# SEASONAL AUTOREGRESSIONS WITH REGIME SWITCHING

R. Paroli, L. Spezia

## 1. INTRODUCTION

Autoregressive processes can be extended to model non-linear and non-normal time series by assuming that different autoregressions, each one depending on a latent regime, alternate according to the regime switching, which is driven by an unobserved Markov chain. These models, widely known as Markov switching autoregressive models (MSARMs), have been introduced in the econometric literature by Hamilton to study economic and financial time series (Hamilton, 1989, 1990, 1993). Krozlig (1997), Kim and Nelson (1999), and Franses and van Dijk (2000), provide generalizations and applications of this class of models. Bayesian analysis of MSARMs has been developed among others by McCulloch and Tsay (1994), Chib (1996), Billio *et al.* (1999), Frühwirth-Schnatter (1999, 2001).

MSARMs are discrete-time stochastic processes $\{Y_t;X_t\}$, so that $\{X_t\}$ is a latent, or hidden, finite-state Markov chain and $\{Y_t\}$, given $\{X_t\}$, satisfies the order-$p$ dependence and the contemporary dependence conditions: we have a sequence of observed random variables $\{Y_t\}$ depending on the $p$ previous observations, whose conditional distributions depend on $\{X_t\}$ only through the contemporary state, or regime, of the Markov chain.

Let $\{X_t\}$ be a discrete-time, first-order, homogeneous, ergodic Markov chain on a finite state-space $S_X$ with cardinality $m$ ($S_X=\{1,\ldots,m\}$). $\boldsymbol{\Gamma}=[\gamma_{ij}]$ is the $(m\times m)$ transition matrix, where $\gamma_{ij}=P(X_t=j\,|\,X_{t-1}=i)$, for any $i$, $j\in S_X$ and any $t=2,\ldots,T$; $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_m)'$ is the stationary distribution, so that $\boldsymbol{\delta}'=\boldsymbol{\delta}'\boldsymbol{\Gamma}$; $\mathbf{x}^T=(x_1,\ldots,x_T)'$ is the sequence of the states of the Markov chain and, for any $t=1,\ldots,T$, $x_t$ has values in $S_X$.

Hence, given the order-$p$ dependence and the contemporary dependence conditions, the equation describing MSARMs is

$$Y_{t(i)} = \mu_i + \sum_{\tau=1}^{p} \varphi_{\tau(i)}\, y_{t-\tau} + E_{t(i)}, \qquad (1)$$

where $Y_{t(i)}$ denotes the conditional variable $Y_t$ when $X_t=i$, for any $1\leq t\leq T$ and for any $i\in S_X$, and $E_{t(i)}$ denotes the Gaussian noise $E_t$ when $X_t=i$, with zero mean and precision $\lambda_i$ ($E_{t(i)}\sim\mathcal{N}(0;\lambda_i)$), for any $i\in S_X$, with the discrete-time process $\{E_t\}$, given $\{X_t\}$, satisfying the conditional independence and the contemporary dependence conditions. Any signal $\mu_i$, any precision $\lambda_i$ and any autoregressive coefficient $\varphi_{\tau(i)}$, for any $\tau = 1,\ldots, p$, depend on the current state $i$ of the Markov chain, for any $i\in S_X$. From equation (1), the conditional distribution of $Y_{t(i)}$, given the $p$ previous observations and the current hidden state, is normal with mean $\mu_i+\sum_{\tau=1}^{p}\varphi_{\tau(i)}\, y_{t-\tau}$ and precision $\lambda_i$, while the marginal distribution of $Y_t$ is a mixture of normals, whose weights are the stationary distribution of the hidden Markov chain.

A special MSARM with a seasonal component is proposed here in the Bayesian framework, giving rise to Seasonal MSARMs (SMSARMs). The paper is organized as follows. SMARMs will be described in Section 2; Bayesian estimators of the parameters of SMARMs will be obtained in Section 3, by means of a Markov chain Monte Carlo (MCMC) algorithm; finally in Section 4 our methodology will be illustrated by analysing a data set about air pollution, also tackling selection of the identifiability constraint, model choice and forecasting.

## 2. SEASONAL MARKOV SWITCHING AUTOREGRESSIVE MODELS

### 2.1. *The basic model*

Our aim is the analysis of seasonal time series characterized by periodic variations with period $s$ and we assume that different seasonalities occur, according to the regime switching. Seasonal components, depending on the hidden states, are modelled by parameters $\beta_{h(i)}$, for any $i\in S_X$ and for any $h=1,\ldots,s$, under the constraint $\sum_{h=1}^{s}\beta_{h(i)}=0$.

It serves our purposes to replace the time $t$ subscript with the $d$ and $h$ subscripts, so that $t=(d-1)s+h$, where $d=1,\ldots,D=T/s$ and $h=1,\ldots,s$. Hence SMSARMs are

$$Y_{[(d-1)s+h](i)}=\mu_i+\sum_{\tau=1}^{p}\varphi_{\tau(i)}\, y_{(d-1)s+h-\tau}+\beta_{h(i)}+E_{[(d-1)s+h](i)}. \qquad (2)$$

### 2.2. *Stationarity*

A sufficient condition for the stationarity of the process (2) is that all the $m$ sub-processes generated by the $m$ states of the chain are stationary, that is, for any

$i \in S_X$, the roots of the auxiliary equations $z^p - \varphi_{1(i)} z^{p-1} - \ldots - \varphi_{p(i)} = 0$, where z is a complex variable, are all inside the unit circle.

To automatically satisfy the constraint on any $\boldsymbol{\varphi}_i = (\varphi_{1(i)}, \ldots, \varphi_{p(i)})'$, we reparametrize $\boldsymbol{\varphi}_i$ in terms of the partial autocorrelations $\mathbf{r}_i = (r_{1(i)}, \ldots, r_{p(i)})'$ of any sub-process, for any $i \in S_X$, according to Barndorff-Nielsen and Schou (1973), and Jones (1987).

The functional relation between $\mathbf{r}_i$ and $\boldsymbol{\varphi}_i$, for any $i \in S_X$, is recursively defined, for $J=1,\ldots,p$:

$$
\begin{aligned}
&g^{1}_{1(i)} = r_{1(i)} && \text{if } J=1 \\
&g^{J}_{K(i)} = g^{J-1}_{K(i)} - r_{J(i)}\, g^{J-1}_{J-K(i)} \text{, for any } K=1,\ldots,J-1 \\
&g^{J}_{J(i)} = r_{J(i)} && \text{if } J>1 \qquad (3) \\
&\boldsymbol{\varphi}_i = (g^{p}_{1(i)}, g^{p}_{2(i)}, \ldots, g^{p}_{p(i)})'.
\end{aligned}
$$

Our inference will be based on the logarithmic transformation $R_{j(i)}$, which maps any partial autocorrelation $r_{j(i)}$ from (-1; 1) to $\mathcal{R}$, for any $j=1,\ldots,p$ and any $i \in S_X$:

$$R_{j(i)} = \ln\left(\frac{1+r_{j(i)}}{1-r_{j(i)}}\right).$$

2.3. *Identification*

Model (2) is unidentifiable in data fitting: when we have $m$ states, we have $m!$ ways to label them; so different models are interchangeable by permuting their labeling. This is the so-called label switching problem (Richardson and Green, 1997, Celeux, Hurn, Robert, 2000, Stephens, 2000, Frühwirth-Schnatter, 2001) and it can be overcome by placing some parameters in increasing or decreasing order. In this paper the special SMSARM with decreasing precisions is analysed ($\lambda_i > \lambda_j$, for any $i, j \in S_X$ so that $i < j$), but the procedures we shall introduce can be easily adapted to any other type of constraint. In Section 4 we shall see how and why we can derive suitable constraints by a data-driven procedure, based on random permutation sampling algorithm (Frühwirth-Schnatter, 2001). Here it is important only to notice that the constraint is chosen ex post after simulations so as to respect the geometry and the shape of the unconstrained posterior distribution, that is different identifiability constraints can be derived by different data sets.

Furthermore to be able to estimate the state-dependent seasonal component, given that $\sum_{h=1}^{s} \beta_{h(i)} = 0$ for any $i \in S_X$, we also need that the hidden chain visits the same hidden state $i$ for all the $s$ times of any sub-period $d$.

2.4. *Prior distributions*

Let $\boldsymbol{\vartheta}$ be the vector of the unknown parameters and latent data of the SMSARM to be estimated,

$$\boldsymbol{\vartheta} = (\boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{x}^D, \mathbf{y}^*)',$$

where $\boldsymbol{\mu}$ is the vector of the $m$ signals $\mu_i$, $\boldsymbol{\lambda}$ is the vector of the $m$ precisions $\lambda_i$, $\mathbf{R}$ is the matrix of the $m$ vectors $\mathbf{R}_i$, i.e. $\mathbf{R}=(\mathbf{R}_1',\ldots,\mathbf{R}_i',\ldots,\mathbf{R}_m')'$ with $\mathbf{R}_i=(R_{1(i)},\ldots,R_{j(i)},\ldots,R_{p(i)})'$, $\boldsymbol{\beta}$ is the matrix of the $m$ seasonal coefficients vectors $\boldsymbol{\beta}_i$, i.e. $\boldsymbol{\beta}=(\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_i',\ldots,\boldsymbol{\beta}_m')'$ with $\boldsymbol{\beta}_i=(\beta_{1(i)},\ldots,\beta_{h(i)},\ldots,\beta_{s(i)})'$, $\mathbf{x}^D=(x_1,\ldots,x_d,\ldots,x_D)'$ is the sequence of the hidden states and $\mathbf{y}^*$ is the vector of all the missing observations $y^*_{(d-1)s+h}$ occuring within the sequence $\mathbf{y}^{sD}=(y_1,\ldots,y_{(d-1)s+h},\ldots,y_{sD})'$.

All the parameters and the latent data will be estimated by simulation, by performing an MCMC algorithm (except for the stationary distribution $\boldsymbol{\delta}$ that will be estimated by the equality $\boldsymbol{\delta}'=\boldsymbol{\delta}'\boldsymbol{\Gamma}$).

For our Bayesian inference, we can place independent Dirichlet priors on each row of $\boldsymbol{\Gamma}$; independent normal priors on each entry of vector $\boldsymbol{\mu}$; independent gamma priors on each entry of vector $\boldsymbol{\lambda}$, under the identifiability constraint; independent normal priors on each entry of matrix $\mathbf{R}$; independent normal priors on the first $s - 1$ entries of any vector $\boldsymbol{\beta}_i$. By the label switching, we need priors that are invariant w.r.t. the $m!$ ways of labeling the states, so that also the posterior distribution from which we sample is relabeling-invariant.

2.5. *Posterior distribution*

The posterior distribution of $\boldsymbol{\vartheta}$ is

$$\begin{aligned}\pi(\boldsymbol{\vartheta}\,|\,\mathbf{y}^{sD}, \mathbf{y}^0) &= f(\boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{x}^D, \mathbf{y}^*\,|\,\mathbf{y}^{sD}, \mathbf{y}^0) \propto \\ &\propto f(\mathbf{y}^{sD}, \mathbf{y}^*\,|\,\boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{x}^D, \mathbf{y}^0)\, f(\mathbf{x}^D\,|\,\boldsymbol{\Gamma})\, p(\boldsymbol{\Gamma})\, p(\boldsymbol{\mu})\, p(\boldsymbol{\lambda})\, p(\mathbf{R})\, p(\boldsymbol{\beta}),\end{aligned}$$

where $\mathbf{y}^0=(y_{-p+1},\ldots,y_0)'$ are the initial values fixed for the $p$-dependence condition,

$$\begin{aligned}&f(\mathbf{y}^{sD}, \mathbf{y}^*\,|\,\boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{x}^D, \mathbf{y}^0) = \\ &\qquad = \prod_{d=1}^{D}\prod_{h=1}^{s} f(y_{(d-1)s+h}\,|\,y_{(d-1)s+h-1},\ldots,y_{(d-1)s+h-p}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, x_d, \mathbf{y}^0), \qquad (4)\end{aligned}$$

by the order-$p$ dependence and the contemporary dependence conditions, with

$$\begin{aligned}&f(y_{(d-1)s+h}\,|\,y_{(d-1)s+h-1},\ldots,y_{(d-1)s+h-p}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, x_d, \mathbf{y}^0) = \\ &\qquad = \sqrt{\frac{\lambda_{x_d}}{2\pi}}\exp\left\{-\frac{\lambda_{x_d}}{2}\left(y_{(d-1)s+h} - \mu_{x_d} - \sum_{\tau=1}^{p}\varphi_{\tau(x_d)}\, y_{(d-1)s+h-\tau} - \beta_{h(x_d)}\right)^2\right\},\end{aligned}$$

for any $d=1,\ldots, D$ and any $h=1,\ldots, s$, and

$$f(\mathbf{x}^D|\boldsymbol{\Gamma}) = \delta_{x_1}\prod_{d=2}^{D}\gamma_{x_{d-1},x_d} = \delta_{x_1}\prod_{i=1}^{m}\prod_{j=1}^{m}\gamma_{i,j}^{D_{i,j}},$$

by the Markov dependence condition, where $D_{i,j}$ is the number of couples of consecutive hidden states $i, j$. Notice that in the r.h.s. of (4) there are no missing observations: if one or more missing observations occur within $\mathbf{y}^{sD}$, any missing observation will be replaced by the corresponding simulated value $y^*_{(d-1)s+h}$.

## 3. PARAMETER ESTIMATION OF SMSARMs

The Metropolis-within-Gibbs algorithm associated with the constrained permutation sampling algorithm is now developed for the special SMSARM with decreasing precisions, noticing that this scheme can be easily rearranged whenever another type of identifiability constraint is imposed. To be able to perform permutation sampling, all the priors must be invariant to relabelling the states, i.e. their hyperparameters must not depend on the hidden states. We shall not describe in detail the iterative scheme of the Metropolis-within-Gibbs algorithm, that can be seen for example in Gamerman, 1997.

Here we can analyse the generic $k$-th iteration of the MCMC sampler only, remembering that at the $(k-1)$-th iteration the vector $\boldsymbol{\vartheta}^{(k-1)}$ has been generated,

$$\boldsymbol{\vartheta}^{(k-1)} = (\boldsymbol{\Gamma}^{(k-1)}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\lambda}^{(k-1)}, \mathbf{R}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \mathbf{x}^{D(k-1)}, \mathbf{y}^{*(k-1)})',$$

and the identifiability constraint on the precision has been chosen, $\lambda_i^{(k-1)} > \lambda_j^{(k-1)}$, for any $i, j \in S_X$ so that $i < j$.

1) The sequence $\mathbf{x}^{D(k)}$ of hidden states is generated in block from the full conditional $\pi(\mathbf{x}^D|\mathbf{y}^{sD}, \boldsymbol{\Gamma}^{(k-1)}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\lambda}^{(k-1)}, \mathbf{R}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \mathbf{y}^{*(k-1)}, \mathbf{y}^0)$, by means of the procedure proposed by Chib, 1996, based on the forward filtering-backward sampling (*ff-bs*) algorithm by Carter and Kohn, 1994, and Frühwirth-Schnatter, 1994, for state-space models. The *ff-bs* algorithm is so called because first the filtered probabilities of the hidden states are computed going forwards; then the conditional probabilities of the hidden states are computed going backwards, sampling the states from the full conditional,

$$\pi(\mathbf{x}^D|\mathbf{y}^{sD}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{y}^*, \mathbf{y}^0) =$$
$$= \pi(x_D|\mathbf{y}^{sD}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{y}^*, \mathbf{y}^0)\prod_{d=1}^{D-1}\pi(x_d|\ x_{d+1}, \mathbf{y}^{sd}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{y}^*, \mathbf{y}^0).$$

Let $\boldsymbol{\xi}_{d+1|d}$ be the $m$-dimensional vector whose generic entry is $P(X_{d+1}=i|\mathbf{y}^{sd},\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\lambda},\mathbf{R},\boldsymbol{\beta},\mathbf{y}^*,\mathbf{y}^0)$, for any $i=1,\ldots, m$; $\boldsymbol{\xi}_{d|d}$ be the $m$-dimensional vector

whose generic entry is $P(X_d=i|\mathbf{y}^{sd},\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\lambda},\mathbf{R},\boldsymbol{\beta},\mathbf{y}^*,\mathbf{y}^0)$, for any $i=1,\ldots, m$; $\boldsymbol{\xi}_d$ be the $m$-dimensional vector whose generic entry is $P(X_d=i|X_{d+1}=x_{d+1},\mathbf{y}^{sd},\boldsymbol{\Gamma},\boldsymbol{\mu},\boldsymbol{\lambda},\mathbf{R},\boldsymbol{\beta},\mathbf{y}^*,\mathbf{y}^0)$, for any $i=1,\ldots, m$. The iterative scheme of the *ff-bs* algorithm is the following.

1.1) Compute

$$\boldsymbol{\xi}_{1|0}^{(k)}=\boldsymbol{\delta}'^{(k-1)}=\boldsymbol{\delta}'^{(k-1)}\boldsymbol{\Gamma}^{(k-1)},$$

that is $\boldsymbol{\delta}^{(k-1)}$ is the left eigenvector of the matrix $\boldsymbol{\Gamma}^{(k-1)}=[\gamma_{i,j}^{(k-1)}]$, associated with the eigenvalue equal to one.

1.2) Compute

$$\boldsymbol{\xi}_{d|d}^{(k)}=\frac{\boldsymbol{\xi}_{d|d-1}^{(k)}\mathbf{F}_d^{(k-1)}}{\mathbf{1}'_{(m)}(\boldsymbol{\xi}_{d|d-1}^{(k)}\mathbf{F}_d^{(k-1)})} \qquad \text{and} \qquad \boldsymbol{\xi}_{d+1|d}^{(k)}=\boldsymbol{\Gamma}'^{(k-1)}\boldsymbol{\xi}_{d|d}^{(k)},$$

for any $d=1,\ldots, D-1$, where $\mathbf{F}_d^{(k-1)}=\text{diag}[\prod_{h=1}^{s} f(y_{(d-1)s+h}|y_{(d-1)s+h-1},\ldots, y_{(d-1)s+h-p},\boldsymbol{\mu}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{R}^{(k-1)},\boldsymbol{\beta}^{(k-1)},\mathbf{y}^0,x_d^{(k-1)}=1), \ldots, \prod_{h=1}^{s} f(y_{(d-1)s+h}|y_{(d-1)s+h-1},\ldots,y_{(d-1)s+h-p},\boldsymbol{\mu}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{R}^{(k-1)},\boldsymbol{\beta}^{(k-1)},\mathbf{y}^0,x_d^{(k-1)}=m)]$ and $\mathbf{1}_{(m)}$ is the $m$-dimensional vector of ones.

1.3) Compute

$$\boldsymbol{\xi}_{D|D}^{(k)}=\frac{\boldsymbol{\xi}_{D|D-1}^{(k)}\mathbf{F}_D^{(k-1)}}{\mathbf{1}'_{(m)}(\boldsymbol{\xi}_{D|D-1}^{(k)}\mathbf{F}_D^{(k-1)})}$$

(for details on the derivation of formulae at steps 1.2 and 1.3, see Hamilton, 1994, pp. 692-693).

1.4) Generate $x_D^{(k)}$ from $\boldsymbol{\xi}_{D|D}^{(k)}$.

1.5) Compute

$$\boldsymbol{\xi}_d^{(k)}=\frac{\boldsymbol{\xi}_{d|d}^{(k)}\boldsymbol{\Gamma}_{\bullet x_{d+1}^{(k)}}^{(k-1)}}{\mathbf{1}'_{(m)}(\boldsymbol{\xi}_{d|d}^{(k)}\boldsymbol{\Gamma}_{\bullet x_{d+1}^{(k)}}^{(k-1)})}$$

and generate $x_d^{(k)}$ from $\boldsymbol{\xi}_d^{(k)}$, for any $d=D-1,\ldots,1$. $\boldsymbol{\Gamma}_{\bullet x_{d+1}^{(k)}}^{(k-1)}$ represents the column of $\boldsymbol{\Gamma}^{(k-1)}$ corresponding to the state previously generated.

2) Placing a gamma prior $\mathcal{G}(\alpha_\Lambda; \beta_\Lambda)$ on any $\lambda_i$, the parameters $\lambda_i^{(k)}$, for any $i \in S_X$, are independently generated from a gamma distribution with parameters

$$\frac{D_i^{(k)}}{2} + \alpha_\Lambda$$

and

$$\frac{1}{2} \sum_{\{d \geq 1:\, x_d^{(k)} = i\}} \sum_{h=1}^{s} \left( y_{(d-1)s+h} - \mu_i^{(k-1)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k-1)} y_{(d-1)s+h-\tau} - \beta_{h(i)}^{(k-1)} \right)^2 + \beta_\Lambda ,$$

where $D_i^{(k)}$ is the number of observations corresponding to the contemporary hidden state $i$ in the sequence $\mathbf{x}^{D(k)}$ generated at step 1).

The entries of the vector $\boldsymbol{\lambda}^{(k)}$ must be in decreasing order to satisfy the identifiability constraint: $\lambda_i^{(k)} > \lambda_j^{(k)}$, for any $i, j \in S_X$, so that $i < j$. If $\boldsymbol{\lambda}^{(k)}$ is not ordered, instead of rejecting the vector and going on sampling till we have an ordered one, we introduce the constrained permutation sampling algorithm (Frühwirth-Schnatter, 2001): we have $m$ couples $(i, \lambda_i^{(k)})$; if the $\lambda_i^{(k)}$'s are unordered, we apply a permutation $\rho(\cdot)$ to order them; consequently also the corresponding $i$'s must be permuted according to the permutation $\rho(\cdot)$, $\rho(S_X) = \{\rho(1), \ldots, \rho(m)\}$; finally the permutation $\rho(S_X)$ is extended to the generated sequence of states $\mathbf{x}^{D(k)}$, $\rho(\mathbf{x}^{D(k)}) = (\rho(x_1^{(k)}), \ldots, \rho(x_d^{(k)}), \ldots, \rho(x_D^{(k)}))'$, and to the switching-parameters previously generated, $\rho(\boldsymbol{\Gamma}^{(k-1)})$, $\rho(\boldsymbol{\mu}^{(k-1)})$, $\rho(\mathbf{R}^{(k-1)})$, $\rho(\boldsymbol{\varphi}^{(k-1)})$, $\rho(\boldsymbol{\beta}^{(k-1)})$, where $\boldsymbol{\varphi}^{(k-1)}$ are obtained from $\mathbf{R}^{(k-1)}$ by means of (3).

Notice that if we had had a different constraint, either on the means or on the diagonal entries of transition matrix, its corresponding full-conditional would have been placed at this step.

3) Placing a normal prior $\mathcal{N}(\mu_M; \lambda_M)$ on any $\mu_i$, the parameters $\mu_i^{(k)}$, for any $i \in S_X$, are independently generated from a normal distribution with mean

$$\frac{\rho(\lambda_i^{(k)}) \sum_{\{d \geq 1:\, \rho(x_d^{(k)}) = i\}} \sum_{h=1}^{s} \left( y_{(d-1)s+h} - \sum_{\tau=1}^{p} \rho(\varphi_{\tau(i)}^{(k-1)}) y_{(d-1)s+h-\tau} - \rho(\beta_{h(i)}^{(k-1)}) \right) + \mu_M \lambda_M}{\rho(D_i^{(k)}) \rho(\lambda_i^{(k)}) + \lambda_M}$$

and precision

$$\rho(D_i^{(k)}) \rho(\lambda_i^{(k)}) + \lambda_M ,$$

where $\rho(D_i^{(k)})$ is the number of observations corresponding to the contemporary hidden state $i$ in the permuted sequence $\rho(\mathbf{x}^{D(k)})$.

4) Placing a normal prior $\mathcal{N}(\mu_R; \lambda_R)$ on any $R_{j(i)}$, the parameters $R_{j(i)}^{(k)}$, for any $j=1,\ldots, p$ and any $i\in S_X$, are independently generated from the random walk $R_{j(i)}^{(k)} = \rho(R_{j(i)}^{(k-1)}) + U$, where $U$ is a Gaussian noise with zero mean and constant precision, for any $k$. Then any vector $\mathbf{R}_i$ is accepted with probability

$$\alpha(\rho(\mathbf{R}_i^{(k-1)}); \mathbf{R}_i^{(k)}) = \min\left\{1, \frac{\pi(\mathbf{R}_i^{(k)} \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \rho(\boldsymbol{\beta}^{(k-1)}), \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k-1)}, \mathbf{y}^0, \mathbf{y}^{sD})}{\pi(\rho(\mathbf{R}_i^{(k-1)}) \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \rho(\boldsymbol{\beta}^{(k-1)}), \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k-1)}, \mathbf{y}^0, \mathbf{y}^{sD})}\right\}$$

for any $i\in S_X$, where

$$\pi(\mathbf{R}_i^{(k)} \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \rho(\boldsymbol{\beta}^{(k-1)}), \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k-1)}, \mathbf{y}^0, \mathbf{y}^{sD}) \propto$$

$$\propto \exp\left\{ \frac{\rho(\lambda_i^{(k)})}{2} \sum_{\{d\geq 1:\, \rho(x_d^{(k)})=i\}} \sum_{h=1}^{s} \left( y_{(d-1)s+h} - \mu_i^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k-1)} y_{(d-1)s+h-\tau} + \right.\right.$$

$$\left.\left. - \rho(\beta_{h(i)}^{(k-1)}) \right)^2 - \frac{\lambda_R}{2} \sum_{j=1}^{p} (R_{j(i)}^{(k)} - \mu_R)^2 \right\}$$

and

$$\pi(\rho(\mathbf{R}_i^{(k-1)}) \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \rho(\boldsymbol{\beta}^{(k-1)}), \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k-1)}, \mathbf{y}^0, \mathbf{y}^{sD}) \propto$$

$$\propto \exp\left\{ \frac{\rho(\lambda_i^{(k)})}{2} \sum_{\{d\geq 1:\, \rho(x_d^{(k)})=i\}} \sum_{h=1}^{s} \left( y_{(d-1)s+h} - \mu_i^{(k)} - \sum_{\tau=1}^{p} \rho(\varphi_{\tau(i)}^{(k-1)}) y_{(d-1)s+h-\tau} + \right.\right.$$

$$\left.\left. - \rho(\beta_{h(i)}^{(k-1)}) \right)^2 - \frac{\lambda_R}{2} \sum_{j=1}^{p} (\rho(R_{j(i)}^{(k-1)}) - \mu_R)^2 \right\}.$$

5) Placing a normal prior $\mathcal{N}(\mu_B; \lambda_B)$ on any $\beta_{h(i)}$, the parameters $\beta_{h(i)}^{(k)}$, for any $h=1,\ldots, s-1$ and any $i\in S_X$, are independently generated from a normal distributions with mean

$$\frac{\rho(\lambda_i^{(k)}) \sum_{\{d\geq 1:\, \rho(x_d^{(k)})=i\}} (y_{(d-1)s+h} - \mu_i^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k)} y_{(d-1)s+h-\tau}) + \mu_B \lambda_B}{\rho(D_{h;i}^{(k)}) \rho(\lambda_i^{(k)}) + \lambda_B}$$

and precision

$$\rho(D_{h;i}^{(k)}) \rho(\lambda_i^{(k)}) + \lambda_B \ ,$$

where $\rho(D_{h;i}^{(k)})$ is the number of observations corresponding to the contemporary hidden state $i$, recorded in time $h$, in the permuted sequence $\rho(\mathbf{x}^{D(k)})$. The last entry of $\boldsymbol{\beta}_i^{(k)}$, i.e. $\beta_{s(i)}^{(k)}$, is obtained by difference w.r.t. the constraint $\sum_{h=1}^{s} \beta_{h(i)} = 0$:

$$\beta_{s(i)}^{(k)} = - \sum_{h=1}^{s-1} \beta_{h(i)}^{(k)} .$$

6) Let $\boldsymbol{\Gamma}_{i\bullet}=(\gamma_{i,1}, \gamma_{i,2},\ldots, \gamma_{i,m})$, be the $i$-th row of $\boldsymbol{\Gamma}$. Placing a Dirichlet prior with parameter $\boldsymbol{\omega}=(\omega_1,\ldots, \omega_m)$ on $\boldsymbol{\Gamma}_{i\bullet}$, each row $\boldsymbol{\Gamma}_{i\bullet}^{(k)}$, for any $i \in S_X$, is independently generated from a Dirichlet $\mathcal{D}(\boldsymbol{\omega} + \rho(D_{i\bullet}^{(k)}))$, where $\rho(D_{i\bullet}^{(k)})=(\rho(D_{i;1}^{(k)}), \ldots, \rho(D_{i;m}^{(k)}))$ and $\rho(D_{i;j}^{(k)})$ is the number of couples of consecutive hidden states $i, j$ in the permuted sequence $\rho(\mathbf{x}^{D(k)})$, for any $i, j \in S_X$.

7) Every missing observation $y^*_{(d-1)s+h}$ is generated from the normal distribution

$$\mathcal{N}\left( \mu_{\rho(x_{d^{(k)}})}^{(k)} + \sum_{\tau=1}^{p} \varphi_{\tau(\rho(x_{d^{(k)}}))}^{(k)} y_{(d-1)s+h-\tau} + \beta_{h(\rho(x_{d^{(k)}}))}^{(k)}, \rho(\lambda_{\rho(x_{d^{(k)}})}^{(k)}) \right). \quad (5)$$

Now, at the end of the $k$-th iteration of the MCMC sampler, the vector $\boldsymbol{\vartheta}^{(k)}$ has been simulated from $\pi(\boldsymbol{\vartheta} \mid \mathbf{y}^{sD}, \mathbf{y}^0)$, if $k$ is large enough. We shall repeat these steps till we have an $N$-dimensional sample. This sample will be used to estimate each entry of $\boldsymbol{\vartheta}$ by means of posterior means, but the sequence of states, estimated through posterior modes.

## 4. APPLICATION TO OZONE DATA

An application of SMSARMs to real data will be studied in the following: we shall analyse the time series of the hourly mean concentrations of Ozone ($O_3$), in micrograms per cubic meter ($\mu g/m^3$), recorded by the air pollution testing station placed in Via San Giorgio, Bergamo (Italy), from April 1st, 1998, 1 a.m., to September 30th, 1998, 12 p.m. (4392 observations).

The analysed data are the natural logarithms of $O_3$ concentrations, while the latent data are made up both by the sequence of the hidden states and by 549 missing values occuring within the observed series. In the series (Figure 1a) a daily periodicity ($s$=24) is evident and it is confirmed by the 120 hours correlogram (Figure 1b); moreover the non-normality of the series emerges from the histogram of the data (Figure 1c). Our data set is collected in the six warmest months, because $O_3$ is a substance which is not directly emitted in the air, but it forms owing to complex chemical reactions in the presence of solar radiation and high temperature.

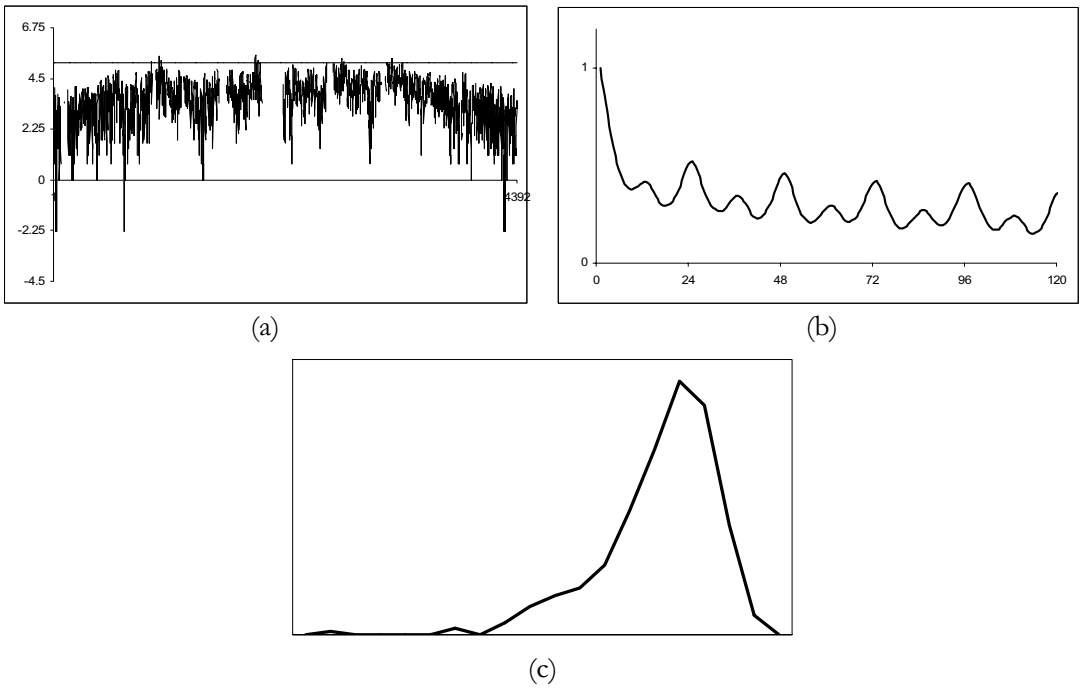



*Figure 1* – Series of the $O_3$ hourly mean log-concentrations with the log of the attention-level (a); the 120 hours correlogram (b) and the continuous approximation of the histogram (c) of the data plotted in (a).

We shall develop our empirical analysis in four steps: *i*) constraint identification, *ii* ) model selection, *iii* ) parameter estimation, *iv* ) forecasting.

The following hyperparameters have been chosen for all the models and used in all the four steps of our empirical analysis:

– $\omega_{ij}=m\cdot I(i=j)+0.6\cdot I(i\neq j)$, for any $i, j=1,\ldots, m$, where $I(A)$ is the indicator function that takes the value of 1 if $A$ is true and the value of 0 otherwise, i.e. the probability of persistence is greater than the probability of transition; the probability of persistence is about 0.7 and it slowly decreases as the number of states increases;

– $\mu_M=\ln(180/2)$ and $\lambda_M=0.3$, i.e. the concentrations of $O_3$ are in the middle of the tolerance interval defined by the attention-level ($180\mu g/m^3$);

– $\alpha_\Lambda=\beta_\Lambda=0.5$, i.e. each precision is assumed a priori to follow a gamma with mean 1 and variance 2, leading to low variabilty within each state;

– $\mu_R=0$ and $\lambda_R=0.1$, i.e the prior information on any $R_{j(i)}$ is quite vague;

– $\mu_B=0$ and $\lambda_B=0.1$, i.e the prior information on any $\beta_{b(i)}$ is quite vague.

### 4.1. *Constraint identification*

At the beginning of our analysis we have to investigate the consistency of the cardinality of the state-space of the hidden Markov chain and to select some suitable identifiability constraint. By plotting the couples of the outputs of the estimates, obtained via unconstrained Metropolis-within-Gibbs algorithm performed associated with random permutation sampling (Frühwirth-Schnatter, 2001), we can check if there are as many groups as the hidden states and if these groups can

suggest special ordering in their labeling. When the MCMC sampler runs unconstrained, it takes advantages from random permutation sampling to explore the whole support of the posterior distribution and the sampling Markov chain randomly switches from the current subspace, defined by the current labeling, to one of the other (*m!* - 1): w.r.t. the procedure described in Section 3, the permutation $\rho(\cdot)$ is selected not to respect a special constraint, but it is randomly generated at each iteration.

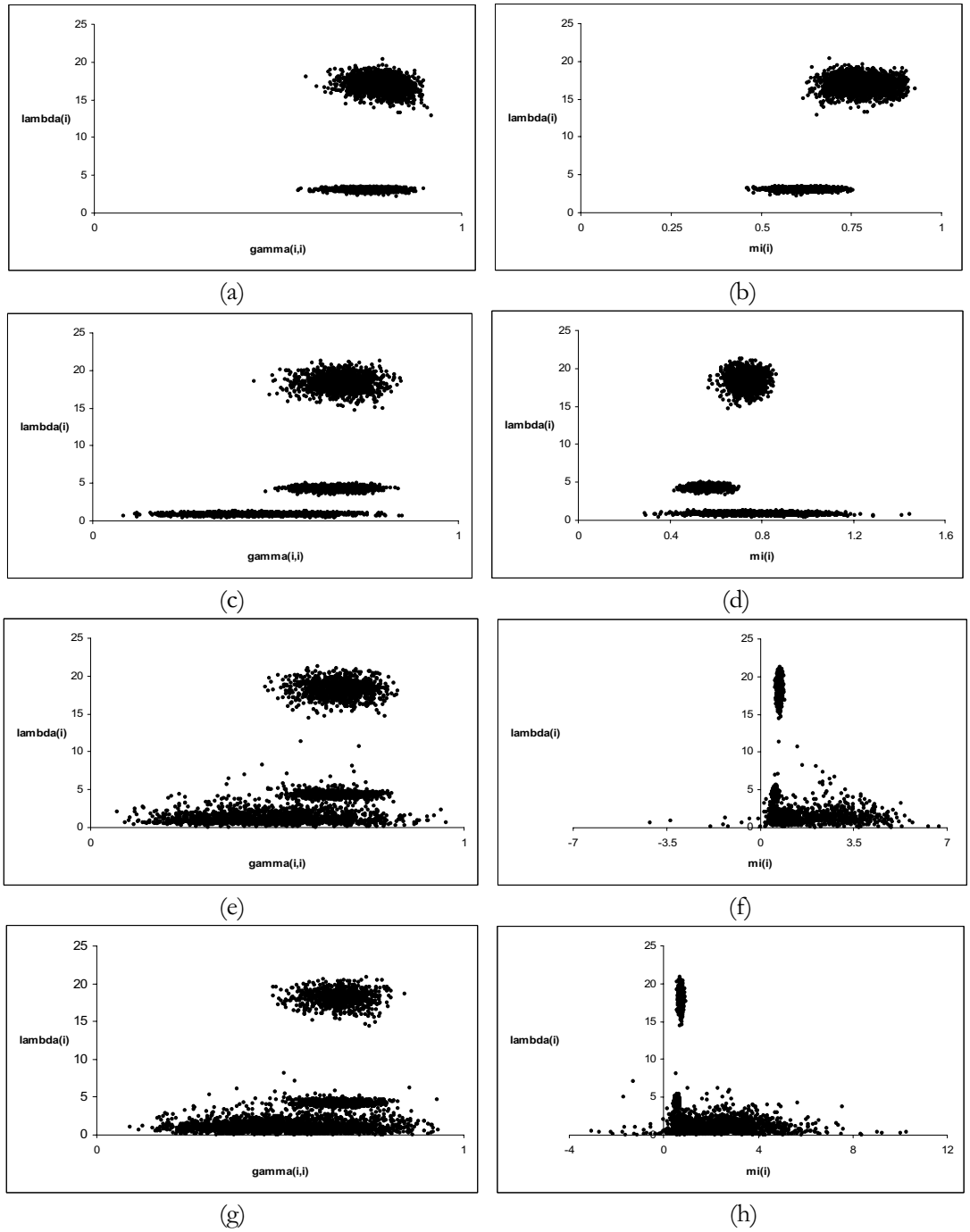



*Figure 2* – Some outputs of unconstrained Metropolis-within-Gibbs algorithm with random permutations for $p$=2 and $m$=2 (a; b), $m$=3 (c; d), $m$=4 (e; f), $m$=5 (g; h).

We can compare the outputs of thirty competing models, henceforth said SMSAR($m;p$), which differ for the cardinality of the state-space of the hidden Markov chain ($m$=1,…,5) and for the order of the autoregressive process ($p$=0,…,5). By graphically analysing the outputs of the unconstrained SMSAR($m;p$), first we can notice that, for any $p$, $m$=4 and $m$=5 are not consistent with the data we are studying, because four or five groups do not emerge in any plot (Figures 2e - 2h), hence we shall develop our analysis for $m$=1;2;3 only. After that we can select the constraint on the precisions ($\lambda_i > \lambda_j$, for any $i, j \in S_X$, so that $i < j$) because the decreasing ordering is evident in any graph (Figures 2a - 2d). Decreasing precisions is a reasonable constraint, because when the low hidden state occurs, the variability of $O_3$ data depending on it is low and the concentrations of pollution are also low; by contrast when the high hidden state occurs, the variability of $O_3$ data depending on it is high and the concentrations of pollution are also high.

4.2. *Model choice*

Model choice will be performed by means of Bayes factors (Kass and Raftery, 1995) in which the marginal likelihoods, i.e. the normalizing constants of the posterior densities, are computed according to Chib (1995), and Chib and Jeliazkov (2001), through the relabeling of the hidden states by means of constrained permutation sampling.

The natural logarithm of the marginal likelihood, $\ln f(\mathbf{y}^{sD} | \mathbf{y}^0)$, is estimated in a special point $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*)'$, the posterior mode of $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{R}, \boldsymbol{\beta}, \boldsymbol{\Gamma})'$, and we obtain the estimate $\ln \hat{f}(\mathbf{y}^{sD} | \mathbf{y}^0)$:

$$\ln \hat{f}(\mathbf{y}^{sD} | \mathbf{y}^0) = \ln f(\mathbf{y}^{sD} | \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*, \mathbf{y}^0) + \ln p(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*) + -\ln \hat{\pi}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^* | \mathbf{y}^{sD}, \mathbf{y}^0). \quad (6)$$

The first expression in the r.h.s. of (6) is

$$\ln f(\mathbf{y}^{sD} | \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*, \mathbf{y}^0) = \sum_{d=1}^{D} \sum_{h=1}^{s} \ln \left( \sum_{i=1}^{m} f(y_{(d-1)s+h} | y_{(d-1)s+h-1}, \ldots, y_{(d-1)s+h-p}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0, x_d = i) \cdot \cdot P(X_d = i | \mathbf{y}^{d-1}, \boldsymbol{\Gamma}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0) \right) \quad (7)$$

where $P(X_d = i | \mathbf{y}^{d-1}, \boldsymbol{\Gamma}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0)$, for any $d$=1,…, $D$ and for any $i$=1,…,$m$, is the filtered probability (see Section 3).

The second expression of r.h.s. of (6) becomes

$$\ln p(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*) = \ln p(\boldsymbol{\mu}^*) + \ln p(\boldsymbol{\lambda}^*) + \ln p(\mathbf{R}^*) + \ln p(\boldsymbol{\beta}^*) + \ln p(\boldsymbol{\Gamma}^*)$$

and the third can be decomposed as

$$\ln \hat{\pi}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^* \mid \mathbf{y}^{sD}, \mathbf{y}^0) =$$
$$= \ln\left[\frac{1}{N}\sum_{k=1}^{N}(\alpha(\mathbf{R}^{(k)};\mathbf{R}^*)\, q(\mathbf{R}^{(k)};\mathbf{R}^* \mid \mathbf{y}^{sD},\mathbf{y}^0,\boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}),\boldsymbol{\beta}^{(k)},\boldsymbol{\Gamma}^{(k)},\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)}))\right] +$$
$$-\ln\left(\frac{1}{N}\sum_{k=1}^{N}\alpha(\mathbf{R}^*;\mathbf{R}^{(k)})\right)+\ln\left(\frac{1}{N}\sum_{k=1}^{N}\pi(\boldsymbol{\mu}^* \mid \mathbf{y}^{sD},\mathbf{y}^0, \rho(\boldsymbol{\lambda}^{(k)}),\mathbf{R}^*,\boldsymbol{\beta}^{(k)},\boldsymbol{\Gamma}^{(k)},\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)})\right)+$$
$$+\ln\left(\frac{1}{N}\sum_{k=1}^{N}\pi(\boldsymbol{\beta}^* \mid \mathbf{y}^{sD},\mathbf{y}^0, \boldsymbol{\mu}^*, \rho(\boldsymbol{\lambda}^{(k)}),\mathbf{R}^*,\boldsymbol{\Gamma}^{(k)},\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)})\right) +$$
$$+\ln\left(\frac{1}{N}\sum_{k=1}^{N}\pi(\boldsymbol{\Gamma}^* \mid \mathbf{y}^{sD},\mathbf{y}^0, \boldsymbol{\mu}^*, \rho(\boldsymbol{\lambda}^{(k)}),\mathbf{R}^*,\boldsymbol{\beta}^*,\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)})\right) +$$
$$+\ln\left(\frac{1}{N}\sum_{k=1}^{N}\pi(\boldsymbol{\lambda}^* \mid \mathbf{y}^{sD},\mathbf{y}^0, \boldsymbol{\mu}^*, \mathbf{R}^*,\boldsymbol{\beta}^*,\boldsymbol{\Gamma}^*,\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)})\right),$$

and estimated using 6·N extra-iterations, labelled by $k$, of the MCMC sampler, where $q(\mathbf{R}^{(k)};\mathbf{R}^* \mid \mathbf{y}^{sD},\mathbf{y}^0,\boldsymbol{\mu}^{(k)},\rho(\boldsymbol{\lambda}^{(k)}),\boldsymbol{\beta}^{(k)},\boldsymbol{\Gamma}^{(k)},\rho(\mathbf{x}^{D(k)}),\mathbf{y}^{*(k)})$ is the probability density function (*pdf*) of multivariate normal with mean $R^{(k)}_{j(i)}$ and precision matrix $U$ evaluated in $R^{*}_{j(i)}$;

$$\alpha(\mathbf{R}^{(k)};\mathbf{R}^*) = \min\left\{1; \frac{\pi(\mathbf{R}^*_i \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \boldsymbol{\beta}^{(k)}, \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k)}, \mathbf{y}^0, \mathbf{y}^{sD})}{\pi(\mathbf{R}^{(k)}_i \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \boldsymbol{\beta}^{(k)}, \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k)}, \mathbf{y}^0, \mathbf{y}^{sD})}\right\}$$

$$\alpha(\mathbf{R}^*;\mathbf{R}^{(k)}) = \min\left\{1; \frac{\pi(\mathbf{R}^{(k)}_i \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \boldsymbol{\beta}^{(k)}, \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k)}, \mathbf{y}^0, \mathbf{y}^{sD})}{\pi(\mathbf{R}^*_i \mid \boldsymbol{\mu}^{(k)}, \rho(\boldsymbol{\lambda}^{(k)}), \boldsymbol{\beta}^{(k)}, \rho(\mathbf{x}^{D(k)}), \mathbf{y}^{*(k)}, \mathbf{y}^0, \mathbf{y}^{sD})}\right\}.$$

Notice that all the values, labelled by $k$, are drawn from their respective full conditionals, except those used to build $\alpha(\mathbf{R}^*; \mathbf{R}^{(k)})$ which are drawn from the multivariate normals $R^{(k)}_{j(i)}=R^{*}_{j(i)}+U$, for any $j$=1,…, $p$ and $i$=1,…, $m$.

When a current observation $y^*_{(d-1)s+b}$ is missing, its corresponding *pdf*s in (7) are replaced with 1, for any $i$=1,…, $m$. Hence, we introduce in the recursive computations of the filtered probabilities the powers of the transition matrix, which, by the Chapman-Kolmogorov equations, represent a skip of the missing observation. By contrast, when missing observations occur among the $p$ previous observations, they are replaced by the expected values

$$E(Y_{(d-1)s+b} \mid \mathbf{y}^{d-1}, \boldsymbol{\Gamma}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0) =$$
$$= \sum_{i=1}^{m} E(Y_{(d-1)s+b} \mid \mathbf{y}^{d-1}, \boldsymbol{\Gamma}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0, x_d=i)\, P(X_d=i \mid \mathbf{y}^{d-1}, \boldsymbol{\Gamma}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \mathbf{R}^*, \boldsymbol{\beta}^*, \mathbf{y}^0),$$

for any $i$=1,…, $m$.

From the values of the marginal likelihoods in Table 1, we can notice that the SMSAR(*3;2*) is the best among all the competing models.

TABLE 1

*Marginal likelihoods of the competing SMSARMs*

| *m**p* | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | -5148.514 | -2349.522 | -2377.124 | -2263.395 | -2363.223 | -2350.491 |
| 2 | -3767.218 | -1819.074 | -1816.076 | -1753.605 | -1797.981 | -1656.859 |
| 3 | -3457.887 | -1651.769 | **-1410.689** | -1746.917 | -1559.064 | -1527.177 |

## 4.3. *Parameter estimation*

Now we can deal with the results we obtained for the best model, SMSAR(*3;2*). The estimates of the transition matrix of the hidden Markov chain are

$$\boldsymbol{\Gamma}=\begin{bmatrix} 0.734 & 0.227 & 0.039 \\ 0.242 & 0.714 & 0.044 \\ 0.151 & 0.514 & 0.335 \end{bmatrix},$$

from which we have the estimate of the stationary initial distribution, by $\boldsymbol{\delta}'=\boldsymbol{\delta}'\boldsymbol{\Gamma}$,

$$\boldsymbol{\delta}=(0.466;\ 0.476;\ 0.058)',$$

while those of the parameters of the three Gaussian *pdf* are

| *i* | $\mu_i$ | $\lambda_i$ | $\varphi_{1(i)}$ | $\varphi_{2(i)}$ |
|---|---|---|---|---|
| 1 | 0.701 | 18.769 | 0.832 | -0.001 |
| 2 | 0.558 | 4.006 | 0.862 | -0.039 |
| 3 | 0.830 | 0.333 | 0.496 | 0.160 |

where the $\varphi_i$'s have been obtained by

| *i* | $R_{1(i)}$ | $R_{2(i)}$ |
|---|---|---|
| 1 | 2.390 | -0.001 |
| 2 | 2.379 | -0.078 |
| 3 | 1.359 | 0.324 |

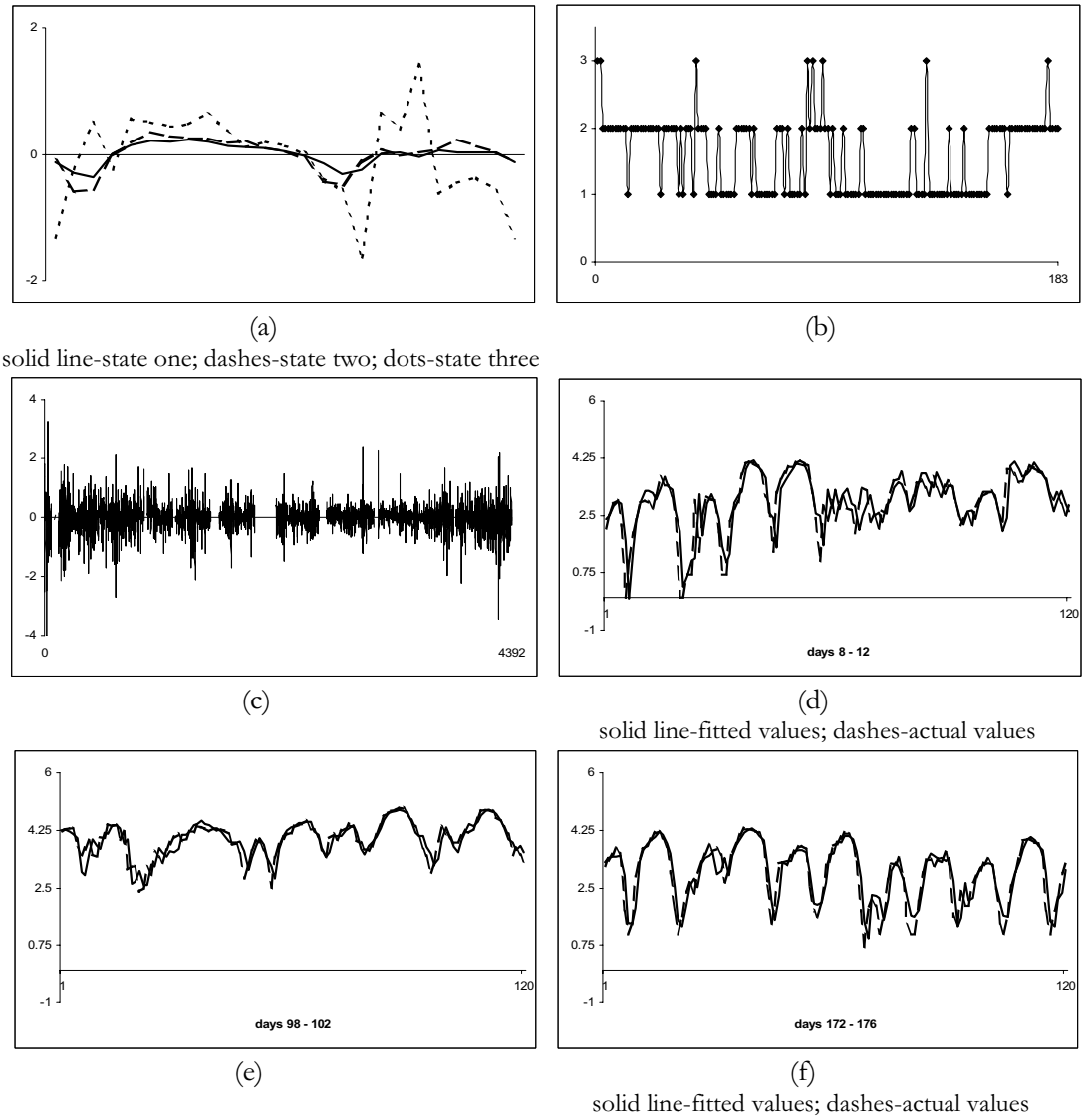
and

| | $\boldsymbol{\beta}_i$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *i**b* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 0.055 | 0.030 | 0.199 | 0.026 | -0.123 | -0.299 | -0.363 | 0.002 |
| 2 | 0.078 | 0.235 | 0.133 | 0.018 | -0.110 | -0.600 | -0.557 | -1.564 |
| 3 | -0.620 | -0.456 | -0.372 | -0.580 | -1.338 | -0.242 | 0.510 | -0.253 |
| *i**b* | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 1 | 0.153 | 0.219 | 0.200 | 0.232 | 0.189 | 0.134 | 0.111 | 0.098 |
| 2 | 0.175 | 0.356 | 0.288 | 0.255 | 0.246 | 0.180 | 0.213 | 0.097 |
| 3 | 0.562 | 0.507 | 0.434 | 0.475 | 0.662 | 0.367 | 0.104 | 0.195 |
| *i**b* | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 1 | 0.044 | -0.023 | -0.150 | -0.312 | -0.256 | 0.015 | 0.032 | -0.037 |
| 2 | 0.038 | -0.077 | -0.433 | -0.486 | -0.121 | 0.073 | -0.022 | 0.031 |
| 3 | 0.136 | 1.628 | -0.390 | -0.551 | -1.640 | 0.667 | 0.388 | 1.415 |

We can see the precisions are ordered and the variability within each state increases as the unobserved level of pollution increases. Moreover the $O_3$ hourly dynamics of the series, described by the $\beta_i$'s, presents peaks which become more evident as the unobserved level of pollution increases (Figure 3a).

The dynamics of the hidden states can be observed in Figure 3b, where we have the sequence of the posterior modes of any generated state $x_d$, for any $d$=1,…, $D$: state 1 underlies the observations with the lowest level of pollution, while state 3 underlies those with the highest level of pollution.

The probability of persistence in states 1 and 2 is greater than the probability of transition; by contrast the probability of persistence in state 3 is less than the probability of transition, that is the mean time of persistence in the dangerous state is the lowest one.




(a)

solid line-state one; dashes-state two; dots-state three

(b)

(c)

(d)

solid line-fitted values; dashes-actual values

(e)

(f)

solid line-fitted values; dashes-actual values

*Figure 3* – The daily component from 5 a.m. to 5 a.m. (a); the sequence of the hidden states (b); the sequence of the residuals (c); actual and fitted values of the sub-series of days 8 - 12 (d), 98 - 102 (e), 172 - 176 (f).

Within the sequence of observations, we have 549 missing values which can be grouped in three sets: 37 missing observations gathered in 21 small blocks of 4 data at maximum, 311 missing observations gathered in 8 medium blocks whose number of elements is between 13 and 75 and 1 huge block of 201 observations. Missing observations are simulated according to (5): we can see from Figures 4a and 4b these simulated values correctly fill the series according to the dynamics of the twenty-four hours.

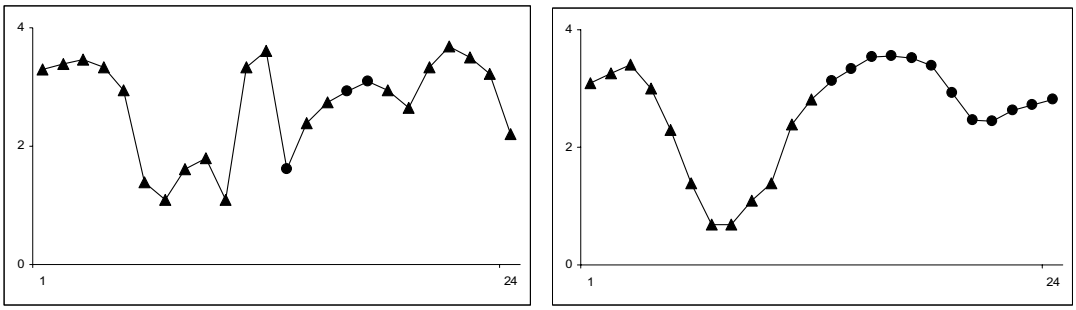



*Figure 4* – Actual (triangles) and fitted (circles) values of days 7 (a) and 161 (b).

In order to assess the fitting accuracy of SMSAR(*3;2*), the fitted and actual values are analysed through two descriptive statistics: the root mean squared error (RMSE) and the mean absolute error (MAE). We obtain RMSE=0.406 and MAE=0.259: by these values we can argue that the fitting ability of the model is satisfactory. By comparing three subseries of five days (Figures 3d - 3f) we can see the dynamics of the fitted values respects the dynamics of the real data.

Another interesting result of our analysis is the comparison between observed and fitted values with regards to the concentrations of $O_3$ exceeding the attention-level (180μg/m$^3$). In the observed series there are 37 values greater than the attention-level, splitted in 10 days, and SMSAR (*3;2*) captures 5 day out of 10 in which the threshold is exceeded.

Residual analysis has been graphically performed through histograms and QQ plots (Figure 5), showing that for any state the normality of the residuals is respected.

### 4.3. *Forecasting*

The model we fitted by the series of $O_3$ concentrations recorded from April 1st, 1998, 1 a.m., to September 30th, 1998, 12 p.m., is now used to obtain the one-day ahead hidden state predictions and the $k$-hour ahead pollutant predictions ($1 \le k \le 24$), for the data from April 1st, 1999, 1 a.m., to September 30th, 1999, 12 p.m.; this problem is tackled considering future values as missing.

Let $d^*$ be the generic future day ($d^* > D$); we have to forecast the future state $x_{d^*}$, given the observations up to day $d^*-1$ ($\mathbf{y}^{24(d^*-1)}$), and the future observations $y_{(d^*-1)24+1},\ldots,y_{24d^*}$, given the future state and the observations up to $k$ times before, assuming parameter $\vartheta$ as known because it has just been estimated.

Gibbs sampling procedure for $k$-hour ahead forecasting is the following.

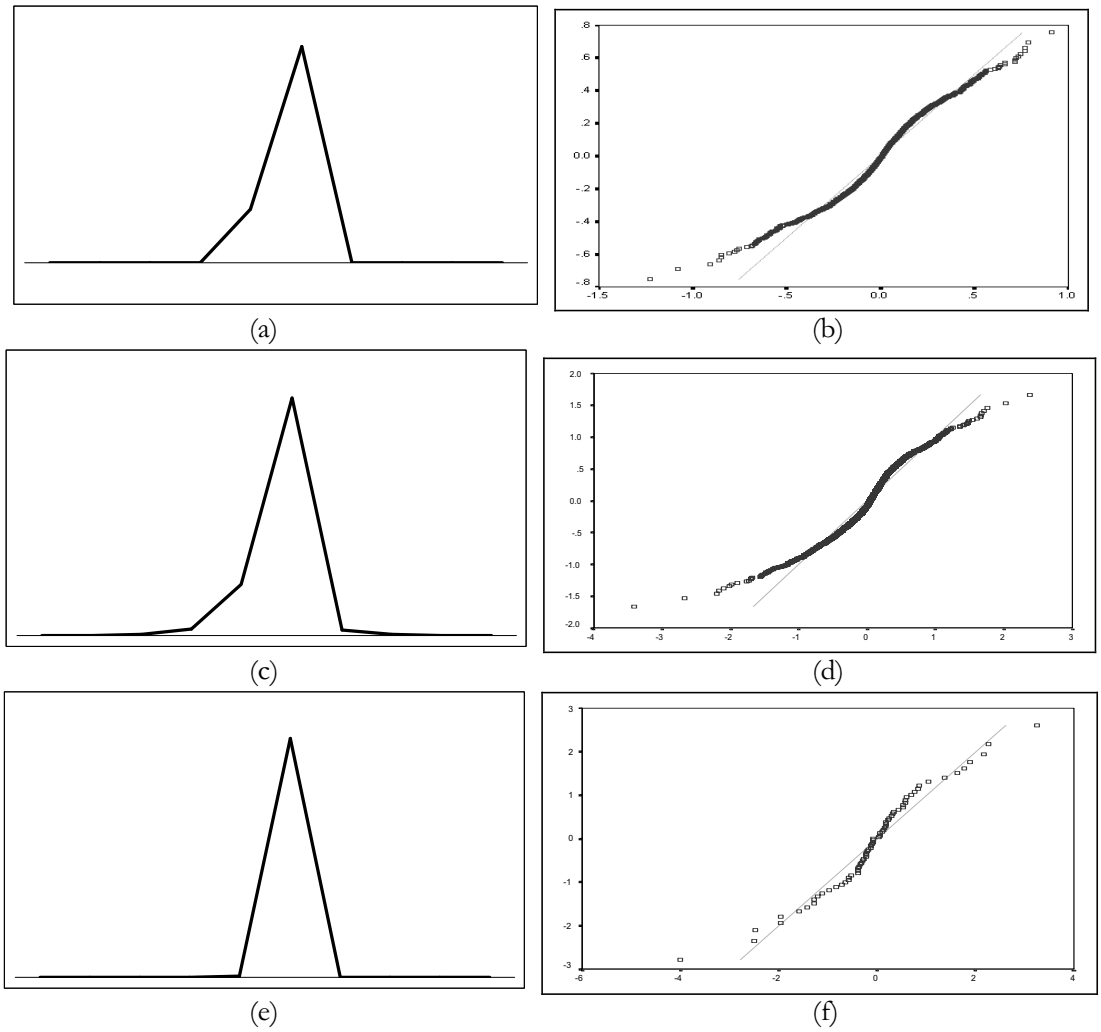



*Figure 5* – Continuous approximations of the histograms and QQ plots of the residuals, given state one (a; b), two (c; d) and three (e; f).

1) Generate state $x_{d^*}$ through *ff-bs* algorithm by means of the observations $\mathbf{y}^{24(d^*-1)}$, considering the observations $y_{(d^*-1)24+1},\ldots,y_{24d^*}$ as missing values, that is their corresponding *pdf*s are replaced with 1.

2) For any $h$=1,…,24, generate the missing observations $y_{(d^*-1)24+h},\ldots,y_{(d^*-1)24+h+k-1}$, given the observations $\mathbf{y}^{(d^*-1)24}$, the previously forecasted values $y_{(d^*-1)24+1},\ldots,$ $y_{(d^*-1)24+h-1}$ and the hidden state contemporary to any prediction; then collect $y_{(d^*-1)24+h+k-1}$ as the $k$-hour ahead prediction.

3) Repeat steps 1) and 2), for any $d^*>D$.

To evaluate the forecasting ability of our model we used the $O_3$ spring and summer 1999 data to compare them with all the $k$-hour ahead predictions. We also considered different values for $k$ ($k$=1,2,3,4,5,6,9,12,18,24) and, for any $k$, we compared the predictive MAE (PMAE) and the predictive RMSE (PRMSE): by Table 2 we can notice the forecasting statistics slowly increase, as $k$ increases.

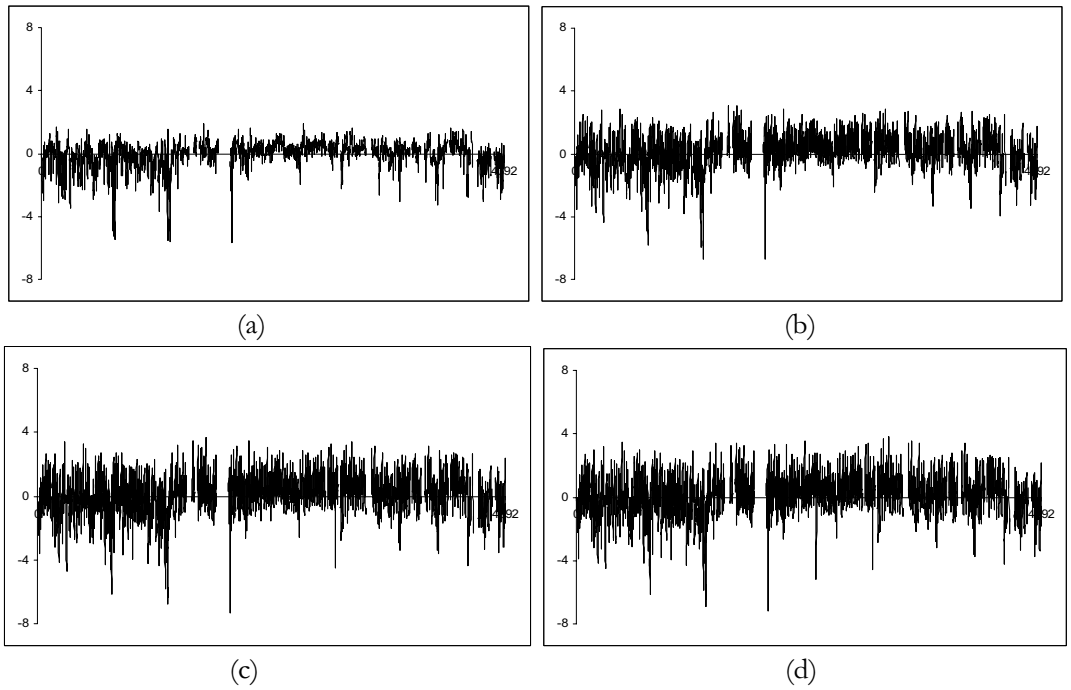



*Figure 6* – The sequence of the predictive residuals for $k$=1 (a); $k$=6 (b); $k$=12 (c); $k$=24 (d).

In Figure 6 we can see the dynamics of the predictive residuals for $k$=1;6;12;24.

Finally we are interested in the ability of the model to forecast the exceedings of the attention-level (180μg/m$^3$), given that the alarm-level has never been reached. The threshold has been exceeded by our observations nine times in five different days: one hour in days 276, 293 and 294; three hours in day 277 and 295. For any $k$ up to six, we have that the threshold is never exceeded by the predictions, by contrast when $k$ is greater than six the predictions exceed the threshold more times than real data do.

TABLE 2

*Predictive mean absolute error (PMAE) and predictive root mean square error (PRMSE) computed between the actual values and the k-hour ahead predictions*

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 9 | 12 | 18 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|
| PMAE | 0.550 | 0.571 | 0.617 | 0.666 | 0.713 | 0.754 | 0.838 | 0.864 | 0.881 | 0.883 |
| PRMSE | 0.798 | 0.822 | 0.872 | 0.932 | 0.987 | 1.036 | 1.133 | 1.171 | 1.194 | 1.198 |

## 5. CONCLUSIONS

We recurred to Bayesian seasonal Markov switching autoregressive models to analyse and predict a time series about the mean concentrations of ozone, whose dynamics is characterized by seasonality, non-normality and non-linearity. Model choice, inference and forecasting have been performed through Metropolis-within-Gibbs algorithm, considering the label switching problem, which has been efficiently tackled by permutation sampling.

The models we considered can be extended in many ways (i.e. time-varying transition matrices, multivariate pollutants and multisite recording analysis) to apply them more extensively to air quality control; these extensions concern the authors' current researches.

ACKNOWLEDGEMENTS

The authors are thankful to the referee for his constructive comments. The second author is thankful to Mary MacBalcon for helpful discussions on model choice. The research of the second author has been supported by "Progetto di Eccellenza 21302BAIPSE *La ricerca economica e sociale a supporto delle politiche locali di sviluppo*, Università degli Studi del Piemonte Orientale". The data set has been provided by the Bergamo ARPA.

*Istituto di Statistica* ROBERTA PAROLI
*Università Cattolica del S. Cuore di Milano*

*Dipartimento di Scienze Economiche e Metodi Quantitativi* LUIGI SPEZIA
*Università degli Studi del Piemonte Orientale, Novara*

REFERENCES

O. BARNDORFF-NIELSEN, G. SCHOU, (1973), *On the parametrization of autoregressive models by partial autocorrelations*, "Journal of Multivariate Analysis", 3, 408-419.

M. BILLIO, A. MONFORT, C.P. ROBERT, (1999), *Bayesian estimation of switching ARMA models*, "Journal of Econometrics", 93, 229-255.

C.K. CARTER, R. KOHN, (1994), *On Gibbs sampling for state space models*, "Biometrika", 81, 541-553.

G. CELEUX, M. HURN, C.P. ROBERT, (2000), *Computational and inferential difficulties with mixture posterior distributions*, "Journal of the American Statistical Association", 95, 957-970.

S. CHIB, (1995), *Marginal likelihood from the gibbs output*, "Journal of the American Statistical Association", 90, 1313-1321.

S. CHIB, (1996), *Calculating posterior distributions and modal estimates in Markov mixture models*, "Journal of Econometrics", 75, 79-97.

S. CHIB, I. JELIAZKOV, (2001), *Marginal likelihoods from the Metropolis-Hastings output*, "Journal of the American Statistical Association", 96, 270-281.

P.H. FRANSES, D. VAN DIJK, (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.

S. FRÜHWIRTH-SCHNATTER, (1994), *Data augmentation and dynamic linear models*, "Journal of Time Series Analysis", 15, 183-202.

S. FRÜHWIRTH-SCHNATTER, (1999), *Model likelihoods and bayes factors for switching and mixture models*, Technical Report 1999-71, Department of Statistics, Vienna University of Economics and Business Administration.

S. FRÜHWIRTH-SCHNATTER, (2001), *Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models*, "Journal of the American Statistical Association", 96, 194-209.

D. GAMERMAN, (1997), *Markov chain Monte Carlo: stochastic simulation for bayesian inference*, Chapman & Hall, London.

J.D. HAMILTON, (1989), *A new approach to the economic analysis of nonstationary time series and the business cycle*, "Econometrica", 57, 357-384.

J.D. HAMILTON, (1990), *Analysis of time series subject to changes in regime*, "Journal of Econometrics", 45, 39-70.

J.D. HAMILTON, (1993), *Estimation, inference and forecasting of time series subject to changes in regime*, In "Handbook of Statistics", vol. 11, eds G. S. Maddala, C. R. Rao, H. D. Vinod, 231-259. North-Holland, Amsterdam.

J.D. HAMILTON, (1994), *Time series analysis*, Princeton University Press, Princeton.

M.C. JONES, (1987), *Randomly choosing parameters from the stationarity and invertible region of autoregressive-moving average models*, "Applied Statistics", 36, 134-138.

R.E. KASS, A.E. RAFTERY, (1995), *Bayes factors*, "Journal of the American Statistical Association", 90, 773-795.

C.J. KIM, C.R. NELSON, (1999), *State-space models with regime switching: classical and gibbs-sampling approaches with applications*, Cambridge: The MIT Press.

H.-M. KROZLIG, (1997), *Markov-switching vector autoregressions: modelling, statistical inference and applications to business cycle analysis*, Springer, Berlin.

R.E. McCULLOCH, R.S. TSAY, (1994), *Statistical analysis of economic time series via markov switching models*, "Journal of Time Series Analysis", 81, 27-40.

S. RICHARDSON, P.J. GREEN, (1997), *On bayesian analysis of mixtures with unknown number of components (with discussion)*, "Journal of the Royal Statistical Society", *Series B*, 59, 731-792.

M. STEPHENS, (2000), *Dealing with label switching in mixture models*, "Journal of the Royal Statistical Society", *Series B*, 62, 795-809.

## RIASSUNTO

*Autoregressioni stagionali con cambiamenti di regime*

I modelli autoregressivi con cambiamenti di regime costituiscono un valido strumento nell'analisi di serie temporali non lineari e non normali. In questo lavoro un particolare modello autoregressivo, con cambiamenti di regime di natura markoviana e con una componente stagionale dipendente dallo stato della catena di Markov latente, viene studiato per l'analisi di serie temporali periodiche. La selezione dei vincoli di identificabilità, la scelta del modello migliore, la stima dei parametri e dei dati latenti e la previsione dei valori futuri sono affrontati utilizzando algoritmi di tipo Metropolis-within-Gibbs, che tengono conto anche dei differenti ordinamenti possibili degli stati latenti. La metodologia proposta è illustrata con l'analisi della dinamica di un inquinante dell'aria.

## SUMMARY

*Seasonal autoregressions with regime switching*

Markov switching autoregressive models (MSARMs) are efficient tools to analyse nonlinear and non-normal time series. A special MSARM with a hidden state-dependent seasonal component is proposed here to analyse periodic time series. We present a complete Metropolis-within-Gibbs algorithm for constraint identification, for model choice and for the estimation of the unknown parameters and the latent data. These three consecutive steps are developed tackling the problem of the hidden states labeling, by means of random permutation sampling and constrained permutation sampling. The missing observations occurring within the observed series and the future values are respectively estimated and forecasted considering them as unknown parameters. We illustrate our methodology with an example about the dynamics of an air pollutant.