36**, 197–242.

Borovkova, S., Lopuhaa, H. and Ruchjana, B. (2008), 'Consistency and asymptotic normality of least squares estimators in generalized star models', *Statistica Neerlandica* pp. 1–27.

Brockwell, P. and Davis, R. (1991), *Time Series: Theory and Methods*, second edn, Springer.

Cliff, A. and Ord, J. (1975), 'Space-time modeling with an application to regional forecasting', *Transactions of the Institute of British Geographers* **64**, 119–128.

Cliff, A. and Ord, J. (1981), *Spatial Processes: Models and Applications*, London: Pion.

Crespo, J., Zorrilla, M., Bernardos, P. and Mora, E. (2007), 'A new image prediction model based on spatio-temporal techniques', *Visual Computer* **23**(6), 419–431.

De-Iaco, S., Myers, D. and Posa, D. (2003), 'The linear coregionalization model and the product-sum space-time variogram', *Mathematical Geology* **35**(1), 25–38.

Deutsch, S. and Ramos, J. (1986), 'Space-time modeling of vector hydrologic sequences', *Water Resources Bulletin* **22**, 967–980.

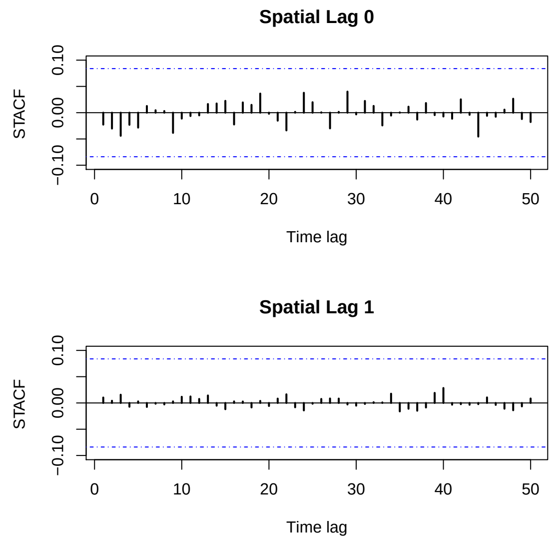


Figure 5 Space-time Autocorrelation Function (STACF) of the residuals from the fitted STARFIMA($2_{10}, \widehat{\boldsymbol{d}}, 0$) model.

Epperson, B. (1994), 'Spatial and space-time correlations in systems of subpopulations with stochastic migration', *Theoretical Population Biology* **46**, 106–197.

Epperson, B. (2000), 'Spatial and space-time correlations in ecological models', *Ecological Modelling* **132**, 63–76.

Gao, X. and Subba Rao, T. (2011), Regression models with starma errors: An application to the study of temperature variations in the antarctic peninsula, *in* M. T. Wells and A. SenGupta, eds, 'Advances in Directional and Linear Statistics', Physica-Verlag HD, pp. 27–50.

Garrido, R. (2000), 'Spatial interaction between the truck flows through the mexico-texas border', *Transportation Research Part A* **34**, 23–33.

Geweke, J. and Porter-Hudak, S. (1983), 'The estimation and application of long memory time series model', *Journal of Time Series Analysis* **4**, 221–238.

Giacinto, V. D. (2006), 'A generalized space-time ARMA model with an application to regional unemployment analysis in italy', *International Regional Science Review* **29**(2), 159–198.

Giacomini, R. and Granger, C. (2004), 'Aggregation of space-time processes', *Journal of Econometrics* **118**, 7–26.

Table 5 Model accuracy measures for both fitted models.

Station	**STARMA**$(2_{10}, 0)$	**STARFIMA**$(2_{10}, \widehat{\boldsymbol{d}}, 0)$
Laranjeiras	5.5767	3.2323
Carapina	2.6455	1.5250
Jardim Camburi	4.6144	2.9156
Enseada do Suá	5.9992	3.3684
Vitória Centro	4.8821	3.2148
Vila Velha Centro	3.3488	2.4147

Glasbey, C. and Allcroft, D. (2008), 'A spatiotemporal auto-regressive moving average model for solar radiation', *Journal of the Royal Statistical Society: Applied Statistics* **57**(3), 343–355.

Granger, C. and Joyeux, R. (1980), 'An introduction to long memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–30.

Hosking, J. (1981), 'Fractional differencing', *Biometrika* **68**, 165–167.

Huerta, G., Sansï¿$\frac{1}{2}$, B. and Stroud, J. (2004), 'A spatiotemporal model for Mexico City ozone levels', *Applied Statistics* **53**(2), 231–248.

Kamarianakis, Y. and Prastacos, P. (2005), 'Space-time modeling of traffic flow', *Computers and Geosciences* **31**, 119–133.

Kunsch, H. (1987), Statistical aspects of self-similar processes, *in* 'Proceedings of the First World Congress of the Bernoulli Society', Prokhorov, Yu., Sazanov, V.V., pp. 67–74.

LaValle, P., Lakhan, V. and Trenhaile, A. (2001), 'Space-time series modelling of beach and shoreline data', *Environmental Modelling & Software* **16**, 299–307.

Lobato, I. N. (1999), 'A semiparametric two-step estimator in a multivariate long-memory model', *Journal of Econometrics* **90**, 129–153.

Monroy, N. J., Reisen, V. A., Subba Rao, T. and Terdik, G. (2015), 'Modeling and forecasting of sulfur dioxide using space-time series models. a case study', *In press* (1), 01–10.

Nielsen, F. (2011), 'Local whittle estimation of multivariate fractionally integrated processes', *Journal of Time Series Analysis* **32**, 317–335.

Pace, R., Barry, R., Clapp, J. and Rodriguez, M. (1998), 'Spatiotemporal autoregressive models of neighborhood effects', *Journal of Real Estate Finance and Economics* **17**(1), 15–33.

Palma, W. (2007), *Long-memory time series: theory and methods*, Vol. 662, Wiley. com.

Pfeifer, P. and Deutsch, S. (1980*a*), 'A comparison of estimation procedures for the parameters of the star model', *Communications in Statistics, Serie B - Simmulation and Computation* **9**(3), 255–270.

Pfeifer, P. and Deutsch, S. (1980*b*), 'Identification and interpretation of first order space-time ARMA models', *Technometrics* **22**(3), 397–408.

Pfeifer, P. and Deutsch, S. (1980*c*), 'A three-stage iterative procedure for space-time modeling', *Technometrics* **22**(1), 35–47.

Reisen, V. (1994), 'Estimation of the fractional difference parameter in the arima (p,d,q) model using the smoothed periodogram', *Journal of Time Series Analysis* **15**, 335–350.

Reisen, V. A., Sarnaglia, A. J. Q., Reis Jr, N. C., Lévy-Leduc, C. and Santos, J. M. (2014), 'Modeling and forecasting daily average PM_{10} concentrations by a seasonal long-memory model with volatility', *Environmental Modelling & Software* **51**(0), 286–295.

Reynolds, K. and Madden, L. (1988), 'Analysis of epidemics using spatio-temporal autocorrelation', *Phytopathology* **78**(2), 240–246.

Reynolds, K., Madden, L. and Ellis, M. (1988), 'Spatio-temporal analysis of epidemic development of leather rot of strawberry', *Phytopathology* **78**(2), 246–252.

Robinson, P. (1995a), 'Gaussian semiparametric estimation of long range dependence', *The annals of statistics* **23**, 1630–1661.

Robinson, P. (1995b), 'Log-periodogram regression of time series with long range dependence', *The annals of statistics* **23**, 1048–1072.

Robinson, P. (2008), 'Multiple local whittle estimation in stationary systems', *The Annals of statistics* **36**, 2508–2530.

Rouhani, S., Ebrahimpour, M., Yaqub, I. and Gianella, E. (1992), 'Multivariate geostatistical trend detection and network evaluation of space-time acid deposition data – I. Methodology', *Atmospheric Environment. Part A. General Topics* **26**(14), 2603 – 2614.

Seber, G. A. (2008), *A matrix handbook for statisticians*, Vol. 15, Wiley. com.

Shimotsu, K. (2007), 'Gaussian semiparametric esimation of multivariate fractionally integrated processes', *Journal of Econometrics* **137**, 277–310.

Soni, P., Chan, Y., Preissl, H., Eswaran, H., Wilson, J., Murphy, P. and Lowery, C. (2004), 'Spatial-temporal analysis of non-stationary fMEG data', *Neurology and Clinical Neurophysiology* **100**, 1–6.

Stoffer, D. (1986), 'Estimation and identification of space-time ARMAX models in the presence of missing data', *Journal of the American Statistical Association* **81**(395), 762–772.

Subba Rao, T. and Antunes, A. (2003), Spatio-temporal modelling of temperature time series: a comparative study, *in* 'Time Series Analysis and Applications to Geophysical Systems', The IMA volumes in Mathematics and its Applications, pp. 123–150.

Terzi, S. (1995), 'Maximum likelihood estimation of a generalized star$(1, 1_p)$ model', *Statistical Methods and Applications* **4**(3), 377–393.

Yu, T.-Y. and Chang, I.-C. (2006), 'Spatiotemporal features of severe air pollution in Northern Taiwan', *Environmental science and pollution research international* **13**(4), 268–275.

Zeri, M., Oliveira-Júnior, J. and Lyra, G. (2011), 'Spatiotemporal analysis of particulate matter, sulfur dioxide and carbon monoxide concentrations over the city of Rio de Janeiro, Brazil', *Meteorology and Atmospheric Physics* **113**, 139–152.

Appendix

To prove Theorem 1, the following results are used.

Definition 3. (Definition 19.3 in Seber (2008)). Let $\{\boldsymbol{A}_n\}$ $(n = 1, 2, ...)$ be a sequence of $N \times N$ matrices and let $a^n_{i,j}$ denote the (i, j)th element of $\{\boldsymbol{A}_n\}$. The sequence $\{\boldsymbol{A}_k\}$ converges to $\boldsymbol{A} = (a_{i,j})$, that is $\lim_{n\to\infty} \boldsymbol{A}_n = \boldsymbol{A}$, if $\lim_{n\to\infty} a^n_{i,j} = a_{i,j}$, $\forall\ i, j$, when $n \to \infty$.

Lemma 1. *Let $\boldsymbol{A}_n$ $(n = 1, 2, ...)$ be a sequence of $N\times N$ matrices. Furthermore, let a_n be a sequence of positive numbers. Then, $\boldsymbol{A}_n = O(a_n)$ if and only if $a^n_{i,j} = O(a_n)$ where $a^n_{i,j}$ denotes the (i, j)th element of $\{\boldsymbol{A}_n\}$.*

Proof of Theorem 1. a. For a fixed location $i = 1, 2, \ldots, N$, let $Y_{i,t} = \sum_{j=0}^{\infty} \eta_j \varepsilon_{i,t-j}$ be a random variable at the site i where η_j are the coefficients of the (i, i)-th entry of the diagonal matrix $[\mathcal{D}(B)]^{-1}$, that is, η_j are the coefficients of $(1 - B)^{-d_i}$ $(\eta^i_j = O(j^{d_i-1}))$ and $\varepsilon_{i,t}$ is the white noise process of the N-dimensional vectors $\boldsymbol{\varepsilon}_t$ with $\mathbb{E}[\varepsilon_{i,t}] = 0$, $t = 1, ...T$ and $\mathbb{E}[\varepsilon^2_{i,t}] = \sigma^2_i$. For $d_i < 1/2$, $\forall i = 1, \ldots, N$, it follows that $\sum_{j=0}^{\infty} \eta^2_j < \infty$ and, therefore, $\sum_{j=0}^{T} \eta_j e^{\iota\omega j}$ converge to $(1 - e^{\iota\omega})^{-d_i}$ as $T \to \infty$ in the Hilbert space $\mathbf{L}^2(d\omega)$ and $d\omega$ denotes the Lebesgue measure. By Theorems 4.10.1 and 1.4 in Brockwell and Davis (1991) and Palma (2007), respectively, the process $Y_{i,t}$ is well-defined. Therefore, by Definition 3 and the above results for $Y_{i,t}$, $\boldsymbol{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{\eta}_j \boldsymbol{\varepsilon}_{t-j}$ where $\sum_{j=0}^{\infty} \|\boldsymbol{\eta}_j\|_2 < \infty$, that is, $\boldsymbol{Y}_t$ is also well-defined. $\|\boldsymbol{A}\|_2$ denotes the 2-norm for the matrix $\boldsymbol{A}$ such as $\|\boldsymbol{A}\|^2_2 = \mathrm{tr}\{\boldsymbol{A}'\boldsymbol{A}\}$.

Note that, by Lemma 1, $\boldsymbol{\eta}_j = O(j^{\max_{i=1,..,N}\{d_i\}})$, the condition

$$\det\left\{\mathbb{I}_N - \sum_{k=1}^{p} \left(\phi_{k0}\mathbb{I}_N + \phi_{k1}\boldsymbol{W}\right) z^k\right\} \neq 0$$

for $|z| \leq 1$ with $z \in \mathbb{C}$ implies that $\exists\ \xi > 0$ such that $\boldsymbol{\Phi}^{-1}_{p,1}(z)$ exits for $|z| < 1 + \xi$. Since each of the N^2 elements of $\boldsymbol{\Phi}^{-1}_{p,1}(z)$ is a rational function of z with no singularities in $|z| < 1 + \xi$, consequently $\boldsymbol{\Phi}^{-1}_{p,1}(z)$ can be written as the absolutely convergent Laurent series, that is, it has the power expansion

$$\boldsymbol{\Phi}^{-1}_{p,1}(z)\boldsymbol{\Theta}_{q,1}(z) = \sum_{j=0}^{\infty} A_j z^j = A(z) \text{ for } |z| < 1 + \xi. \tag{10}$$

Thus, by Theorem 1.5(a) in Palma (2007), the process

$$\boldsymbol{Z_t} = \boldsymbol{\Phi}^{-1}_{p,1}(B)\boldsymbol{\Theta}_{q,1}(B)\boldsymbol{Y_t} = \sum_{j=0}^{\infty} \boldsymbol{A}_j B^j. \tag{11}$$

is a stationary vector process. Consequently $\boldsymbol{A}_j(1+\xi) \to 0$ as $j \to \infty$, so there exists $K \in (0, \infty)$, independent of j, such that all components of $\boldsymbol{A}_j$ are bounded in absolute value by $K(1+\xi/2)^{-j}$, $j = 0, 1, \ldots$. This implies absolute summability of the components of the matrices $\boldsymbol{A}_j$. Moreover, by Theorem 1.5(b) in Palma (2007), the vector $\boldsymbol{Z_t}$ can be written as Equation (4), that is,

$$\boldsymbol{Z}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j} \tag{12}$$

where $\boldsymbol{\Psi}(B) = \boldsymbol{\Phi}_{p,1}^{-1}(B)\boldsymbol{\Theta}_{q,1}(B)\boldsymbol{\eta}(B)$.

Now, premultiplying Equation 12 by $\boldsymbol{\Phi}_{p,1}(B)$ and applying Theorem 1.5 in Palma (2007), then

$$\boldsymbol{\Phi}_{p,1}(B)\boldsymbol{Z}_t = \boldsymbol{\Theta}_{q,1}(B)\boldsymbol{\eta}(B)\boldsymbol{\varepsilon}_t, \tag{13}$$

which shows that $\boldsymbol{Z}_t$ is a stationary vector process that satisfies Equations 1 and 4.

b. The proof of the casual property follows the same lines of the univariate case as Theorem 3.4(b) given in Palma (2007).

c. The proof that $\boldsymbol{Z}_t$ is invertible can be obtained using similar arguments of the proof in (a), excepts that conditions are required on the convergence of $[\mathcal{D}(z)]\ \dfrac{\Phi_{p,1}(z)}{\boldsymbol{\Theta}_{q,1}(z)}$.

□

According to the space-time covariance function of the standard STARMA model introduced by Pfeifer and Deutsch (1980*c*), the space-time covariance function of the STARFIMA model is defined as follows.

Definition 4. Let $\boldsymbol{Z}_t$ be the STARFIMA process defined in Eq. 3 with $d_i \in (-1, 0.5)$, $i = 1, 2, \ldots, N$ and satisfying the conditions of Theorem 1. Assuming that $\mathbb{E}\{Z_{i,t}\} = 0$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, the space-time covariance function can be expressed as

$$\gamma_{lk}(s) = \mathbb{E}\left\{\frac{[\boldsymbol{W}_l\boldsymbol{Z}_t]'[\boldsymbol{W}_k\boldsymbol{Z}_{t+s}]}{N}\right\}, \quad k, l = 0, 1, \ldots \tag{14}$$

Lemma 2. *Let $\boldsymbol{Z}_t$ be the space-time process defined in Eq. 3 with $d_i \in (-1, 0.5)$, $i = 1, 2, \ldots, N$ and satisfying the conditions of Theorem 1, that is, $\boldsymbol{Z}_t$ is a stationary i-dimensional process that has an infinite-order moving average representation as:*

$$\boldsymbol{Z}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j},$$

where $\boldsymbol{\Psi}_j$ are the $N \times N$ matrix of coefficients of $\boldsymbol{\Psi}(z) = (\boldsymbol{\Phi}_{p,1}(z)\mathcal{D}(z))^{-1}\boldsymbol{\Theta}_{q,1}(z)$. Then, 1.

$$\mathbf{\Psi}_s \sim diag\{\Gamma(d)^{-1}s^{d-1}\}\mathbf{\Pi}, \qquad as\ s \to \infty,$$

where $\Gamma(\cdot)$ *is the Gamma function and* $\mathbf{\Pi}$ *is a nonsingular* $N \times N$ *matrix of constants that are independent of* s*. The notation* $diag\{s^{d-1}/\Gamma(d)\}$ *represents a diagonal matrix* $N \times N$ *with* $s^{d_1-1}/\Gamma(d_1), \ldots, s^{d_N-1}/\Gamma(d_N)$ *on the diagonal.*

2. *Given the approximation in item (1), then*

$$\sum_{k=0}^{s} \mathbf{\Psi}_k \sim diag\{\Gamma(d+1)^{-1}s^{d}\}\mathbf{\Pi}, \qquad as\ s \to \infty.$$

Proof of Lema 2. 1. To prove item 1, it is only necessary to find the order of the diagonal elements of $\mathcal{D}(z)$ since the $\mathbf{\Pi}(z) = (\mathbf{\Phi}_{p,1}(z))^{-1}\mathbf{\Theta}_{q,1}(z)$ is a nonsingular $N \times N$ matrix of constants that are independent of s. Following the same lines part (a) of the proof of Theorem 1, the N memory parameters $d_1, d_2, ..., d_N$, with values lie in (-0.5, 0.5), the (i,i)-th entry of the diagonal matrix $[\mathcal{D}(B)]^{-1}$ are the coefficients of $(1-B)^{-d_i}$ of order $O(s^{d_i-1})$, that is, for each $i = 1, ..., N$, the (i,i) coefficient is $s^{d_i-1}/\Gamma(d_i)$. In addition, the notation $\sim$ is defined as follows: given two sequences of matrices $\boldsymbol{A}_s$ and $\boldsymbol{B}_s$ of the same dimensions, $\boldsymbol{A}_s \sim \boldsymbol{B}_s$, as $s \to \infty$, if $\frac{a^s_{i,j}}{b^s_{i,j}} \to 1$, where $a^s_{i,j}$ and $b^s_{i,j}$ are the $(i,j)th$ elements of $\boldsymbol{A}_s$ and $\boldsymbol{B}_s$, respectively.

2. To prove item 2, one can use the fact that since the series $\sum_{k=0}^{s} k^{d_i-1}/\Gamma(d_i)$ is divergent as $s \to \infty$, then this series can be approximated by $\int \frac{x^{d_i-1}}{\Gamma(d_i)}dx = \frac{j^{d_i}}{\Gamma(d_i+1)}$. By the fact that $[\mathcal{D}(B)]^{-1}$ is a $N \times N$ diagonal matrix with the (i,i) coefficient equal to $s^{d_i-1}/\Gamma(d_i)$, then $\sum_{k=0}^{s} \mathbf{\Psi}_k \sim diag\{\Gamma(d+1)^{-1}s^{d}\}\mathbf{\Pi}$ as $s \to \infty$.

□

Lemma 3. *Given the assumptions of Theorem 2,*

$$diag\{s^{0.5-d}\}\left(\sum_{k=0}^{\infty}\mathbf{\Psi}_k\mathbf{\Sigma}_\varepsilon\mathbf{\Psi}'_{k+s}\right)diag\{s^{0.5-d}\}$$
$$= diag\{s^{0.5-d}\}\left\{\sum_{k=1}^{\infty} diag\{\Gamma(d)^{-1}k^{d-1}\}\mathbf{\Pi}\boldsymbol{\sigma}_\varepsilon\mathbf{\Pi}'diag\{\Gamma(d)^{-1}(k+s)^{d-1}\}\right\}diag\{s^{0.5-d}\}$$
$$+\, o(1), \ \ as\ s \to \infty$$

Proof of Theorem 2. a. By Definition 4,

$$
\begin{aligned}
\gamma_{lk}(s) &= \frac{1}{N}\mathbb{E}\left\{[\boldsymbol{W}_l\boldsymbol{Z}_t]^{'}[\boldsymbol{W}_k\boldsymbol{Z}_{t+s}]\right\} \quad k,l=0,1,\\
&= \frac{1}{N}\mathbb{E}\left\{\boldsymbol{Z}_t^{'}\boldsymbol{W}_l^{'}\boldsymbol{W}_k\boldsymbol{Z}_{t+s}\right\} = \frac{1}{N}\mathbb{E}\left\{\mathrm{tr}(\boldsymbol{Z}_t^{'}\boldsymbol{W}_l^{'}\boldsymbol{W}_k\boldsymbol{Z}_{t+s})\right\}\\
&= \frac{1}{N}\mathbb{E}\left\{\mathrm{tr}(\boldsymbol{W}_k^{'}\boldsymbol{W}_l\boldsymbol{Z}_t\boldsymbol{Z}_{t+s}^{'})\right\} = \frac{1}{N}\mathrm{tr}\left\{\mathbb{E}(\boldsymbol{W}_k^{'}\boldsymbol{W}_l\boldsymbol{Z}_t\boldsymbol{Z}_{t+s}^{'})\right\}\\
&= \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{W}_k^{'}\boldsymbol{W}_l\mathbb{E}(\boldsymbol{Z}_t\boldsymbol{Z}_{t+s}^{'})\right\} = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{W}_k^{'}\boldsymbol{W}_l\boldsymbol{\Gamma}(s)\right\}\\
&= tr\left[\frac{\boldsymbol{W}_k^{'}\boldsymbol{W}_l\boldsymbol{\Gamma}(s)}{N}\right]
\end{aligned}
$$

where $\boldsymbol{\Gamma}(s) = \mathbb{E}(\boldsymbol{Z}_t\boldsymbol{Z}_{t+s}^{'})$.

From Lemmas 2 and 3, we have for the covariances $\boldsymbol{\Gamma}(s) \equiv \mathrm{cov}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+s})$ that

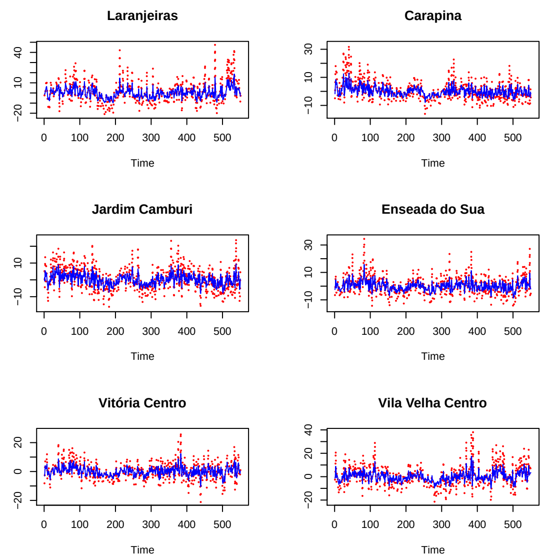
$$
\begin{aligned}
&diag\{s^{0.5-d}\}\mathrm{cov}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+s})diag\{s^{0.5-d}\}\\
&= diag\{s^{0.5-d}\}\mathrm{cov}\left(\sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\varepsilon}_{t-k}, \sum_{k=-s}^{\infty}\boldsymbol{\Psi}_{k+s}\boldsymbol{\varepsilon}_{t-k}\right)diag\{s^{0.5-d}\}\\
&= diag\{s^{0.5-d}\}\left(\sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Psi}_{k+s}^{'}\right)diag\{s^{0.5-d}\}.
\end{aligned}
$$

From Lemma 3 it follows

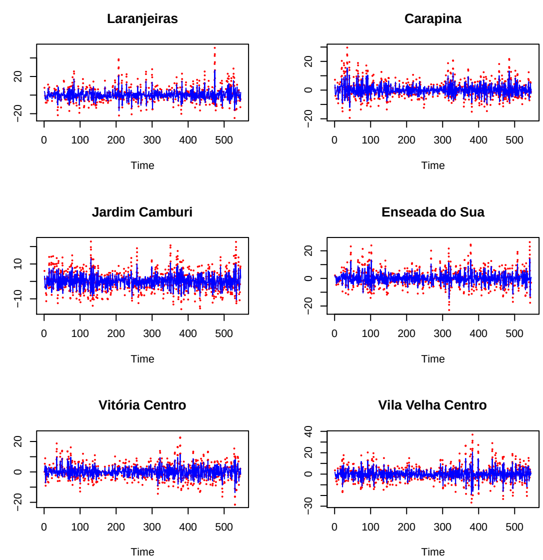
$$
\begin{aligned}
&diag\{s^{0.5-d}\}\mathrm{cov}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+s})diag\{s^{0.5-d}\}\\
&\to \int_0^{\infty} diag\{\Gamma(d)^{-1}z^{d-1}\}\boldsymbol{\Pi}\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Pi}^{'}diag\{\Gamma(d)^{-1}(z+1)^{d-1}\}dz \quad \text{as } s\to\infty\\
&= \left[(\pi_i^{'}\Sigma_\varepsilon\pi_k)\frac{1}{\Gamma d_i\Gamma d_k}\int_0^{\infty} z^{d_i-1}(z+1)^{d_k-1}dz\right], \quad i,k=1,\dots,N.
\end{aligned}
$$

b. The proof is straightforward and is omitted here.

□

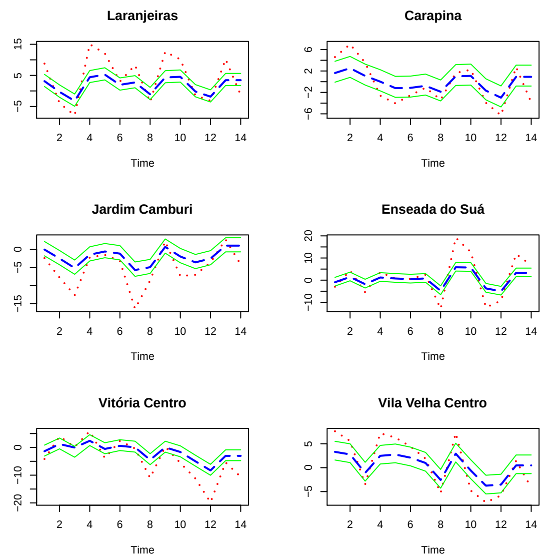


(a) STARMA$(2_{10}, 0)$

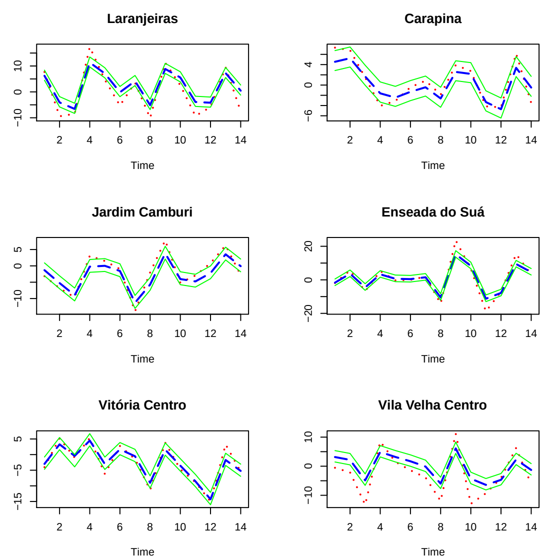


(b) STARFIMA$(2_{10}, \widehat{\boldsymbol{d}}, 0)$

Figure 6 Within-sample prediction (· · · Observed concentrations — Predicted concentrations).



(a) STARMA$(2_{10}, 0)$



(b) STARFIMA$(2_{10}, \widehat{\boldsymbol{d}}, 0)$

Figure 7 Out-of-sample one-step-ahead forecasts for the transformed SO_2 time series ($\cdots$ Observed data $--$ Forecasted data — 95% confidence limits for prediction interval).