Consistency of the maximum likelihood estimate for Non-homogeneous Markov-switching models

Pierre Ailliot
Françoise Pène
Laboratoire de Mathématiques, UMR 6205, Université de Brest, France

June 7, 2013

Abstract

Many nonlinear time series models have been proposed in the last decades. Among them, the models with regime switchings provide a class of versatile and interpretable models which have received a particular attention in the literature. In this paper, we consider a large family of such models which generalize the well known Markov-switching AutoRegressive (MS-AR) by allowing non-homogeneous switching and encompass Threshold AutoRegressive (TAR) models. We prove various theoretical results related to the stability of these models and the asymptotic properties of the Maximum Likelihood Estimates (MLE). The ability of the model to catch complex nonlinearities is then illustrated on various time series.

Keywords: Markov-switching autoregressive process, non-homogeneous hidden Markov process, maximum likelihood, consistency, stability, lynx data, wind direction

Introduction

Recent decades have seen extensive interest in time series models with regime switchings. One of the most influential paper in this field is the one by Hamilton in 1989 (see [**?**]) where Markov-Switching AutoRegressive (MS-AR) models were introduced. It became one of the most popular nonlinear time series model. MS-AR models combine several autoregressive models to describe the evolution of the observed process $\{Y_k\}$ at different periods of time, the transition between these autoregressive models being controlled by a hidden Markov chain $\{X_k\}$. In most applications, it is assumed that $\{X_k\}$ is an homogeneous Markov chain. In this work, we relax this assumption and let the evolution of $\{X_k\}$ depend on lagged values of $\{Y_k\}$ and exogenous covariates.

More formally, we assume that X_k takes its values in a compact metric space E endowed with a finite Borel measure m_E and that Y_k takes its values in a complete separable metric space K endowed with a non-negative Borel σ-finite measure m_K and we set $\mu_0 := m_E \times m_K$. It will be useful to denote $Y_k^{k+\ell} := (Y_k, ..., Y_{k+\ell})$, $y_k^{k+\ell} := (y_k, ..., y_{k+\ell})$ (and to use analogous notations $X_k^{k+\ell}$, $x_k^{k+\ell}$) for integer k and $\ell \geq 0$. The Non-Homogeneous Markov-Switching AutoRegressive (NHMS-AR) model of order $s > 0$ considered in this work is characterized by Hypothesis 1 below.

Hypothesis 1. *The sequence $\{X_k, Y_k\}_k$ is a Markov process of order s with values in $E \times K$ such that, for some parameter θ belonging to some subset Θ of $\mathbb{R}^d$,*

- *the conditional distribution of X_k (wrt m_E) given the values of $\{X_{k'} = x_{k'}\}_{k'<k}$ and $\{Y_{k'} = y_{k'}\}_{k'<k}$ only depends on x_{k-1} and y_{k-s}^{k-1} and this conditional distribution has a probability density function (pdf) denoted*
$$p_{1,\theta}(x_k|x_{k-1}, y_{k-s}^{k-1})$$
with respect to m_E.

- *the conditional distribution of Y_k given the values of $\{Y_{k'} = y_{k'}\}_{k'<k}$ and $\{X_{k'} = x_{k'}\}_{k' \leq k}$ only depends on x_k and y_{k-s}^{k-1} and this conditional distribution has a pdf*

  $$p_{2,\theta}\left(y_k | x_k, y_{k-s}^{k-1}\right)$$

  with respect to m_K.

Let us write $q_\theta(\cdot | x_{k-1}, y_{k-s}^{k-1})$ for the conditional pdf (with respect to μ_0) of (X_k, Y_k) given $(X_{k-1} = x_{k-1}, Y_{k-s}^{k-1} = y_{k-s}^{k-1})$. Hypothesis 1 implies that

$$q_\theta(x, y | x_{k-1}, y_{k-s}^{k-1}) = p_{1,\theta}(x | x_{k-1}, y_{k-s}^{k-1}) p_{2,\theta}(y | x, y_{k-s}^{k-1}).$$

The various conditional independence assumptions of Hypothesis 1 are summarized by the directed acyclic graph (DAG) below when $s = 1$.

$$\begin{array}{llccccccccc}
\textbf{Hidden Regime} & \cdots & \rightarrow & X_{k-1} & \rightarrow & X_k & \rightarrow & X_{k+1} & \rightarrow & \cdots \\
 & & & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & & \\
\textbf{Observed time series} & \cdots & \rightarrow & Y_{k-1} & \rightarrow & Y_k & \rightarrow & Y_{k+1} & \rightarrow & \cdots
\end{array}$$

This defines a general family of model which encompasses the most usual models with regime switchings.

- When $p_{1,\theta}(x_k | x_{k-1}, y_{k-s}^{k-1})$ does not dependent on y_{k-s}^{k-1}, the evolution of the hidden Markov chain $\{X_k\}$ is homogeneous and independent of the observed process and we retrieve the usual MS-AR models. If we further assume that $p_{2,\theta}\left(y_k | x_k, y_{k-s}^{k-1}\right)$ does not depend of y_{k-s}^{k-1}, we obtain the Hidden Markov Models (HMMs).
- When $p_{1,\theta}(x_k | x_{k-1}, y_{k-s}^{k-1})$ does not dependent on x_{k-1} and is parametrized using indicator functions, we obtain the Threshold AutoRegressive (TAR) models which is an other important family of models with regime switching in the literature (see e.g. [**?**]).

HMMs, MS-AR and TAR models have been used in many fields of applications and their theoretical properties have been extensively studied (see e.g. [**?**], [**?**] and [**?**]).

Models with non-homogeneous Markov switchings have also been considered in the literature. In particular, they have been used to describe breaks associated with events such as financial crises or abrupt changes in government policy in econometric time series (see [**?**] and references therein). They are also popular for meteorological applications (see e.g. [**?**], [**?**], [**?**]) with the regimes describing the so-called "weather types". In most cases it is assumed that the evolution of $\{X_k\}$ depends not only on lagged values of the process of interest but also on strictly exogenous variables. In order to handle such situation, we will denote $Y_k = (Z_k, R_k)$ with $\{Z_k\}$ the time series of covariates and $\{R_k\}$ the output time series to be modeled. Besides Hypothesis 1, various supplementary conditional independence assumptions can be made for specific applications. For example, in [**?**] it is assumed that the switching probabilities of $\{X_k\}$ only depend on the exogenous covariates

$$p_{1,\theta}(x_k | x_{k-1}, r_{k-s}^{k-1}, z_{k-s}^{k-1}) = p_{1,\theta}(x_k | x_{k-1}, z_{k-1})$$

that the evolution of $\{Z_k\}$ is independent of $\{X_k\}$ and $\{R_k\}$ and that R_k is conditionally independent of Z_{k-s}^k and R_{k-s}^{k-1} given X_k

$$p_{2,\theta}\left(z_k, r_k | x_k, z_{k-s}^{k-1}, r_{k-s}^{k-1}\right) = p_{R,\theta}\left(r_k | x_k\right) p_Z\left(z_k | z_{k-s}^{k-1}\right).$$

This model, which dependence structure is summarized by the DAG below when $s = 1$ is often referred as Non-Homogeneous Hidden Markov Models (NHMMs) in the literature.

$$\begin{array}{llccccccccc}
\textbf{Covariates} & \cdots & \rightarrow & Z_{k-1} & \rightarrow & Z_k & \rightarrow & Z_{k+1} & \rightarrow & \cdots \\
 & & & & \searrow & & \searrow & & \searrow & \\
\textbf{Hidden Regime} & \cdots & \rightarrow & X_{k-1} & \rightarrow & X_k & \rightarrow & X_{k+1} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\textbf{Output time series} & \cdots & & R_{k-1} & & R_k & & R_{k+1} & & \cdots
\end{array}$$

The most usual method to fit such models consists in computing the Maximum Likelihood Estimates (MLE). It is indeed relatively straightforward to adapt the standard numerical estimation procedure which are available for the homogeneous models, such as the forward-backward recursions or the EM algorithm, to the non-homogeneous models (see e.g. [?], [?], [?]). However, we could not find any theoretical results on the asymptotic properties of the MLE for these models and this paper aims at filling this gap.

The paper is organized as follows. In Section 1, we give general conditions ensuring the consistency of the MLE. They include conditions on the ergodicity of the model and the identifiability of the parameters. This is further discussed in Section 2 where we show that our general conditions apply to various specific but representative NHMS-AR models. In Section 3, we discuss the results obtained with the model on several time series in order to illustrate that NHMS-AR provide a wide class of flexible and interpretable models which is able to reproduce complex features of real data sets. The proofs of some results are given in the appendices.

1 A general consistency result of MLE for NHMS-AR models

We aim at estimating the true parameter $\theta^* \in \Theta$ of a NHMS-AR process $(X_k, Y_k)_k$ for which only the component $\{Y_k\}$ is observed. For that we consider the Maximum Likelihood Estimator (MLE) $\hat{\theta}_{n,x_0}$ which is defined as the maximizer of $\theta \mapsto \ell_n(\theta, x_0)$ for a fixed $x_0 \in E$ with

$$\ell_n(\theta, x_0) = \log p_\theta(Y_1^n | X_0 = x_0, Y_{-s+1}^0) = \sum_{k=1}^n \log \frac{p_\theta(Y_1^k | X_0 = x_0, Y_{-s+1}^0)}{p_\theta(Y_1^{k-1} | X_0 = x_0, Y_{-s+1}^0)},$$

where $p_\theta(Y_1^k | X_0 = x_0, Y_{-s+1}^0)$ is the conditional pdf of Y_1^k given $(X_0 = x_0, Y_{-s+1}^0)$ evaluated at Y_1^k, i.e.

$$p_\theta(Y_1^k | X_0 = x_0, Y_{-s+1}^0) := \int_{E^k} \prod_{\ell=1}^k q_\theta(x_\ell, Y_\ell | x_{\ell-1}, Y_{\ell-s}^{\ell-1})\, dm_E^{\otimes k}(x_1^k).$$

Before stating our main result, let us precise some notations. As usual, we define the associated transition operator Q_θ as an operator acting on the set of bounded measurable functions of $E \times K^s$ (it may also act on other Banach spaces $\mathcal{B}$) by

$$\begin{aligned} Q_\theta g(x_0, y_{-s+1}^0) &= \mathbb{E}_\theta[g(X_1, Y_{-s+2}^1) | X_0 = x_0, Y_{-s+1}^0 = y_{-s+1}^0] \\ &= \int_{E\times K} g(x_1, y_{-s+2}^1) q_\theta(x_1, y_1 | x_0, y_{-s+1}^0)\, d\mu_0(x_1, y_1). \end{aligned}$$

We denote by Q_θ^* the adjoint operator of Q_θ defined on $\mathcal{B}'$ the dual space of $\mathcal{B}$ (if Q_θ acts on $\mathcal{B}$) by

$$\forall \nu \in \mathcal{B}',\ \forall f \in \mathcal{B},\quad Q_\theta^*(\nu)(f) = \nu(Q_\theta(f)).$$

For every integer $k \geq 0$, the measure $(Q_\theta^*)^k(\nu)$ corresponds to the distribution of (X_k, Y_{k-s+1}^k) if $\{X_l, Y_l\}_l$ is the Markov chain with transition operator Q_θ such that the distribution of (X_0, Y_{-s+1}^0) is ν.

If $\nu \in \mathcal{B}'$ has a pdf h with respect to $\mu := m_E \times m_K^{\otimes s}$, then $Q_\theta^* \nu$ is also absolutely continuous with respect to μ and its pdf, written $Q_\theta^* h$, is given by

$$Q_\theta^* h(x_0, y_{-s+1}^0) := \int_{E\times K} q_\theta(x_0, y_0 | x_{-1}, y_{-s}^{-1}) h(x_{-1}, y_{-s}^{-1})\, d\mu_0(x_{-1}, y_{-s}).$$

Observe that, due to the particular form of q_θ, for every integer $k \geq s$ and every $P = (x_{-k}, y_{-k-s+1}^{-k}) \in E \times K^s$, the measure $(Q_\theta^*)^k \delta_P$ (where δ_P is the Dirac measure at P) is absolutely continuous with respect to $\mu := m_E \times m_K^{\otimes s}$; its pdf $Q_\theta^{*k}(\cdot|P)$ is given by

$$Q_\theta^{*k}(x_0, y_{-s+1}^0 | P) = \int_{E^{k-1}\times K^{k-s}} \prod_{i=1-k}^0 q_\theta(x_i, y_i | x_{i-1}, y_{i-s}^{i-1})\, dm_E^{\otimes(k-1)}(x_{-k+1}^{-1}) dm_K^{\otimes(k-s)}(y_{-k+1}^{-s}).$$

More generally, for every initial measure ν and every $k \geq s$, $Q_\theta^{*k}\nu$ is absolutely continuous with respect to μ and its pdf $[Q_\theta^{*k}\nu]$ is given by

$$[Q_\theta^{*k}\nu](\cdot) = \int_{E\times K^s} Q_\theta^{*k}(\cdot|P)\, d\nu(P). \tag{1}$$

We suppose that, for every $\theta \in \Theta$, there exists an invariant probability measure $\bar{\nu}_\theta$ for Q_θ^*. Observe that, due to (1), $\bar{\nu}_\theta$ admits a pdf h_θ with respect to μ.

We identify $(X_k, Y_k)_k$ with the canonical Markov chain $\{(X_0, Y_0) \circ \tau^k\}_k$ defined on $\Omega_+ := (E \times K)^{\mathbb{N}}$ by $X_0((x_k, y_k)_k) = x_0$, $Y_0((x_k, y_k)_k) = y_0$, τ_+ being the shift $(\tau_+((x_k, y_k)_k) = (x_{k+1}, y_{k+1})_k)$. We endow Ω_+ with its Borel σ-algebra $\mathcal{F}_+$. We denote by $\bar{\mathbb{P}}_\theta$ the probability measure on $(\Omega_+, \mathcal{F}_+)$ associated to the invariant measure $\bar{\nu}_\theta$ and by $\bar{\mathbb{E}}_\theta$ the corresponding expectation.

The question of consistency of the MLE has been studied by many authors in the context of usual HMMs (see e.g. [**?**, **?**, **?**]) and MS-AR models (see [**?**] and references therein). The aim of this section is to state consistency results of MLE for general NHMS-AR. The proof of the following theorem is a direct but careful adaptation of the proof of [**?**, Thm. 1 & 5]. This proof is given in appendix A.

Theorem 2. *Assume that Θ is compact, that $(\Omega, \mathcal{F}, \bar{\mathbb{P}}_{\theta^*}, \tau)$ is ergodic, that there exists an invariant probability measure for every $\theta \in \Theta$, that $\bar{\mathbb{P}}_{\theta^*}$ is absolutely continuous with respect to $\bar{\mathbb{P}}_\theta$ for every $\theta \in \Theta$, that p_1 and p_2 are continuous in θ. Assume also that the following conditions hold true*

$$0 < p_{1,-} := \inf_{\theta, x_1, x_0, y_0} p_{1,\theta}(x_1|x_0, y_0) \leq p_{1,+} := \sup_{\theta, x_1, x_0, y_0} p_{1,\theta}(x_1|x_0, y_0) < \infty, \tag{2}$$

$$B_- := \bar{\mathbb{E}}_{\theta^*}\left[\left|\log\left(\inf_\theta \int_E p_{2,\theta}(Y_0|x_0, Y_{-s}^{-1})\, dm_E(x_0)\right)\right|\right] < \infty, \tag{3}$$

$$B_+ := \bar{\mathbb{E}}_{\theta^*}\left[\left|\log\left(\sup_\theta \int_E p_{2,\theta}(Y_0|x_0, Y_{-s}^{-1})\, dm_E(x_0)\right)\right|\right] < \infty, \tag{4}$$

$$\forall \theta \in \Theta, \quad \sup_{y_{-s}^{-1}} \int_E p_{2,\theta}(Y_0|x, y_{-s}^{-1})\, dm_E(x) < \infty, \quad \bar{\mathbb{P}}_{\theta^*} - a.s, \tag{5}$$

$$\forall \theta \in \Theta, \quad for\ \mu - a.e.\ P \in E \times K^s, \lim_{k\to+\infty} ||Q_\theta^{*k}(P'|P) - h_\theta(P')||_{L^1(\mu)} = 0. \tag{6}$$

Then, for every $x_0 \in E$, the limit values of $(\hat{\theta}_{n,x_0})_n$ are $\bar{\mathbb{P}}_{\theta^*}$-almost surely contained in $\{\theta \in \Theta \ :\ \bar{\mathbb{P}}_\theta^Y = \bar{\mathbb{P}}_{\theta^*}^Y\}$.

If, moreover, Q_{θ^*} is positive Harris recurrent and aperiodic, then, for every $x_0 \in E$ and every initial probability ν, the limit values of $(\hat{\theta}_{n,x_0})_n$ are almost surely contained in $\{\theta \in \Theta \ :\ \bar{\mathbb{P}}_\theta^Y = \bar{\mathbb{P}}_{\theta^*}^Y\}$.

Our hypotheses are very close to those of [**?**]. Let us point out the main differences. First, in [**?**] $p_{1,\theta}(x|x', y')$ does not depend on y'. Second, (4) and (5) are slightly weaker than

$$\sup_{\theta, y_{-s}^{-1}, y_0, x} p_{2,\theta}(y_0|x, y_{-s}^{-1}) < \infty$$

assumed in [**?**]. This is illustrated below in Section 2.3 where the parametrization of p_2 uses Gamma pdf which may not be bounded close to the origin depending on the values of the parameters. The results given in [**?**] do not apply directly to this model whereas we will show that (4) applies. Third, to prove the result in the stationary case, we replace Harris recurrence by (6) which is equivalent to each one of the two following properties

- for any initial measure ν on $E \times K^s$, we have $\lim_{n\to+\infty} \|Q_\theta^{*n}\nu - \nu_\theta\|_{TV} = 0$, where $\|\cdot\|_{TV}$ stands for the total variation norm,
- for any initial measure ν on $E \times K^s$, we have $\lim_{n\to+\infty} \sup_{\nu\in\mathcal{P}(E\times K)} ||[Q_\theta^{*n}\nu] - h_\theta||_{L^1(m_E\times m_K^s)} = 0$, with $\mathcal{P}(E \times K)$ the set of probability measures on $E \times K$.

Remark 3. *Observe that, if* $q_\theta > 0$ *and if* ν_θ *exists for every* $\theta \in \Theta$*, then the pdf* h_θ *of* ν_θ *satisfies* $h_\theta > 0$ *(μ-a.e.). In this case,* $\bar{\mathbb{P}}_{\theta^*}$ *is absolutely continuous with respect to* $\bar{\mathbb{P}}_\theta$ *for every* $\theta \in \Theta$.

Observe also that the ergodicity of the dynamical system $(\Omega, \mathcal{F}, \bar{\mathbb{P}}_{\theta^*}, \tau)$ *is satisfied as soon as the transition operator is strongly ergodic with respect some Banach space* $\mathcal{B}$ *satisfying general assumptions (see for example [?, Proposition 2.2]).*

2 Ergodicity and consistency of MLE for specific NHMS-AR models

In this section, we discuss how the general results given in the previous section apply to specific but typical NHMS-AR models. We discuss in particular the recurrent and ergodic properties of the models since it is a key step to prove the consistence of the MLE (see Theorem 2).

2.1 NHMS-AR model with gaussian linear autoregressive models

2.1.1 Model

In this section, we focus on a simple NHMS-AR model with only two regimes and linear Gaussian autoregressive models. The model has no exogenous variable but the transition kernel depends on lagged values of the observed process as in the Self Exciting Threshold AutoRegressive (SETAR) models. The model is introduced more formally below.

Hypothesis 4. *We assume that* $E = \{1, 2\}$ *(endowed with the counting measure),* $K = \mathbb{R}$ *(endowed with the Lebesgue measure) and* $\{Y_k\}$ *satisfies*

$$Y_k = \beta_0^{(x_k)} + \sum_{\ell=1}^{s} \beta_\ell^{(x_k)} Y_{k-\ell} + \sigma^{(x_k)} \epsilon_k$$

with $\{\epsilon_k\}$ *an iid sequence of standard Gaussian random variables, with* $\sigma^{(x)} > 0$ *and* $\beta_l^{(x)} \in \mathbb{R}$ *for every* $\ell \in \{0, ..., s\}$ *and every* $x \in \{1, 2\}$,

$$i.e. \qquad p_{2,\theta}(y_k | x_k, y_{k-s}^{k-1}) = \mathcal{N}\left(y_k; \beta_0^{(x_k)} + \sum_{\ell=1}^{s} \beta_\ell^{(x_k)} y_{k-\ell}, \sigma^{(x_k)} \right),$$

where $\mathcal{N}(\cdot; m, \sigma)$ *stands for the gaussian pdf with mean* m *and standard deviation* σ.

The transition probabilities of $\{X_k\}$ *are parametrized using the logistic function as follows when* $x_k = x_{k-1}$

$$p_{1,\theta}(x_k | x_{k-1}, y_{k-s}^{k-1}) = \pi_-^{(x_{k-1})} + \frac{1 - \pi_-^{(x_{k-1})} - \pi_+^{(x_{k-1})}}{1 + \exp\left(\lambda_0^{(x_{k-1})} + \lambda_1^{(x_{k-1})} y_{k-r} \right)} \tag{7}$$

with $r \leq s$ *a positive integer and the unknown parameters* $\pi_-^{(x)}, \pi_+^{(x)}, \lambda_0^{(x)}, \lambda_1^{(x)}$ *for* $x \in \{1, 2\}$.
The unknown parameter θ *corresponds to*

$$\theta = \left((\beta_i^{(x)}), (\sigma^{(x)}), (\pi_-^{(x)}), (\pi_+^{(x)}), (\lambda_i^{(x)}) \right).$$

We write $\tilde{\Theta}$ *for the set of such parameters* θ *satisfying, for every* $x \in \{1, 2\}$, $\sigma^{(x)} > 0$ *and* $0 < \pi_-^{(x)} < 1 - \pi_+^{(x)} < 1$ *(this last constraint is added in order to ensure that (2) holds).*

Although very simple, this model encompasses the homogeneous gaussian MS-AR model when $\lambda_1^{(1)} = \lambda_1^{(2)} = 0$ and the SETAR model with two regimes (SETAR(2)) as a limit case. Indeed, if $s = -\frac{\lambda_0^{(x)}}{\lambda_1^{(x)}}$ is fixed for $x \in \{1,2\}$, $\lambda_1^{(1)} \to +\infty$, $\lambda_1^{(2)} \to -\infty$, $\pi_-^{(x)} \to 0$ and $\pi_+^{(x)} \to 0$ then

$$p_1(X_k = 1|x_{k-1}, y_{k-s}^{k-1}) \to \mathbb{1}(y_{k-r} \leq s) \text{ and } p_1(X_k = 2|x_{k-1}, y_{k-s}^{k-1}) \to \mathbb{1}(y_{k-r} \geq s)$$

Both models have been extensively studied in the literature.

The model can be generalized in several ways to handle $M \geq 3$ regimes or include covariates, for example through a linear function in the logistic term (see e.g. [?]). Other link functions such as the probit model used in [?] or a Gaussian kernel (see (18)) could also be considered.

2.1.2 Properties of this Markov chain

Various authors have studied the ergodicity of MS-AR ([?], [?], [?]) and TAR ([?], [?]) models. A classical approach to prove the ergodicity of a non-linear time series consists in establishing a drift condition. Here we will use a strict drift condition. Let $\|\cdot\|$ be some norm on $\mathbb{R}^s$. For any $R > 0$, we consider the set $E_R := \{(x, y_{-s+1}^0) : \|y_{-s+1}^0\| \leq R\}$. Recall that μ is here the product of the counting measure on E and of the Lebesgue measure on $\mathbb{R}^s$.

Proposition 5. *Assume hypothesis 4.*

The Markov chain is ψ-irreducible (with $\psi = \mu$).

Let $R > 0$. The set E_R is ν_s-small and ν_{s+1}-small with ν_s and ν_{s+1} equivalent to μ. Hence, the markov chain is aperiodic.

Proof. The ψ-irreducibility comes from the positivity of q_θ.

Let us prove that E_R is ν_s-small with $\nu_s = h_s \cdot \mu$ and

$$h_s(x_s, y_1^s) = \inf_{(x_0, y_{-s+1}^0) \in E_R} \int_{E^s} \prod_{\ell=1}^{s} q_\theta(x_\ell, y_\ell | x_{\ell-1}, y_{\ell-s}^{\ell-1}) \, dx_1^{s-1} > 0.$$

Indeed $p_{1,\theta}$ is uniformly bounded from below by some $p_{1,-}$, $\sigma^{(x)}$ are uniformly bounded from above by some σ_+ and from below by some σ_- and, for every $\ell \in \{1, ..., s\}$, we have

$$\forall Z \in \mathbb{R}, \quad g_\ell(Z) := \sup_{(x_\ell, y_{-s+1}^0) \in E_R} \left| Z - \beta_0^{(x_\ell)} - \sum_{j=\ell}^{s} \beta_j^{(x_\ell)} y_{\ell-j} \right|^2 < \infty.$$

So

$$h_s(x_s, y_1^s) \geq \inf_{x_1, \ldots, x_s \in \{1,2\}} \frac{(p_{1,-})^s}{(2\pi\sigma_-)^{\frac{s}{2}}} \exp\left(-\frac{1}{2\sigma_+} \sum_{\ell=1}^{s} g_\ell \left(y_\ell - \sum_{j=1}^{\ell-1} \beta_j^{(x_\ell)} y_{\ell-j} \right) \right).$$

The proof of the ν_{s+1}-smallness of E_R (with ν_{s+1} equivalent to μ) uses the same ideas. $\square$

Now, to obtain the other properties related to the ergodicity of the process for practical applications (see Section 3.1 for an example), we can use the following strict drift property.

Hypothesis 6. *There exist three real numbers $K < 1$, $L > 0$ and $R > 0$ such that, for every $(x_0, y_{-s+1}^0) \in \{1,2\} \times \mathbb{R}^s$,*

$$\mathbb{E}[\left\|Y_{-s+2}^1\right\|^2 | Y_{-s+1}^0 = y_{-s+1}^0, X_0 = x_0] \leq K \left\|y_{-s+1}^0\right\|^2 + L\mathbb{1}_{E_R}(y_{-s+1}^0). \tag{8}$$

Recall that this property has several classical consequences (see [?, Chapters 11 and 15] for more details). Hypothesis 6 (combined with the irreducibility and aperiodicity coming from Hypothesis 4) implies in particular

- the existence of a (unique) stationary measure admitting a moment of order 2;

- the V-geometric ergodicity with $V(x, y_{-s+1}^0) = \|y_{-s+1}^0\|^2$ and so the ergodicity of the Markov chain (see for example [?, Proposition 2.2] for this last point);
- the positive Harris recurrence.

We end this section with some comments on (8). Let us write

$$\Lambda^{(x)} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ \beta_s^{(x)} & \beta_{s-1}^{(x)} & \cdots & \cdots & \cdots & \beta_1^{(x)} \end{pmatrix}$$

for the companion matrix associated to the AR model in regime x,

$$\Phi^{(x)} := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \beta_0^{(x)} \end{pmatrix}, \quad \Sigma^{(x)} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \sigma^{(x)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \varepsilon_1 \end{pmatrix}.$$

There exist $A, B > 0$ such that, for every $(x_0, y_{-s+1}^0) \in \{1, 2\} \times \mathbb{R}^s$, we have

$$\begin{aligned} \mathbb{E}[\left\|Y_{-s+2}^1\right\|^2 | Y_{-s+1}^0 = y_{-s+1}^0, X_0 = x_0] &= \sum_{x_1=1}^{M} p_{1,\theta}(x_1|x_0, y_{-s+1}^0) \mathbb{E}[\left\|\Lambda^{(x_1)} y_{-s+1}^0 + \Phi^{(x_1)} + \Sigma^{(x_1)} \boldsymbol{\varepsilon}\right\|^2] \\ &\leq \sum_{x_1=1}^{M} p_{1,\theta}(x_1|x_0, y_{-s+1}^0) \left\|\Lambda^{(x_1)}\right\|^2 \left\|y_{-s+1}^0\right\|^2 + A \left\|y_{-s+1}^0\right\| + B \end{aligned}$$

where $\|.\|$ denotes abusively the matrix norm associated to the vector norm. We deduce the following.

Remark 7. *The strict drift condition (8) is satisfied when there exists $R > 0$ such that for all $x_0 \in \{1, 2\}$ and all $y_{-s+1}^0 \in \mathbb{R}^s$*

$$\|y_{-s+1}^0\| > R \quad \Rightarrow \quad \sum_{x_1=1}^{2} p_{1,\theta}(x_1|x_0, y_{-s+1}^0) \left\|\Lambda^{(x_1)}\right\|^2 < 1. \tag{9}$$

This is true in particular if

$$\forall x \in E, \ \left\|\Lambda^{(x)}\right\| < 1. \tag{10}$$

The use of condition (10) will be illustrated on a specific example in Section 3.1. It implies that all the regimes are stable. However, it is also possible to construct models which satisfy (9) with some unstable regimes if the instability is controlled by the dynamics of $\{X_k\}$.

2.1.3 Consistency of MLE

The results given in this section generalize the results given in [?, ?] for homogeneous MS-AR models with linear Gaussian autoregressive models.

Corollary 8. *Assume that Hypotheses 4 and 6 hold true for every θ. Let Θ be a compact subset of $\tilde{\Theta}$. Then, for all $\theta \in \Theta$ there exists a unique invariant probability distribution and, for every $x_0 \in \{1, 2\}$ and every initial probability distribution ν, the limit values of $(\hat{\theta}_{n,x_0})_n$ are $\bar{\mathbb{P}}_{\theta^*}$-almost surely contained in $\{\theta \in \Theta \ : \ \bar{\mathbb{P}}_\theta = \bar{\mathbb{P}}_{\theta^*}\}$.*

Proof. This corollary is a direct consequence of Theorem 2 and of the previous section. As already noticed in section 1, the invariant measure has a positive pdf with respect to μ. As seen in the previous section, the Markov chain is aperiodic positive Harris recurrent (which implies (6)) and the stationary process is square integrable, which implies (3) and (4). In this example, $p_{2,\theta}$ is bounded from above and so (5) holds. $\square$

In the sequel, we explicit the limit set $\{\theta \in \Theta \ : \ \bar{\mathbb{P}}_\theta = \bar{\mathbb{P}}_{\theta^*}\}$ under the supplementary condition

$$\left(\beta_0^{(1)}, \beta_1^{(1)}, ..., \beta_s^{(1)}, \sigma^{(1)}\right) \neq \left(\beta_0^{(2)}, \beta_1^{(2)}, ..., \beta_s^{(2)}, \sigma^{(2)}\right) \tag{11}$$

that the dynamics in the two regimes are distinct. Note that this condition is not sufficient in order to ensure identifiability. First, it can be easily seen that the homogeneous MS-AR model can be written in many different ways using the parametrization (7). It led us to add one of the following constraints on the parameters

$$\forall x \in \{1,2\}, \lambda_1^{(x)} \neq 0 \tag{12}$$

which does not include the homogeneous model as a particular case or

$$\forall x \in \{1,2\}, \pi_-^{(x)} = \pi_+^{(x)} = \pi_0 \text{ where } 0 < \pi_0 < 1/2 \text{ is a fixed constant} \tag{13}$$

in order to solve this problem. A practical motivation for (13) is given in Section 3.1. Let Θ' be the set of $\theta \in \tilde{\Theta}$ satisfying (12) and let Θ'' be the set of $\theta \in \tilde{\Theta}$ satisfying (13). Then, a permutation of the two states also leads different parameters values but to the same model. This problem can be solved by ordering the regimes or by allowing a permutation of the states as discussed below.

Proposition 9 (Identifiability). *Let θ_1 and θ_2 belong to Θ' (resp. Θ'') with $\theta_i = \left(\theta_i^{(1)}, \theta_i^{(2)}\right)$ and*

$$\theta_i^{(x)} = \left((\beta_{j,(i)}^{(x)})_{j\in\{0,...s\}}, \sigma_i, (\lambda_{j,(i)}^{(x)})_{j\in\{0,1\}}\right)$$

the parameters associated with the regime $x \in \{1,2\}$.

Assume that θ_1 satisfies (11). Then $\bar{\mathbb{P}}_{\theta_1}^Y = \bar{\mathbb{P}}_{\theta_2}^Y$ if and only if θ_1 and θ_2 define the same model up to a permutation of indices, i.e. there exists a permutation τ of $\{1,2\}$ such that

$$\theta_1^{(x)} = \theta_2^{(\tau(x))}$$

The proof of Proposition 9 is postponed to appendix C.

Now due to Corollary 8 and Proposition 9, we directly get Theorem 10.

Theorem 10. *Assume that Hypotheses 4 and 6 hold true for every θ. Let Θ be a compact subset of Θ' or Θ''. Assume that θ^* satisfies (11). Then, for every $x_0 \in \{1,2\}$ and any initial probability distribution ν, on a set of probability one, the limit values θ of the sequence of random variables $(\hat{\theta}_{n,x_0})_n$ are equal to θ^* up to a permutation of indices.*

2.2 NHMS-AR with von Mises autoregressive models

2.2.1 Model

The motivations which led us to consider the NHMS-AR model introduced below are given in Section 3.2.

Hypothesis 11. *Let M be a positive integer. We suppose that $E = \{1, ..., M\}$ (endowed with the counting measure) and $K = \mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ (endowed with the Lebesgue measure), that $p_{1,\theta}$ and $p_{2,\theta}$ are given by*

$$p_{1,\theta}(x_k|x_{k-1}, y_{k-s}^{k-1}) = \frac{q_{x_{k-1},x_k}\left|\exp\left(\tilde{\lambda}_{x_{k-1},x_k} e^{-iy_{k-1}}\right)\right|}{\sum_{x''=1}^M q_{x_{k-1},x''}\left|\exp\left(\tilde{\lambda}_{x_{k-1},x''} e^{-iy_{k-1}}\right)\right|} \tag{14}$$

and

$$p_{2,\theta}(y_k|x_k, y_{k-s}^{k-1}) = \frac{1}{b(x_k, y_{k-s}^{k-1})} \left| \exp\left((\gamma_0^{(x_k)} + \sum_{\ell=1}^{s} \gamma_\ell^{(x_k)} e^{iy_{k-\ell}}) e^{-iy_k} \right) \right|.$$

with respect to the Lebesgue measure m on $\mathbb{T}$ and with

$$b(x_k, y_{k-s}^{k-1}) := \frac{1}{2\pi} \int_{\mathbb{T}} \exp\left(\left| \gamma_0^{(x_k)} + \sum_{\ell=1}^{s} \gamma_\ell^{(x_k)} e^{iy_{k-\ell}} \right| \cos(y) \right) dy.$$

The parameter θ belongs to the set $\tilde{\Theta}$ of $\theta = (\gamma, Q, \tilde{\lambda})$ with

$$\gamma := (\gamma_0^{(x)}, ..., \gamma_s^{(x)})_{x \in \{1,...,M\}}, \quad Q := (q_{x,x'})_{x,x' \in \{1,...,M\}}, \quad \tilde{\lambda} = (\tilde{\lambda}_{x,x'})_{x,x' \in \{1,...,M\}},$$

such that for every $x, x' \in E$, $\gamma_j^{(x)} \in \mathbb{C}$, $q_{x,x'} > 0$, $\sum_{x''} q_{x,x''} = 1$, $\tilde{\lambda}_{x,x'} \in \mathbb{C}$.

2.2.2 Properties of this Markov chain

Assume Hypothesis 11 holds. This model defines an ergodic process for any parameters values. Since, for every $(\theta, x, y) \in \tilde{\Theta} \times E \times K$, $q_\theta(x, y|\cdot, \cdot)$ is continuous on the compact set $E \times K^s$, we have

$$\alpha = \int_{E \times K} \gamma(x, y)\, d\mu_0(x, y) > 0, \quad \text{with} \quad \gamma(x, y) := \inf_{x', y_{-s}^{-1}} q_\theta(x, y|x', y_{-s}^{-1}).$$

Now we consider the pdf (w.r.t. μ_0) β given by

$$\beta(x, y) := \frac{\gamma(x, y)}{\alpha}.$$

For every $x_0, x_{-1} \in E$ and every $y_{-s}^0 \in K^{s+1}$, we have

$$q_\theta(x_0, y_0|x_{-1}, y_{-s}^{-1}) \geq \alpha\beta(x_0, y_0).$$

Due to classical results [**?**], this implies the ψ-irreducibility, the strong aperiodicity (the whole space is ν_s-small and ν_{s+1}-small, with ν_s and ν_{s+1} equivalent to μ), the Harris recurrence (since we can decompose the whole set in a union of uniformly accessible sets from the whole set), positive (the invariant measure being unique and finite).

2.2.3 Consistency of MLE

The aperiodicity and positive recurrence imply (6). The positivity of q_θ implies that the invariant distribution is equivalent to μ. Since $p_{1,\theta}(x_1|x_0, y_0)$ and $p_{2,\theta}(y_0|x_0, y_{-1})$ are continuous in (θ, x_1, x_0, y_0) and in $(\theta, x_0, y_0, y_{-1})$ (respectively), assumptions ((2), (3), (4) and (5)) of Theorem 2 are satisfied for any compact subset of $\tilde{\Theta}$. Hence, due to Theorem 2, we have the following corollary.

Corollary 12. *Assume that Hypothesis 11 holds true. Assume that Θ is a compact subset of $\tilde{\Theta}$. Then, for all $\theta \in \Theta$, there exists a unique invariant probability and, for every $x_0 \in E$ and every initial probability distribution ν, the limit values of $(\hat{\theta}_{n,x_0})_n$ are $\bar{\mathbb{P}}_{\theta^*}$-almost surely contained in $\{\theta \in \Theta \ : \ \bar{\mathbb{P}}_\theta = \bar{\mathbb{P}}_{\theta^*}\}$.*

Observe that the replacement of $(\tilde{\lambda}_{x,x'})_{x,x'}$ with $(\tilde{\lambda}_{x,x'} - a_{\theta,x})_{x,x'}$ (for some $(a_{\theta,x})_x$) does not change $p_{1,\theta}$. Therefore, to ensure parameter identifiability, we assume that $\theta = (\gamma, Q, \tilde{\lambda})$ satisfies (with the notations of Hypothesis 11) one of the following assumptions

$$\forall x \in E, \quad \tilde{\lambda}_{x,x} = 0 \tag{15}$$

or

$$\forall x \in E, \quad \sum_{x' \in E} \tilde{\lambda}_{x,x'} = 0. \tag{16}$$

Let Θ' be the set of $\theta \in \tilde{\Theta}$ satisfying (15) and let Θ'' be the set of $\theta \in \tilde{\Theta}$ satisfying (16).

The proposition below states that these conditions ensure the identifiability of the model "up to a permutation of indices" if the parameters are distinct in the different regimes.

Proposition 13 (Identifiability)**.** *Let θ_1 and θ_2 belong to Θ' (resp. Θ'') with*

$$\theta_i = \left((\gamma_{j,(i)}^{(x)})_{j,x}, (q_{x,x',(i)})_{x,x'}, (\tilde{\lambda}_{x,x',(i)})_{x,x'} \right).$$

Assume that

$$x \neq x' \;\Rightarrow\; (\gamma_{0,(1)}^{(x)}, ..., \gamma_{s,(1)}^{(x)}) \neq (\gamma_{0,(1)}^{(x')}, ..., \gamma_{s,(1)}^{(x')}). \tag{17}$$

Then $\bar{\mathbb{P}}_{\theta_1}^Y = \bar{\mathbb{P}}_{\theta_2}^Y$ if and only θ_1 and θ_2 are equal up to a permutation of indices, i.e. there exists a permutation τ of $\{1, ..., M\}$ such that, for every $x, x' \in \{1, ..., M\}$, for every $j = 0, ..., s$, the following relations hold true

$$\gamma_{j,(1)}^{(x)} = \gamma_{j,(2)}^{(\tau(x))}, \quad q_{x,x',(1)} = q_{\tau(x),\tau(x'),(2)} \quad \textit{and} \quad \tilde{\lambda}_{x,x',(1)} = \tilde{\lambda}_{\tau(x),\tau(x'),(2)}.$$

The proof of Proposition 13 is postponed to appendix B. Now due to Corollary 12 and Proposition 13, we directly get Theorem 14.

Theorem 14. *Assume Hypothesis 11. Assume that Θ is a compact subset of Θ' or of Θ'' and that θ^* satisfies (17). Then, for every $x_0 \in \{1, ..., M\}$ and any initial probability distribution ν on $\{1, ..., M\} \times \mathbb{T}^s$, on a set of probability one, the limit values $\theta = (\gamma, Q, \tilde{\lambda})$ of the sequence of random variables $(\hat{\theta}_{n,x_0})_n$ are equal to $\theta^* = (\gamma_*, Q_*, \tilde{\lambda}_*)$ up to a permutation of indices.*

2.3 Non-homogeneous Hidden Markov Models with exogenous variables

2.3.1 Model

In this part, we consider a typical example of NHMM with finite hidden state space and strictly exogenous variables and show that the theoretical results proven in this paper apply to this model. We focus on a model initially introduced in [?] for downscaling rainfall. It is an extension of the model proposed in [?] (see also [?] for more recent references). The results given in this section can be adapted to other NHMM with finite hidden state space such as the one proposed in [?] which is widely used in econometrics.

The model is described more precisely hereafter.

Hypothesis 15. *Let M be a positive integer and Σ be a $m \times m$ positive definite symmetric matrix. We suppose that $E = \{1, ..., M\}$ (endowed with the counting measure m_E on E) and that the observed process has two components $Y_k = (Z_k, R_k)$. For every time k, $Z_k \in \mathcal{Z} \subseteq \mathbb{R}^m$ is a vector of m large scale atmospheric variables (covariates) and $R_k \in ([0, +\infty[)^\ell$ is the daily accumulation of rainfall measured at ℓ meteorological stations (output time series) with the value 0 corresponding to dry days. The model aims at describing the conditional distribution of $\{R_k\}$ given $\{Z_k\}$. For this, we assume that*

$$p_{1,\theta}(x_k|x_{k-1}, y_{k-1}) = \frac{q_{x_{k-1},x_k} \exp\left(-1/2\left(z_{k-1} - \mu_{x_{k-1},x_k}\right)' \Sigma^{-1} \left(z_{k-1} - \mu_{x_{k-1},x_k}\right)\right)}{\sum_{x"=1}^M q_{x_{k-1},x"} \exp\left(-1/2\left(z_{k-1} - \mu_{x_{k-1},x"}\right)' \Sigma^{-1} \left(z_{k-1} - \mu_{x_{k-1},x"}\right)\right)}, \tag{18}$$

with $q_{x,x'} > 0$, $\mu_{x,x'} \in \mathbb{R}^m$ and

$$p_{2,\theta}(y_k|x_k, y_{k-s}^{k-1}) = p_Z\left(z_k|z_{k-1}\right) p_{R,\theta}\left(r_k|x_k\right)$$

with respect to $m_{\mathcal{Z}} \otimes m_0^{\otimes \ell}$, where $m_{\mathcal{Z}}$ is the Lebesgue measure on $\mathcal{Z}$ and where m_0 is the sum of the Dirac measure δ_0 and of the Lebesgue measure on $(0, +\infty[$. We observe that $\{Z_k\}_k$ is a Markov chain which transition kernel depends neither on the current weather type nor on the unknown parameter θ (typically Z_k is the output of an atmospheric model and is considered as an input to the Markov switching model) *and that the conditional distribution of R_k given X_k and $\{Y_{k'}\}_{k'<k}$ only depends on X_k as in usual HMMs. Finally the rainfall at the different locations is assumed to be conditionally independent given the weather type*

$$p_{R,\theta}\left(r_k(1),...,r_k(l)|x_k\right)=\prod_{i=1}^{\ell}p_{R_i,\theta}\left(r_k(i)|x_k\right)$$

and the rainfall at the different locations is given by the product of Bernoulli and Gamma variables

$$p_{R_i,\theta}(r_k(i)|x_k)=\begin{cases}1-\pi_i^{(x_k)} & (r_k(i)=0)\\ \pi_i^{(x_k)}\gamma(r_k(i);\alpha_i^{(x_k)},\beta_i^{(x_k)}) & (r_k(i)>0)\end{cases}\tag{19}$$

where $0<\pi_i^{(x)}<1$, $\alpha_i^{(x)}>0$, $\beta_i^{(x)}>0$ and $\gamma(.;\alpha,\beta)$ denotes the pdf of a Gamma distribution with parameters α, β:

$$\gamma(r;\alpha,\beta)=r^{\alpha-1}\frac{\beta^\alpha e^{-\beta r}}{\Gamma(\alpha)}.$$

The parameter θ corresponds to

$$\theta=\left((q_{x,x'}),(\mu_{x,x'}),(\pi_i^{(x)}),(\alpha_i^{(x)}),(\beta_i^{(x)})\right).$$

We write $\tilde\Theta$ for the set of such parameters θ satisfying, for every $x\in\{1,...,M\}$ and every $i\in\{1,...,\ell\}$,

$$\sum_{x'=1}^{M}q_{x,x'}=1,\quad 0<q_{x,x'}<1,\ \sum_{x'=1}^{M}\mu_{x,x'}=0,\ 0<\pi^{(x)}<1,\ \alpha_i^{(x)}>0,\ \textit{and}\ \beta_i^{(x)}>0.$$

The conditions $\sum_{x'=1}^M q_{x,x'}=1$ and $\sum_{x'=1}^M\mu_{x,x'}=0$ come from [**?**]. These conditions are not restrictive. Indeed, q_θ is unchanged if we replace $\mu_{x,x'}$ by $\mu_{x,x'}-\sum_{x"}\mu_{x,x"}$ and $q_{x,x'}$ by $\frac{q_{x,x'}\exp(-(\mu_{x,x'})\Sigma^{-1}\mu_x)}{\sum_{x"}q_{x,x"}\exp(-(\mu_{x,x"})\Sigma^{-1}\mu_x)}$ (with $\mu_x:=\sum_{x"}\mu_{x,x"}$).

Observe that the fact that, if $\mu_{x,x'}=0$ for every x,x', then $\{X_k\}_k$ is an homogeneous Markov chain and $\{Z_k\}_k$ does not plays any role in the dynamics of $\{X_k,R_k\}_k$.

2.3.2 Properties of this Markov chain

We start by recalling a classical result ensurig (6) in the context of HMM (a proof of this result is given in Appendix E for completness).

Lemma 16 (HMM)**.** *Fix θ. Assume that $p_{1,\theta}(x|x',y')=p_{1,\theta}(x|x')$ does not depend on y', $\{X_k\}_k$ is a Markov chain with transition kernel $Q_{1,\theta}$ admitting an invariant pdf $h_{1,\theta}$ (wrt m_E) such that*

$$\lim_{n\to+\infty}\sup_{\nu\in\mathcal{P}(E)}||[Q_{1,\theta}^{*n}\nu]-h_{1,\theta}||_{L^1(m_E)}=0.$$

Assume moreover that $s=0$ (this means that we can take $s=1$ with $p_{2,\theta}(y|x,y')=p_{2,\theta}(y|x)$). Then there exists an invariant measure ν_θ with pdf h_θ (wrt $m_E\times m_K$) given by $h_\theta(x,y):=h_{1,\theta}(x)p_{2,\theta}(y|x)$ and

$$\lim_{n\to+\infty}\sup_{\nu\in\mathcal{P}(E\times K)}||[Q_\theta^{*n}\nu]-h_\theta||_{L^1(m_E\times m_K)}=0.$$

Moreover, if $p_{2,\theta}>0$ and if $\{X_k\}_k$ is an aperiodic positive Harris recurrent Markov chain, then the Markov chain $\{X_k,Y_k\}_k$ is positive Harris recurrent and aperiodic.

Due to this lemma, assumption (6) holds true and $\{X_k,Y_k\}_k$ is aperiodic positive Harris recurrent as soon as $\{X_k,Z_k\}_k$ is aperiodic positive Harris recurent.

The ergodicity of $\{X_k,Y_k\}_k$ will also follow from the ergodicity of $\{X_k,Z_k\}_k$.

2.3.3 Consistency of MLE

Corollary 17. *Assume Hypothesis 15. Assume that Θ is a compact subset of $\tilde{\Theta}$ and that, for every $\theta \in \Theta$, the transition kernel $Q_{0,\theta}$ of the Markov chain $\{X_k, Z_k\}_k$ admits an invariant pdf $h_{0,\theta} > 0$ (wrt $m_E \times m_{\mathcal{Z}}$) such that*

$$\lim_{n\to+\infty} \sup_{\nu\in\mathcal{P}(E\times\mathcal{Z})} \|[Q_{0,\theta}^{*n}\nu] - h_{0,\theta}\|_{L^1(m_E\times m_{\mathcal{Z}})} = 0. \tag{20}$$

Assume moreover that $\mathcal{Z}$ is compact, that

$$\forall z \in \mathcal{Z}, \quad \sup_{z_{-1}\in\mathcal{Z}} p_Z(z|z_{-1}) < \infty \tag{21}$$

and that

$$\bar{\mathbb{E}}_{\theta^*}[|\log p_Z(Z_0|Z_{-1})|] < \infty. \tag{22}$$

Then, for every $x_0 \in \{1, ..., M\}$, on a set of probability one (for $\bar{\mathbb{P}}_{\theta^*}$), the limit values θ of the sequence of random variables $(\hat{\theta}_{n,x_0})_n$ are $\bar{\mathbb{P}}_{\theta^*}$-almost surely contained in $\{\theta \in \Theta \ : \ \bar{\mathbb{P}}_\theta = \bar{\mathbb{P}}_{\theta^*}\}$.

If, moreover, $\{X_k, Z_k\}_k$ is aperiodic and positive Harris recurrent then this result holds true for any initial probability distribution.

Proof. Due to the previous section, we know that (20) implies (6) and that the aperiodicity and positive Harris recurrence of $\{X_k, Z_k\}_k$ implies the positive Harris recurrence of $\{X_k, Y_k\}_k$.

The fact that Θ is a compact subset of $\tilde{\Theta}$ directly implies (2).

Assumption (5) holds true since E is finite, since $p_{R,\theta}(r|x) < \infty$ for every $(x, y) \in E \times K$ and according to (21).

Now according to (22), (3) and (4) will follow from the fact that, for every $x_0 \in X$ and every $i \in \{1, ..., \ell\}$,

$$\bar{\mathbb{E}}_{\theta^*}\left[\left|\log\left(\inf_\theta p_{R_i,\theta}(R_i|x_0)\right)\right|\right] + \bar{\mathbb{E}}_{\theta^*}\left[\left|\log\left(\sup_\theta p_{R_i,\theta}(R_i|x_0)\right)\right|\right] < \infty.$$

Now we observe that if $R_i = 0$, then

$$0 < 1 - \pi_+ \le p_{R_i,\theta}(R_i|x_0) \le 1 - \pi_-,$$

where π_- and π_+ are the minimal and maximal possible values of $\pi_i^{(x)}$ (for $x \in X$, $i \in \{1, ..., \ell\}$ and θ in the compact set Θ). Analogously, let us write α_-, α_+ for the minimal and maximal possible values of $\alpha_i^{(x)}$ and β_-, β_+ for the minimal and maximal possible values of $\beta_i^{(x)}$. Since, all this quantities are positive and finite, due to the expression of $\log(p_{R_i,\theta}(R_i|x_0))$, to prove (3) and (4), it is enough to prove that

$$\bar{\mathbb{E}}_{\theta^*}[R_i] < \infty \quad \text{and} \quad \bar{\mathbb{E}}_{\theta^*}[|\log(R_i)|\mathbf{1}_{\{R_i>0\}}] < \infty.$$

Observe that, under the stationary distribution, the pdf h_i of R_i satisfies:

$$\forall r > 0, \quad h_i(r) \le (r^{\alpha_- -1}\mathbf{1}_{\{r\le 1\}} + r^{\alpha_+ -1}\mathbf{1}_{\{r>1\}})\frac{\max(\beta_+^{\alpha_+}, \beta_+^{\alpha_-})e^{-r\beta_-}}{\Gamma(\alpha_-)}.$$

Therefore, (3) and (4) come from the facts that $r \mapsto r^{\alpha_+ -1}e^{-r\beta_-}$ is integrable at $+\infty$ (since $\beta_- > 0$) and that $r \mapsto |\log r|r^{\alpha_- -1}$ is integrable at 0 (since $\alpha_- > 0$). ☐

Now we will add an asumption on θ to ensure the identifiability of the parameter. If we assume $\pi_i^{(x)} = 0$ for every i and every x, then identifiability follows easily if we assume moreover that

$$x \ne x' \ \Rightarrow \ (\alpha_i^{(x)}, \beta_i^{(x)})_i \ne (\alpha_i^{(x')}, \beta_i^{(x')})_i. \tag{23}$$

But, if we do not assume $\pi_i^{(x)} = 0$, (23) does not ensure identifiability anymore. We give now an explicit counter-example.

Remark 18. *Assume $M = \ell = 2$. We consider two models A_1 and A_2 associated to θ_1 and θ_2 respectively, with*

$$\theta_j = \Big((q_{x,x',(j)}), (\mu_{x,x',(j)}), (\pi_i^{(x,(j))}), (\alpha_i^{(x,(j))}), (\beta_i^{(x,(j))}) \Big),$$

and

- $q_{x,x',(1)} = 0.5$, $\mu_{x,x',(1)} = 0$, $\pi_i^{(x,(1))} = 0.5$, $\alpha_i^{(x,(1))} = 1$, $\beta_1^{(x,(1))} = 1$, $\beta_2^{(1,(1))} = 2$, $\beta_2^{(2,(1))} = 3$,
- $q_{x,1,(2)} = 0.6$, $q_{x,2,(2)} = 0.4$, $\mu_{x,x',(2)} = 0$, $\pi_1^{(x,(2))} = 0.5$, $\pi_2^{(1,(2))} = \frac{0.25}{0.6}$, $\pi_2^{(2,(2))} = \frac{0.25}{0.4}$, $\alpha_i^{(x,(2))} = 1$, $\beta_1^{(x,(1))} = 1$, $\beta_2^{(1,(1))} = 2$, $\beta_2^{(2,(1))} = 3$.

For model A_1 (under the stationary measure), $\{X_k\}$ is an iid sequence on $\{1,2\}$ with $\mathbb{P}(X_1 = 1) = 0.5$ and the distribution of R_k given $\{X_k = 1\}$ is $(0.5\delta_0 + 0.5\Gamma(1,1)) \otimes (0.5\delta_0 + 0.5\Gamma(1,2))$ whereas the distribution of R_k taken $\{X_k = 2\}$ is $(0.5\delta_0 + 0.5\Gamma(1,1)) \otimes (0.5\delta_0 + 0.5\Gamma(1,3))$. Hence, for the model A_1, the R_k are iid with distribution

$$(0.5\delta_0 + 0.5\Gamma(1,1)) \otimes (0.5\delta_0 + 0.25\Gamma(1,2) + 0.25\Gamma(1,3)). \tag{24}$$

For model A_2 (under the stationary measure), $\{X_k\}$is an iid sequence on $\{1,2\}$ with $\mathbb{P}(X_1 = 1) = 0.6$ and the distribution of R_k given $\{X_k = 1\}$ is $(0.5\delta_0 + 0.5\Gamma(1,1)) \otimes \big((1 - \frac{0.25}{0.6})\delta_0 + \frac{0.25}{0.6}\Gamma(1,2)\big)$ whereas the distribution of R_k taken $\{X_k = 2\}$ is $(0.5\delta_0 + 0.5\Gamma(1,1)) \otimes \big((1 - \frac{0.25}{0.4})\delta_0 + \frac{0.25}{0.4}\Gamma(1,3)\big)$. Hence, for the model A_2, the R_k are iid with distribution (24).

Observe that the distribution of $\{Y_k\}$ under the stationary measure is the same for models A_1 and A_2.

The next result (proved in appendix D) states that the following condition ensures identifiability

$$x \neq x' \ \Rightarrow \ \forall i \in \{1, ..., \ell\},\ (\alpha_{i,\theta_1}^{(x)}, \beta_{i,\theta_1}^{(x)}) = (\alpha_{i,\theta_1}^{(x')}, \beta_{i,\theta_1}^{(x')}). \tag{25}$$

Proposition 19. *Assume Hypothesis 15. Let θ_1 and θ_2 in $\tilde{\Theta}$, with*

$$\theta_j = \Big((q_{x,x',(j)}), (\mu_{x,x',(j)}), (\pi_i^{(x,(j))}), (\alpha_i^{(x,(j))}), (\beta_i^{(x,(j))}) \Big).$$

Assume that θ_1 satisfies (25).

Then $\bar{\mathbb{P}}_{\theta_1}^Y = \bar{\mathbb{P}}_{\theta_2}^Y$ if and only θ_1 and θ_2 are equal up to a permutation of indices, i.e. there exists a permutation τ of $\{1, ..., M\}$ such that, for every $x, x' \in \{1, ..., M\}$ and every $i \in \{1, ..., \ell\}$, we have $q_{x,x',(1)} = q_{\tau(x),\tau(x'),(2)}$, $\mu_{x,x',(1)} = \mu_{\tau(x),\tau(x'),(2)}$, $\pi_i^{(x,(1))} = \pi_i^{(\tau(x),(2))}$, $\alpha_i^{(x,(1))} = \alpha_i^{(\tau(x),(j))}$, $\beta_i^{(x,(1))} = \beta_i^{(\tau(x),(2))}$.

Now the following result is a direct consequence of Corollary 17 and Proposition 19.

Theorem 20. *Assume Hypothesis 15. Assume that Θ is a compact subset of $\tilde{\Theta}$ and that, for every $\theta \in \Theta$, the transition kernel $Q_{0,\theta}$ of the Markov chain $(X_k, Z_k)_k$ admits an invariant pdf $h_{0,\theta}$ (wrt $m_E \times m_{\mathcal{Z}}$) satisfying (20). Assume that θ^* satisfies (25). Assume moreover that $\mathcal{Z}$ is compact, that (21) and (22) hold true. Then, for every $x_0 \in \{1, ..., M\}$, on a set of probability one (for $\bar{\mathbb{P}}_{\theta^*}$), the limit values θ of the sequence of random variables $(\hat{\theta}_{n,x_0})_n$ are equal to θ^* up to a permutation of indices.*

If, moreover, $(X_k, Z_k)_k$ is aperiodic and positive Harris recurrent then this result holds true for any initial probability distribution.

3 Applications to real data

3.1 MacKenzie River Lynx Data

In this section we discuss the results obtained when fitting the model introduced in Subsection 2.1 to the annual number of Canadian lynx trapped in the Mackenzie River district of northwest Canada from 1821 to 1934. This time series is a benchmark dataset to test nonlinear time series model (see e.g. [?], [?]). In order to facilitate the comparison with the other works on this time series, we analyze the data at the logarithm scale with the base 10 shown on Figure 1. This time series exhibits periodic fluctuations (it may be due to the competition between several species, predator-prey interaction,...) with asymmetric cycles (increasing phase are slower than decreasing phase) which makes it challenging to model.

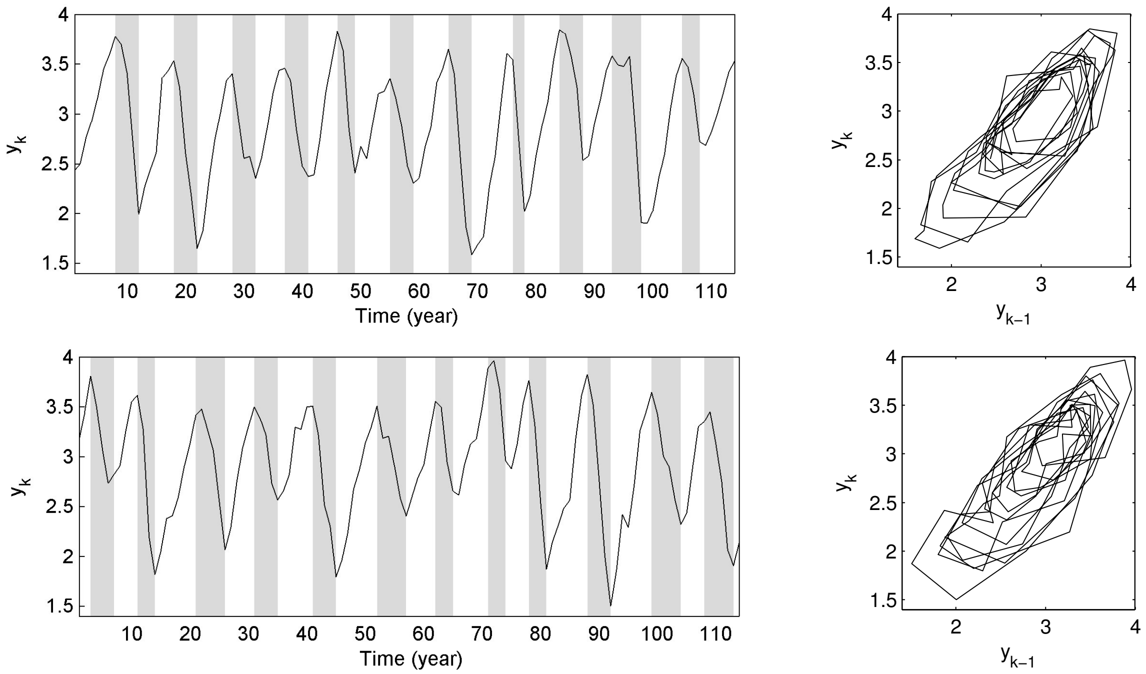

Figure 1: Top left panel: time plot of log Canadian lynx data. The color indicates the most likely regimes identified by the fitted NHMS-AR model. The first [resp. second] regime is the most likely when the color is white [resp. gray]. Top right panel: directed scatter plot of log Canadian lynx data. Bottom left panel: time plot of a sequence simulated with the fitted NHMS-AR model data. The color indicates the simulated regime (first regime in white, second regime in gray). Bottom right panel: directed scatter plot of the simulated sequence shown on the bottom left panel.

In [?], it was proposed to fit a SETAR(2) model to this time series. The fitted model is the following

$$Y_k = \begin{cases} 0.51 + 1.23Y_{k-1} - 0.37Y_{k-2} + 0.18\epsilon_k & (Y_{k-2} \leq 3.15) \\ 2.32 + 1.53Y_{k-1} - 1.27Y_{k-2} + 0.23\epsilon_k & (Y_{k-2} > 3.15) \end{cases} . \tag{26}$$

The two regimes have a nice biological interpretation in terms of prey-predator interaction, with the upper regime ($Y_{t-2} > 3.15$) corresponding to a population decrease whereas the population tends to increase in the lower regime.

A NHMS-AR model has been fitted to this time series. In practice, we have used the EM algorithm to compute the MLE. The recursions of this algorithm are relatively similar to the ones of the MS-AR model (see [?], [?]). To facilitate the comparison with (26), we have also considered AR models of order $s = 2$ and a lag $r = 2$ for the transition probabilities. The fitted model is the following

$$Y_k = \begin{cases} \begin{array}{lllllll} 0.54 & +1.11 & Y_{k-1} & -0.24 & Y_{k-2} & +0.14 & \epsilon_t \quad (X_k = 1) \\ \textit{(0.31,0.80)} & \textit{(0.96,1.27)} & & \textit{(-0.43,-0.05)} & & \textit{(0.11,0.17)} & \\ 1.03 & +1.49 & Y_{k-1} & -0.87 & Y_{k-2} & +0.22 & \epsilon_t \quad (X_k = 2) \\ \textit{(-0.12,1.86)} & \textit{(1.23,1.69)} & & \textit{(-1.20,-0.39)} & & \textit{(0.14,0.26)} & \end{array} \end{cases} \tag{27}$$

with

$$P\left(X_k = i | X_{k-1} = i, Y_{k-2} = y_{k-2}\right) = \begin{cases} \begin{array}{lllll} (1 + exp(& -42.4 & +12.8 & y_{k-2}))^{-1} & (X_k = 1) \\ & \textit{(-587,-16.3)} & \textit{(4.77,176)} & & \\ (1 + exp(& 9.07 & -3.33 & y_{k-2}))^{-1} & (X_k = 2) \\ & \textit{(2.25,178)} & \textit{(-64.1,-1.12)} & & \end{array} \end{cases} \tag{28}$$

where the italic values in parenthesis below the parameter values correspond to 95% confidence intervals computed using parametric bootstrap (see e.g. [?]). The estimate of $\pi_-^{(x)}$ and $\pi_+^{(x)}$ are not given because they are very close to 0. It means that these technical parameters have no practical importance and can be fixed equal to an arbitrary small value (here we used the machine epsilon 2^{-52}). There are small differences between the AR coefficients (26) and (27) but the dynamics inside the regimes of the SETAR(2) and NHMS-AR models are broadly similar. The models differ mainly in the mechanism used to govern the switchings between the two regimes. For the SETAR model the regime is a deterministic function of a lagged value of the observed process. The NHMS-AR model can be seen as a fuzzy extension of the SETAR model where the regime has its own Markovian evolution influenced by the lagged value of the observed process. This is illustrated on Figure 2 which shows the transition probabilities (28) and the threshold of the SETAR(2) model. According to this figure, it seems reasonable to model the transition from regime 1 to regime 2 by a step function at the level $y_{k-2} \approx 3.15$ but the values of y_{k-2} for which the transition from regime 2 to regime 1 occurs seem to be more variable and the step function approximation less realistic.

The asymmetries in the cycle imply that the system spends less time in the second regime (decreasing phase) than in the first one. It may explain the larger confidence intervals in the second regime compared to the first one (see (27)). Figure 2 shows that there is an important sampling variability in the estimate of the transition kernel of the hidden process. This is probably due to the low number of transitions among regimes (see Figure 1) which makes it difficult to estimate the associated parameters. A similar behavior has been observed when fitting the model to other time series.

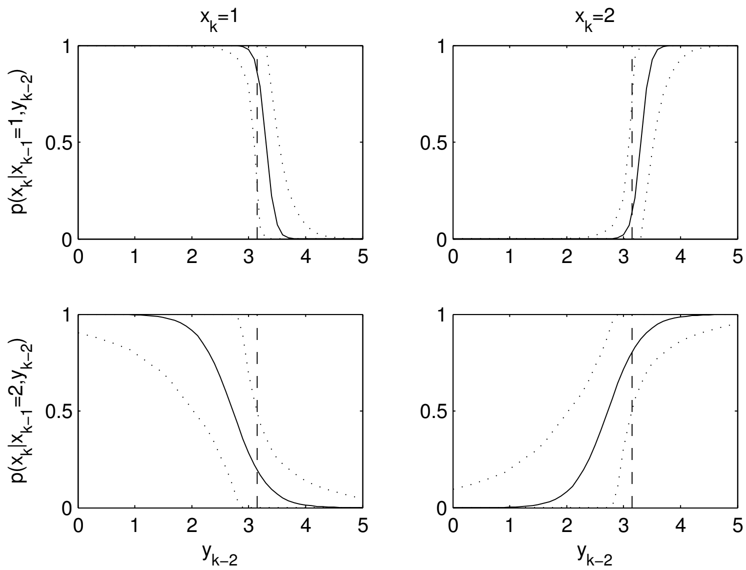

Figure 2: Transition probabilities $P\left(X_k = j | X_{k-1} = i, Y_{k-2} = y_{k-2}\right)$ as a function of y_{k-2}. The dotted lines correspond to 95% confidence intervals computed using parametric bootstrap. The dashed vertical line corresponds to the threshold (3.15) of the SETAR(2) model.

Table 1 gives the AIC and BIC values defined as

$$AIC = -2logL + 2npar, \quad BIC = -2logL + npar\log(N)$$

and L is the likelihood of the data, $npar$ is the number of parameters and N is the number of observations. The values for the NHMS-AR and SETAR models are relatively similar. The NHMS-AR models has a slightly better AIC value but BIC selects the SETAR model. As expected, these two models clearly outperform the homogeneous MS-AR which does not include information on the past values in the switching mechanism.

	AIC	BIC	npar
SETAR ($s = 2$)	-28.33	-3.70	9
MS-AR ($s = 2$)	-0.2063	27.15	10
NHMS-AR ($r = s = 2$)	-30.83	2.00	12

Table 1: AIC and BIC values for the fitted SETAR, homogeneous MS-AR and NHMS-AR models

The simulated sequence shown on Figure 1 exhibits a similar cyclical behavior than the data. A more systematic validation was performed but the results are hard to analyze because of the low amount of data available. Note that the fitted model is stable since it satisfies (10) for the matrix norm defined as

$$\|A\| = \left\|P^{-1}AP\right\|_\infty$$

with P the matrix containing the eigenvectors of the companion matrix for the second regime and $\|.\|_\infty$ the infinity norm.

A more systematic validation is performed on a longer time series in the next section.

3.2 Wind direction

Various approaches have been proposed in the literature for modeling time series of wind speed (see [**?**] and references therein). In comparison, there exist only very few models for time series of wind direction which is an important meteorological parameter for many applications. Some models have been proposed in the literature for circular time series (see [**?**], [**?**],[**?**], [**?**]) and some of them have been applied to time series of wind direction. However they are not able to catch the complex features of the time series of wind direction considered in this work.

We use data from the ERA-40 data set which consists in a global reanalysis with 6-hourly data covering the period from 1958 to 2001. It can be freely downloaded and used for scientific purposes at the URL: *http://data.ecmwf.int/data*
We have extracted the wind data for the point with geographical coordinates (47.5^0 N, 5^0 W) from this data set. It is located off the Brittany coast (northwest of France).

It leads to a long time series which is non-stationary since it exhibits an important seasonal component but also diurnal and interannual components. A classical approach for treating seasonality in meteorological time series consists in blocking the data, typically by period ranging from a month to a trimester depending on the amount of data available, and in fitting a separate model for each period in the year. This approach is used in the present paper and we have chosen to focus on the months of January. It leads to 44 time series of length 124 (31 days with 4 observations per day), each time series describing the wind conditions during the months of January for a particular year. In the sequel, we assume that these time series are independent realizations of a stationary process. It seems realistic according to the results given in [**?**] for the wind speed at the same location since the diurnal components can be neglected during the winter season. Following [**?**], another approach would consist in letting some of the coefficients of the model introduced below to evolve in time with periodic functions for the diurnal and seasonal components and eventually a trend.

The marginal distribution of the time series of wind direction considered in this work is shown on Figure 3. It clearly exhibits two modes, each one corresponding to a meteorological regime: the prevailing mode corresponds to westerlies cyclonic conditions with low pressure systems coming from the Atlantic ocean whereas the second mode is associated to anticyclonic conditions and wind blowing from the east. This is an usual feature of meteorological time series. A classical approach for modeling these meteorological regimes (or "weather types") consists in introducing a hidden (or latent) variable. This idea goes back to [**?**] where HMMs were proposed for modelling the space-time evolution of daily rainfall (see [**?**] for

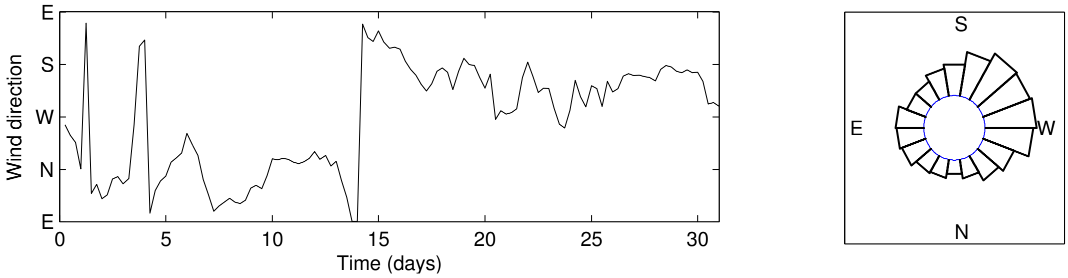

Figure 3: Wind direction for the month of January 1968 (left panel) and rose plot (right panel) of the marginal distribution of the wind direction in January (results obtained with the 44 years of data).

more recent references on this topic). HMMs have also been proposed for modeling time series of wind direction (see [?], [?]). However HMMs assume that successive observations are conditionally independent given the latent weather type and fail in reproducing the strong relation which exists between the wind conditions at successive time steps (see Figure 3).

Several autoregressive models have been proposed in the literature for directional time series (see [?] and references therein). They are all candidates to model the dynamics of the wind direction in the weather types but in this work we have chosen to focus on the von Mises process initially introduced in [?]. It is based on the von Mises distribution which is a natural distribution for circular variables (see [?]) admitting a pdf f_γ (with respect to the Lebesgue measure on $\mathbb{T}$) given by

$$\forall y \in \mathbb{T}, \quad f_\gamma(y) = \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa \cos(y-\phi)\right) = \frac{1}{2\pi I_0(\kappa)} \left| e^{\gamma e^{-iy}} \right|, \tag{29}$$

for some complex parameter $\gamma := \kappa e^{i\phi}$ (with $\kappa \geq 0$ and $\phi \in \mathbb{T}$), where I_0 denotes the modified Bessel function of order 0 defined as

$$I_0(\kappa) := \frac{1}{2\pi} \int_{\mathbb{T}} \exp(\kappa \cos(y))\, dy.$$

In (29), $\phi \in \mathbb{T}$ corresponds to the circular mean of the distribution and $\kappa \geq 0$ describes the concentration of the distribution: when $\kappa = 0$ we get the uniform distribution whereas when κ increases the distribution is more and more concentrated around ϕ. We assume that

$$p_2(y_k | x_k, y_{k-s}^{k-1}) = f_{\gamma_0^{(x_k)} + \sum_{\ell=1}^{s} \gamma_\ell^{(x_k)} e^{i y_{k-\ell}}}(y_k) \tag{30}$$

with $\gamma_0 = \kappa_0 e^{i\phi_0} \in \mathbb{C}$. In [?], the autoregressive parameters $\gamma_\ell^{(x_k)}$ for $\ell \geq 1$ are assumed to have real values. In this work, we extend this model by assuming that $\gamma_\ell^{(x)} = \kappa_\ell^{(x)} e^{i\phi_\ell^{(x)}} \in \mathbb{C}$. We will see in the sequel that it helps modeling the prevailing clockwise rotation of the wind direction.

The parametrization used to model the dependence of the weather change with the previous wind direction is also based on the pdf of von Mises distribution since we assume that

$$p_1(x_k | x_{k-s}^{k-1}, y_{t-1}) \propto q_{x_{k-1}, x_k} \exp\left(\lambda_{x_{k-1}, x_k} \cos\left(y_{k-1} - \psi_{x_{k-1}, x_k}\right)\right), \tag{31}$$

where $Q = (q_{x,x'})_{x,x' \in \{1,\dots,M\}}$ is a stochastic matrix and, for $x, x' \in \{1, \dots, M\}$, $\lambda_{x,x'} \geq 0$ and $\psi_{x,x'} \in \mathbb{T}$ are unknown parameters. Loosely speaking, the probability that the hidden Markov chain $\{X_t\}$ switches from x to x' will increase when the wind direction y_{k-1} is close to $\psi_{x,x'}$ and $\lambda_{x,x'}$ models the directional spreading in which this transition is likely to occur. When $\lambda_{x,x'} = 0$ for every $x, x' \in \{1, \dots, M\}$ then $p_1(x_k | x_{k-1}, y_{k-1}) = q_{x_{k-1}, x_k}$ does not dependent on y_{k-1} and we obtain again the homogeneous MS-AR models. Observe that (31) is the same as (14) with $\tilde{\lambda}_{x,x'} = \lambda_{x,x'} e^{i\psi_{x,x'}}$.

The model, which theoretical properties are discussed in Section 2.2, was fitted using the EM algorithm with a number of regimes M varying from 1 to 6. We also varied the order s of the autoregressive models from $s = 0$ to $s = 5$ and considered various reduced models. The BIC values together with various diagnostic plots (see discussion below) led us to focus on the model with $M = 4$ regimes and autoregressive models of order $s = 2$. The results discussed below have been obtained with (15) and the additional constraints

- $\tilde{\lambda}_{x,x''} = \tilde{\lambda}_{x',x''}$ for every $x, x', x'' \in \{1, ..., M\}$ such that $x \neq x''$ and $x' \neq x''$
- $\phi_\ell^{(x)} = \phi_1^{(x)}$ for every $\ell \geq 2$, $x \in \{1, ..., M\}$

These two constraints were considered in order to get more parsimonious and interpretable models and are justified a posteriori by the AIC and BIC criteria. The consistency results of section 2.2 remain valid with these constraints.

The BIC and AIC values for a few representative models which have been fitted are given in Table 2. These criteria clearly select the most sophisticated NHMS-AR model with an important improvement against the HMM and AR models which were proposed before in the literature. There is also a small improvement over the homogeneous MS-AR model. This is further discussed below.

	M	s	Constraints for $\ell \geq 1$, $x, x' \in \{1, ..., M\}$	BIC	AIC	npar
HMM	4	0	$\gamma_\ell^{(x)} = 0$ $\lambda_{x,x'} = 0$	11619	11751	20
NHMM	4	0	$\gamma_\ell^{(x)} = 0$	10289	10474	28
AR	1	2		7528	7568	4
MS-AR	4	2	$\lambda_{x,x'} = 0,$	5724	5918	28
NHMS-AR	4	2		5607	5854	36

Table 2: BIC and AIC values for the NHMS-AR model and various reduced models.

The main motivation of this work is to develop stochastic models which can be used to generate realistic time series of wind direction (stochastic weather generators). In this context, it seems natural to validate the model by simulating a large number of artificial sequences of the model and comparing the statistical properties of these simulations with the ones of the original data. According to Figure 4, the NHMS-AR model clearly improves the description of the marginal distribution of the process compared to the MS-AR model which is not able to reproduce the second mode of the distribution associated to easterlies. There are also important improvements as concerns the description of the dynamics of the process although it remains some significant discrepancies. In particular, the fitted NHMS-AR model slightly underestimates the circular autocorrelation function defined as (see [**?**])

$$\rho(k) = \frac{E[\cos(Y_0)\cos(Y_k)] + E[\sin(Y_0)\sin(Y_k)] - E[\sin(Y_0)\cos(Y_k)] - E[\cos(Y_0)\sin(Y_k)]}{E[\cos(Y_0)^2]E[\sin(Y_0)^2] - E[\sin(Y_0)\cos(Y_0)]^2}$$

for lags between 2 and 5 days and some coefficients of the cross-correlation function between $\{\cos(Y_k)\}$ and $\{\sin(Y_k)\}$. The sample cross-correlation function computed on the data is at its maximum value for a lag between 18 hours and 24 hours, with the time series $\{\sin(Y_t)\}$ being in advance of the time series $\{\cos(Y_t)\}$ because of the prevailing clockwise rotation of the wind direction. The NHMS-AR model is able to reproduce the shape of this cross-correlation function but slightly underestimates the maximum correlation. Similar plots were done for the more usual HMM and AR models and we obtained substantially less good results compared to MS-AR and NHMS-AR models.

A Consistency : proof of Theorem 2

We follow the proof of [**?**, Thm. 1] with slight modifications due to our assumptions

(see Lemmas 28 and 29). We do not give all the details of the proofs since some of them are a direct rewriting of [**?**]. First, we consider the stationary case. Let τ be the full shift on $\Omega := (E \times K)^{\mathbb{Z}}$. For every $k \in \mathbb{Z}$, we identify X_k with $X_0 \circ \tau^k$ and Y_k with $Y_0 \circ \tau^k$, where $X_0((x_m, y_m)_{m\in\mathbb{Z}}) := x_0$ and $Y_0((x_m, y_m)_{m\in\mathbb{Z}}) = y_0$.

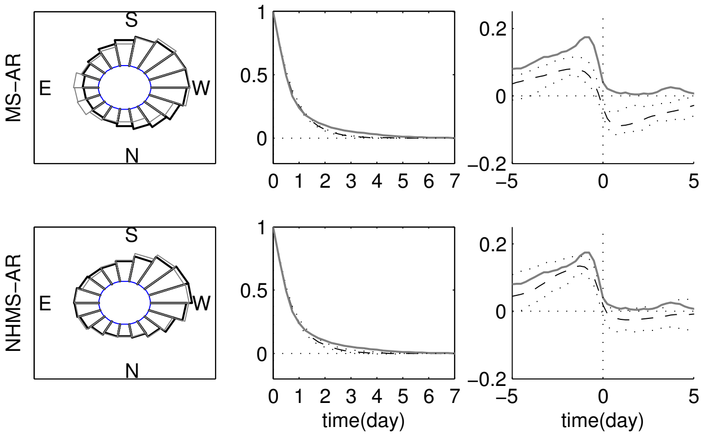

Figure 4: Rose plot of the marginal distribution (left panels), circular autocorrelation functions (middle panels), cross-correlation functions between the time series $\{\cos(Y_t)\}$ and $\{\sin(Y_t)\}$ for the fitted MS-AR (top panels) and NHMS-AR models (bottom panels). The full grey line corresponds to the sample functions and the dashed line to the fitted model with a 95% prediction intervals (dotted line). The distribution for the fitted model was obtained by simulation.

A.1 Likelihood and stationary likelihood

We start by recalling a classical fact in the context of Markov chains (and the proof of which is direct).

Fact 21. *Let m and n belong to $\mathbb{Z}$ with $m \leq n$. Under $\bar{\mathbb{P}}_\theta$, conditionally to (Y_{m-s+1}^n), $(X_k)_{k\in\{m,\dots,n\}}$ is a (possibly nonhomogeneous) Markov chain. Moreover, under $\bar{\mathbb{P}}_\theta$, the conditional pdf (wrt m_E) of X_k given (X_m^{k-1}, Y_{m-s+1}^n) is given by*

$$p_\theta(X_k = x_k | X_m^{k-1}, Y_{m-s+1}^n) = \frac{p_\theta(Y_k^n, X_k = x_k | X_{k-1}, Y_{k-s}^{k-1})}{p_\theta(Y_k^n | X_{k-1}, Y_{k-s}^{k-1})} \quad \bar{\mathbb{P}}_\theta - a.s., \tag{32}$$

with

$$p_\theta(Y_k^n, X_k = x_k | X_{k-1} = x_{k-1}, Y_{k-s}^{k-1}) := \int_{E^{n-k}} \prod_{j=k}^{n} q_\theta(x_j, Y_j | x_{j-1}, Y_{j-s}^{j-1})\, dm_E^{\otimes(n-k)}(x_{k+1}^n) \tag{33}$$

and

$$p_\theta(Y_k^n | X_{k-1}, Y_{k-s}^{k-1}) := \int_E p_\theta(Y_k^n, X_k = x_k | X_{k-1}, Y_{k-s}^{k-1})\, dm_E(x_k). \tag{34}$$

Using (2), (3) and (4), we observe that the quantities appearing in this fact are well-defined. Due to Fact 21, the quantity $\bar{p}_\theta(X_k = x_k | X_{k-1}, Y_{m-s+1}^n)$ is equal to

$$\frac{\int_{E^{n-k+1}} (\prod_{j=k+1}^n a_j) p_{1,\theta}(\tilde{x}_k | X_{k-1}, Y_{k-s}^{k-1}) p_{2,\theta}(Y_k | \tilde{x}_k, Y_{k-s}^{k-1})\, d\delta_{x_k}(\tilde{x}_k)\, dm_E^{\otimes(n-k)}(\tilde{x}_{k+1}^n)}{\int_{E^{n-k+1}} (\prod_{j=k+1}^n a_j) p_{1,\theta}(\tilde{x}_k | X_{k-1}, Y_{k-s}^{k-1}) p_{2,\theta}(Y_k | \tilde{x}_k, Y_{k-s}^{k-1})\, dm_E^{\otimes(n-k+1)}(\tilde{x}_k^n)},$$

with $a_j := q_\theta(\tilde{x}_j, Y_j | \tilde{x}_{j-1}, Y_{j-s}^{j-1})$. Therefore

$$\bar{p}_\theta(X_k = x_k | X_{k-1}, Y_{m-s+1}^n) \geq \frac{p_{1,-}}{p_{1,+}} \beta(x_k), \text{ with } \beta(x_k) := \frac{p_\theta(Y_k^n | X_k = x_k, Y_{k-s}^{k-1})}{\int_E p_\theta(Y_k^n | X_k = \tilde{x}_k, Y_{k-s}^{k-1})\, dm_E(\tilde{x}_k)}. \tag{35}$$

From this last inequality (since $0 < p_{1,-} < p_{1,+} < \infty$), we directly get the following (from [?]).

Corollary 22. *(as [?, Cor. 1]) For all $m \le k \le n$ and every probability measures m_1 and m_2 on E, we have, $\bar{\mathbb{P}}_\theta - a.s.$*

$$\left\| \int_E \bar{\mathbb{P}}_\theta(X_k \in \cdot | X_m = x_m, Y_{m-s+1}^n)\, dm_1(x_m) - \int_E \bar{\mathbb{P}}_\theta(X_k \in \cdot | X_m = x_m, Y_{m-s+1}^n)\, dm_2(x_m) \right\|_{TV} \le \rho^{k-m},$$

with $\rho := 1 - \frac{p_{1,-}}{p_{1,+}}$.

Observe that the log-likelihood $\ell_n(\theta, x_0)$ satisfies

$$\ell_n(\theta, x_0) = \sum_{k=1}^n \log p_\theta(Y_k | X_0 = x_0, Y_{-s+1}^{k-1}) \quad \bar{\mathbb{P}}_\theta - a.s.,$$

with

$$\begin{aligned} p_\theta(Y_k|X_0 = x_0, Y_{-s+1}^{k-1}) &:= \frac{p_\theta(Y_1^k | X_0 = x_0, Y_{-s+1}^0)}{p_\theta(Y_1^{k-1}|X_0 = x_0, Y_{-s+1}^0)} \\ &= \int_{E^2} q_\theta(x_k, Y_k | x_{k-1}, Y_{k-s}^{k-1}) p_\theta(X_{k-1} = x_{k-1} | X_0 = x_0, Y_{-s+1}^{k-1})\, dm_E^{\otimes 2}(x_k, x_{k-1}). \end{aligned}$$

Let us now define the stationary log-likelihood $\ell_n(\theta)$ by

$$\ell_n(\theta) := \sum_{k=1}^n \log \bar{p}_\theta(Y_k | Y_{-s+1}^{k-1}),$$

with

$$\bar{p}_\theta(Y_k|Y_{-s+1}^{k-1}) := \int_{E^2} q_\theta(x_k, Y_k | x_{k-1}, Y_{k-s}^{k-1}) \bar{p}_\theta(X_{k-1} = x_{k-1} | Y_{-s+1}^{k-1})\, dm_E^{\otimes 2}(x_k, x_{k-1})$$

and

$$\bar{p}_\theta(X_{k-1} = x_{k-1}|Y_{-s+1}^{k-1}) := \int_E p_\theta(X_{k-1} = x_{k-1} | X_0 = x_0, Y_{-s+1}^{k-1}) \bar{p}_\theta(X_0 = x_0 | Y_{-s+1}^{k-1})\, dm_E(x_0).$$

Lemma 23. *(as [?, Lem. 2]) We have*

$$\sup_{x_0 \in E} \sup_{\theta \in \Theta} |\ell_n(\theta, x_0) - \ell_n(\theta)| \le \frac{1}{(1-\rho)^2} \quad \bar{\mathbb{P}}_{\theta^*} - a.s., \tag{36}$$

Proof. We have

$$\begin{aligned} &\sup_{x_0 \in E} |p_\theta(Y_k | X_0 = x_0, Y_{-s+1}^{k-1}) - \bar{p}_\theta(Y_k|Y_{-s+1}^{k-1})| \le \\ &\qquad \le p_{1,+} \int_{E^3} p_{2,\theta}(Y_k | x_k, Y_{k-s}^{k-1}) D(x_{k-1}, x_0, x) \bar{p}_\theta(X_0 = x | Y_{-s+1}^{k-1})\, dm_E^{\otimes 3}(x, x_{k-1}, x_k), \end{aligned}$$

with $D(x_{k-1}, x_0, x) := |p_\theta(X_{k-1} = x_{k-1} | X_0 = x_0, Y_{-s+1}^{k-1}) - p_\theta(X_{k-1} = x_{k-1} | X_0 = x, Y_{-s+1}^{k-1})|$. Due to Corollary 22, we have

$$|p_\theta(Y_k|X_0 = x_0, Y_{-s+1}^{k-1}) - \bar{p}_\theta(Y_k | Y_{-s+1}^{k-1})| \le p_{1,+}\rho^{k-1} \int_E p_{2,\theta}(Y_k | x_k, Y_{k-s}^{k-1})\, dm_E(x_k).$$

Since $|p_\theta(Y_k|X_0, Y_{-s+1}^{k-1})|$ and $|p_\theta(Y_k|Y_{-s+1}^{k-1})|$ are both larger than or equal to

$$p_{1,-} \int_E p_{2,\theta}(Y_k | x_k, Y_{k-s}^{k-1})\, dm_E(x_k),$$

we obtain that

$$\begin{aligned} \left|\log p_\theta(Y_k | X_0 = x_0, Y_{-s+1}^{k-1}) - \log \bar{p}_\theta(Y_k | Y_{-s+1}^{k-1})\right| &\le \frac{|p_\theta(Y_k|X_0 = x_0, Y_{-s+1}^{k-1}) - \bar{p}_\theta(Y_k|Y_{-s+1}^{k-1})|}{p_{1,-}\int_E p_{2,\theta}(Y_k|x_k, Y_{k-s}^{k-1})\, dm_E(x_k)} \\ &\le \rho^{k-1} \frac{p_{1,+}}{p_{1,-}} = \frac{\rho^{k-1}}{1-\rho} \quad \bar{\mathbb{P}}_\theta - a.s. \end{aligned} \tag{37}$$

and so (36) since $\bar{\mathbb{P}}_{\theta^*}$ is absolutely continuous with respect to $\bar{\mathbb{P}}_\theta$ (for all θ). $\square$

A.2 Asymptotic behavior of the log-likelihood

The idea is to approximate $n^{-1}\ell_n(\theta)$ by $n^{-1}\sum_{k=1}^n \log p_\theta(Y_k|Y_{-\infty}^{k-1})$. To this end, we define, for any $k\geq 0$, any $m\geq 0$ and any $x_0\in E$, the following quantities

$$\Delta_{k,m,x}(\theta) := \log \bar{p}_\theta(Y_k|Y_{-m-s+1}^{k-1}, X_{-m}=x) \quad\text{and}\quad \Delta_{k,m}(\theta) := \log \bar{p}_\theta(Y_k|Y_{-m-s+1}^{k-1}).$$

With these notations, we have

$$\ell_n(\theta) = \sum_{k=1}^n \Delta_{k,0}(\theta) \quad\text{and}\quad \ell_n(\theta,x_0) = \sum_{k=1}^n \Delta_{k,0,x_0}(\theta). \tag{38}$$

Lemma 24. *(as [?, Lemma 3]) With the notation ρ introduced in Corollary 22, we have $\bar{\mathbb{P}}_{\theta^*}$-almost surely*

$$\forall m,m'\geq 0,\quad \sup_{\theta\in\Theta}\sup_{x,x'\in E} |\Delta_{k,m,x}(\theta)-\Delta_{k,m',x'}(\theta)| \leq \rho^{k+\min(m,m')-1}/(1-\rho) \tag{39}$$

$$\forall m\geq 0,\quad \sup_{\theta\in\Theta}\sup_{x\in E} |\Delta_{k,m,x}(\theta)-\Delta_{k,m}(\theta)| \leq \rho^{k+m-1}/(1-\rho) \tag{40}$$

$$\sup_\theta \sup_{m\geq 0}\sup_{x\in E} |\Delta_{k,m,x}(\theta)| \leq \max(|\log(p_{1,+}b_+(Y_{k-s}^k))|, |\log(p_{1,-}b_-(Y_{k-s}^k))|) \tag{41}$$

with

$$b_-(y_{k-s}^k) := \inf_\theta \int_E p_{2,\theta}(y_k|x, y_{k-s}^{k-1})\, dm_E(x)$$

and

$$b_+(y_{k-s}^k) := \sup_\theta \int_E p_{2,\theta}(y_k|x, y_{k-s}^{k-1})\, dm_E(x).$$

Proof. Assume that $m\leq m'$. We have

$$e^{\Delta_{k,m,x}(\theta)} = \int_{E^2} q_\theta(x_k, Y_k|x_{k-1}, Y_{k-s}^{k-1}) p_\theta(X_{k-1}=x_{k-1}|X_{-m}=x, Y_{-m-s+1}^{k-1})\, dm_E^{\otimes 2}(x_k,x_{k-1}).$$

Observe moreover that, due to Fact 21, we have

$$e^{\Delta_{k,m',x'}(\theta)} = \int_E e^{\Delta_{k,m,x''}(\theta)} p_\theta(X_{-m}=x''|X_{-m'}=x', Y_{-m'-s+1}^{k-1})\, dm_E(x'').$$

Therefore, according to Corollary 22, we obtain

$$\begin{aligned}\left|e^{\Delta_{k,m,x}(\theta)} - e^{\Delta_{k,m',x'}(\theta)}\right| &\leq \sup_{x"\in E} |e^{\Delta_{k,m,x}(\theta)} - e^{\Delta_{k,m,x"}(\theta)}| \\ &\leq p_{1,+}\rho^{k+m-1}\int_E p_{2,\theta}(Y_k|x_k, Y_{k-s}^{k-1})\, dm_E(x_k).\end{aligned}$$

Since

$$\left|e^{\Delta_{k,m,x}(\theta)}\right| \geq p_{1,-}\int_E p_{2,\theta}(Y_k|x_k, Y_{k-s}^{k-1})\, dm_E(x_k),$$

we get the first point. The proof of the second point follows exactly the same scheme with the use of the following formula

$$e^{\Delta_{k,m}(\theta)} = \int_E e^{\Delta_{k,m,x_{-m}}(\theta)} \bar{p}_\theta(X_{-m}=x_{-m}|Y_{-m-s+1}^{k-1})\, dm_E(x_{-m}).$$

The last point comes from the fact that

$$p_{1,-}\int_E p_{2,\theta}(Y_k|x_k, Y_{k-s}^{k-1})\, dm_E(x_k) \leq e^{\Delta_{k,m,x}(\theta)} \leq p_{1,+}\int_E p_{2,\theta}(Y_k|x_k, Y_{k-s}^{k-1})\, dm_E(x_k).$$

□

Due to (39), we get that, $\bar{\mathbb{P}}_{\theta^*}$-a.s., $(\Delta_{k,m,x}(\theta))_m$ is a (uniform in (k,x,θ)) Cauchy sequence and so converges uniformly in (k,x,θ) to some $\Delta_{k,\infty,x}(\theta)$.

Due to (39) and (40), $\Delta_{k,\infty,x}(\theta)$ does not depend on x and will be denoted by $\Delta_{k,\infty}(\theta)$. Moreover we have $\Delta_{k,\infty}(\theta) = \Delta_{0,\infty}(\theta) \circ \tau^k$.

Due to (41), (2), (3) and (4), $(\Delta_{k,m,x}(\theta))_{k,m,x}$ is uniformly bounded in $\mathbb{L}^1(\bar{\mathbb{P}}_{\theta^*})$. Therefore $\Delta_{k,\infty}(\theta)$ is in $\mathbb{L}^1(\bar{\mathbb{P}}_{\theta^*})$. Let us write

$$\ell(\theta) := \bar{\mathbb{E}}_{\theta^*}[\Delta_{0,\infty}(\theta)].$$

Since $(\Omega, \mathcal{F}, \bar{\mathbb{P}}_{\theta^*}, \tau)$ is ergodic, from the Birkhoff-Khinchine ergodic theorem, we have

$$\lim_{n\to+\infty} n^{-1} \sum_{k=1}^{n} \Delta_{k,\infty}(\theta) = \ell(\theta) \quad \bar{\mathbb{P}}_{\theta^*} - a.s. \text{ and in } \mathbb{L}^1(\bar{\mathbb{P}}_{\theta^*}). \tag{42}$$

Now, due to (39) and (40) applied with $m = 0$, we obtain

$$\sum_{k=1}^{n} \sup_{\theta} |\Delta_{k,0}(\theta) - \Delta_{k,\infty}(\theta)| \le \frac{2}{(1-\rho)^2} \quad \bar{\mathbb{P}}_{\theta^*} - a.s.. \tag{43}$$

Now, putting together (38), (40), (42) and (43), we have

Corollary 25.

$$\lim_{n\to+\infty} n^{-1}\ell_n(\theta, x_0) = \lim_{n\to+\infty} n^{-1}\ell_n(\theta) = \ell(\theta), \quad \bar{\mathbb{P}}_{\theta^*} - a.s..$$

Still following [**?**], we have the next lemma insuring the continuity of $\theta \mapsto \ell(\theta)$.

Lemma 26. *(as [**?**, Lemma 4]) For all $\theta \in \Theta$,*

$$\lim_{\delta\to 0} \bar{\mathbb{E}}_{\theta^*}[\sup_{|\theta-\theta'|\le\delta} |\Delta_{0,\infty}(\theta) - \Delta_{0,\infty}(\theta')|] = 0.$$

Proof. We recall that $\Delta_{0,\infty} = \lim_{m\to\infty} \Delta_{0,m,x}(\theta)$ (for every $x \in E$) with

$$\Delta_{0,m,x}(\theta) = \log \frac{\int_{E^m} \prod_{\ell=-m+1}^{0} q_\theta(x_\ell, Y_\ell | x_{\ell-1}, Y_{\ell-s}^{\ell-1})\, dm_E^{\otimes m}(x_{-m+1}^0)\, d\delta_x(x_{-m})}{\int_{E^{m-1}} \prod_{\ell=-m+1}^{-1} q_\theta(x_\ell, Y_\ell | x_{\ell-1}, Y_{\ell-s}^{\ell-1})\, dm_E^{\otimes(m-1)}(x_{-m+1}^{-1})\, d\delta_x(x_{-m})}.$$

Since the maps $\theta \mapsto q_\theta(x_\ell, y_\ell | x_{\ell-1}, y_{\ell-s}^{\ell-1}, y_\ell)$ are continuous, $\Delta_{0,m,x}$ is $\bar{\mathbb{P}}_{\theta^*}$-almost surely continuous. The uniform convergence result proved above insures that $\Delta_{0,\infty}$ is also $\bar{\mathbb{P}}_{\theta^*}$-almost surely continuous. Hence

$$\forall\theta, \ \lim_{\delta\to 0} \sup_{\theta':|\theta-\theta'|\le\delta} |\Delta_{0,\infty}(\theta) - \Delta_{0,\infty}(\theta')| = 0 \ \ \bar{\mathbb{P}}_{\theta^*} - a.s..$$

Now, the result follows from the Lebesgue dominated convergence theorem, due to (41), (2), (3) and (4). □

Lemma 27. *(as [**?**, Prop. 2]) We have*

$$\lim_{n\to+\infty} \sup_{\theta\in\Theta} |n^{-1}\ell_n(\theta, x_0) - \ell(\theta)| = 0, \quad \bar{\mathbb{P}}_{\theta^*} - a.s..$$

Lemma 27 can be deduced exactly as in the proof of [**?**, Prop. 2]. We do not rewrite the proof, but mention that it uses (36), the compacity of Θ, the continuity of ℓ, (43), the ergodicity of $\bar{\mathbb{P}}_{\theta^*}$ and Lemma 26.

Lemma 28. *(as [**?**, Lemma 5]) For every $k \le \ell$, we have*

$$\lim_{j\to-\infty} \sup_{i\le j} |\bar{p}_\theta(Y_k^\ell | Y_{i-s+1}^j) - \bar{p}_\theta(Y_k^\ell)| = 0 \ \ \textit{in } \bar{\mathbb{P}}_{\theta^*} - \textit{probability}.$$

Proof. Let us write $G(y_{-s}^0) := \int_E p_{2,\theta}(y_0|x, y_{-s}^{-1})\, dm_E(x)$ and $\tilde{G}(y_0) := \sup_{y_{-s}^{-1}} G(y_{-s}^0)$. As in the proof of [?, Lemma 5], we observe that, by stationarity, it is enough to prove that

$$\forall \ell > 0, \quad \lim_{k\to+\infty} \sup_{i\geq 0} \left|\bar{p}_\theta(Y_k^{k+\ell}|Y_{-i-s+1}^0) - \bar{p}_\theta(Y_k^{k+\ell})\right| = 0 \ \ in\ \bar{\mathbb{P}}_{\theta^*} - probability$$

and we write

$$\left|\bar{p}_\theta(Y_k^{k+\ell}|Y_{-i-s+1}^0) - \bar{p}_\theta(Y_k^{k+\ell})\right| = \left|\int_{E^2\times K^{2s}} A_k(B_k' - B_k'')C_i\, dm_E^{\otimes 2}(x_s, x_{k-1}) dm_K^{\otimes 2s}(y_1^s, y_{k-s}^{k-1})\right|,$$

with

$$A_k := p_\theta(Y_k^{k+\ell}|X_{k-1} = x_{k-1}, Y_{k-s}^{k-1} = y_{k-s}^{k-1}) \leq \tilde{A}_k := p_{1,+}^{\ell+1} \prod_{j=k+s}^{k+\ell} G(Y_{j-s}^j) \prod_{j=k}^{k+s-1} \tilde{G}(Y_j),$$

(due to (34) and to (2)) with

$$B_k' := p_\theta(X_{k-1} = x_{k-1}, Y_{k-s}^{k-1} = y_{k-s}^{k-1}|X_s = x_s, Y_1^s = y_1^s) = Q_\theta^{*(k-s-1)}(x_{k-1}, y_{k-s}^{k-1}|x_s, y_1^s),$$

with

$$B_k'' := \bar{p}_\theta(X_{k-1} = x_{k-1}, Y_{k-s}^{k-1} = y_{k-s}^{k-1}) = h_\theta(x_{k-1}, y_{k-s}^{k-1})$$

and with

$$C_i := \bar{p}_\theta(X_s = x_s, Y_1^s = y_1^s|Y_{-i-s+1}^0).$$

Let us write

$$B_k := \int_{E\times K^s} |B_k' - B_k''|\, d\mu(x_{k-1}, y_{k-s}^{k-1}).$$

We have

$$\left|\bar{p}_\theta(Y_k^{k+\ell}|Y_{i-s+1}^0) - \bar{p}_\theta(Y_k^{k+\ell})\right| \leq \tilde{A}_k \int_{E\times K^s} B_k C_i\, d\mu(x_s, y_1^s).$$

On the one hand, due to (6), $B_k = B_k(x_s, y_1^s)$ converges to 0 as k goes to infinity, for μ-almost every (x_s, y_1^s) (and this quantity is bounded by 1). On the other hand, on $\{Y_{-i-s+1}^0 = y_{-i-s+1}^0\}$, we have

$$\begin{aligned} C_i &= \int_{E^s} \prod_{j=1}^s q_\theta(x_j, y_j|x_{j-1}, y_{j-s}^{j-1}) \bar{p}_\theta(X_0 = x_0|Y_{-i-s+1}^0 = y_{-i-s+1}^0)\, dm_E^{\otimes s}(x_0^{s-1}) \\ &\leq p_{1,+} H(x_s, y_{-s+1}^s), \end{aligned}$$

with

$$H(x_s, y_{-s+1}^s) := \int_{E^{s-1}} \prod_{j=2}^s p_{1,\theta}(x_j|x_{j-1}, y_{j-1}) \prod_{j=1}^s p_{2,\theta}(y_j|x_j, y_{j-s}^{j-1})\, dm_E^{\otimes s}(x_1^{s-1})$$

and

$$\forall y_{-s+1}^0, \quad \int_{E\times K^s} H(x_s, y_{-s+1}^s)\, d\mu(x_s, y_1^s) = 1.$$

Therefore, by the Lebesgue dominated convergence theorem, we obtain

$$\lim_{k\to+\infty} \sup_{i\leq 0} \int_{E\times K^s} B_k C_i\, d\mu(x_s, y_1^s) = 0 \ \ \bar{\mathbb{P}}_{\theta^*} - a.s..$$

Of course, this convergence also holds in $\bar{\mathbb{P}}_{\theta^*}$-probability. Now, since, for every k, $\tilde{A}_k$ is a real valued random variable (see (5)) with the same distribution as $p_{1,+}^{\ell+1} \prod_{j=s}^{\ell} G(Y_{j-s}^j) \prod_{j=0}^{s-1} \tilde{G}(Y_j)$, we obtain the result. □

Lemma 29. *([?, Lem. 6 & 7, Prop. 3]) For every $\theta \in \Theta$, $\ell(\theta) \leq \ell(\theta^*)$. Furthermore*

$$\ell(\theta) = \ell(\theta^*) \quad \Rightarrow \quad \bar{\mathbb{P}}_\theta^Y = \bar{\mathbb{P}}_{\theta^*}^Y.$$

Elements of the proof. We do not rewrite the proof of this lemma, the reader can follow the proofs of [?, Lem. 6-7, Prop. 3] (using Lemma 28 and Kullback-Leibler divergence functions). The only adaptations to make concern the proof of [?, Lem. 7] which, due to our slightly weaker hypothesis (5), are the following facts. Following the proof of Lemma 28, observe that, due to (2), (34) and (33), on $\{Y_{-s+1}^p = y_{-s+1}^p,\ Y_{-m-s+1}^{-k} = y_{-m-s+1}^{-k}\}$, $\bar{p}_\theta(Y_{-s+1}^p | Y_{-m-s+1}^{-k})$ is between

$$p_{1,-}^{p+s} \int_{E\times K^s} \prod_{j=-s+1}^{p} G(y_{j-s}^j) \bar{p}_\theta(X_{-s} = x_{-s}, Y_{-2s+1}^{-s} = y_{-2s+1}^{-s} | Y_{-m-s+1}^{-k})\, d\mu(x_{-s}, y_{-2s+1}^{-s})$$

and

$$p_{1,+}^{p+s} \int_{E\times K^s} \prod_{j=-s+1}^{p} G(y_{j-s}^j) \bar{p}_\theta(X_{-s} = x_{-s}, Y_{-2s+1}^{-s} = y_{-2s+1}^{-s} | Y_{-m-s+1}^{-k})\, d\mu(x_{-s}, y_{-2s+1}^{-s}),$$

with $G(y_{-s}^0) := \int_E p_{2,\theta}(y_0|x, y_{-s}^{-1})\, dm_E(x)$. Therefore we have

$$\frac{p_{1,-}^{p+s}}{p_{1,+}^s} \prod_{j=1}^{p} G(Y_{j-s}^j) \le \bar{p}_\theta(Y_1^p | Y_{-s+1}^0, Y_{-m-s+1}^{-k}) = \frac{\bar{p}_\theta(Y_{-s+1}^p | Y_{-m-s+1}^{-k})}{\bar{p}_\theta(Y_{-s+1}^0 | Y_{-m-s+1}^{-k})} \le \frac{p_{1,+}^{p+s}}{p_{1,-}^s} \prod_{j=1}^{p} G(Y_{j-s}^j).$$

Due to (3) and (4), we obtain

$$\bar{\mathbb{E}}_{\theta^*}\left[\sup_k \sup_{m\ge k} |\log(\bar{p}_\theta(Y_1^p | Y_{-s+1}^0, Y_{-m-s+1}^{-k}))|\right] < \infty,$$

which enables the adaptation of the proof of [?, Lem. 7]. □

Proof of Theorem 2. Let $x_0 \in E$. We know that, $\bar{\mathbb{P}}_{\theta^*}$-almost surely, $(n^{-1}\ell_n(\cdot, x_0))_n$ converges uniformly to ℓ which admits a maximum $\ell(\theta^*)$. Since $\ell_n(\cdot, x_0)$ is continuous on Θ and since Θ is compact, $\hat{\theta}_{n,x_0}$ is well defined. Moreover, the limit values of $(\hat{\theta}_{n,x_0})_n$ are contained in

$$\{\theta \in \Theta \ :\ \ell(\theta) = \ell(\theta^*)\} \subseteq \{\theta \in \Theta \ :\ \bar{\mathbb{P}}_\theta^Y = \bar{\mathbb{P}}_{\theta^*}^Y\}.$$

Assume now that Q_{θ^*} is aperiodic and positive Harris recurrent, following the proof of [?, Thm. 5], we have $\lim_{n\to+\infty} \ell(\hat{\theta}_{n,x_0}) = \ell(\theta^*)$ almost surely for any initial measure and we conclude as above. □

B Identifiability for the von Mises model: proof of proposition 13

Assume that $\bar{\mathbb{P}}_{\theta_1}^Y = \bar{\mathbb{P}}_{\theta_2}^Y$. In particular, we have

$$\bar{p}_{\theta_1}(Y_k = y_k | Y_{k-s}^{k-1} = y_{k-s}^{k-1}) = \bar{p}_{\theta_2}(Y_k = y_k | Y_{k-s}^{k-1} = y_{k-s}^{k-1}),\ \text{for}\ \bar{\mathbb{P}}_{\theta_1}^{Y_{k-s}^k} - a.e.\ y_{k-s}^k$$

and thus

$$\sum_{x=1}^{M} \bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) p_{2,\theta_1}(y_k | x, y_{k-s}^{k-1}) = \sum_{x=1}^{M} \bar{\mathbb{P}}_{\theta_2}(X_k = x | y_{k-s}^{k-1}) p_{2,\theta_2}(y_k | x, y_{k-s}^{k-1}),\ \text{for}\ \bar{\mathbb{P}}_{\theta_1}^{Y_{k-s}^k} - a.e.\ y_{k-s}^k.$$

Since $\bar{p}_{\theta_1}(y_{k-s}^k) > 0$ (the invariant pdf h_1 satisfies $h_1 > 0$ and the transition pdf q_θ satisfies $q_\theta > 0$ by construction) and due to (30), we deduce that, for $m^{\otimes(s+1)}$-a.e. y_{k-s}^k, we have

$$\sum_{x=1}^{M} \bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) f_{\gamma_{0,(1)}^{(x)} + \sum_{\ell=1}^s \gamma_{\ell,(1)}^{(x)} e^{i y_{k-\ell}}}(y_k) = \sum_{x=1}^{M} \bar{\mathbb{P}}_{\theta_2}(X_k = x | y_{k-s}^{k-1}) f_{\gamma_{0,(2)}^{(x)} + \sum_{\ell=1}^s \gamma_{\ell,(2)}^{(x)} e^{i y_{k-\ell}}}(y_k)$$

with f_γ defined by (29) where m denotes the Lebesgue measure on $\mathbb{T}$.

According to [?], finite mixtures of von Mises distribution are identifiable. This implies in particular that if

$$\sum_{x=1}^{M} \pi_1^{(x)} f_{\gamma_1^{(x)}}(y) = \sum_{x=1}^{M} \pi_2^{(x)} f_{\gamma_2^{(x)}}(y) \text{ for } m-a.e.\ y$$

with $\gamma_1^{(x)} \neq \gamma_1^{(x')}$ for $x \neq x'$ and $\pi_1^{(x)} > 0$ for $x \in \{1, ..., M\}$ then there exists a permutation $\tau : \{1, ..., M\} \to \{1, ..., M\}$ such that $\gamma_1^{(x)} = \gamma_2^{(\tau(x))}$ and $\pi_1^{(x)} = \pi_2^{(\tau(x))}$.

Recall that we have assumed that $(\gamma_{0,(1)}^{(x)}, ..., \gamma_{s,(1)}^{(x)}) \neq (\gamma_{0,(1)}^{(x')}, ..., \gamma_{s,(1)}^{(x')})$ if $x \neq x'$, which implies that

$$\gamma_{0,(1)}^{(x)} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x)} e^{iy_{k-\ell}} \neq \gamma_{0,(1)}^{(x')} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x')} e^{iy_{k-\ell}}, \quad \text{for } m^{\otimes(s+1)} - a.e.\ y_{k-s}^{k}.$$

Therefore, since for every $x \in \{1, ..., M\}$ and for $m^{\otimes s}$-almost every y_{k-s}^{k-1}, $\bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) > 0$ (since $h_{\theta_1} > 0$), for $m^{\otimes s}$-almost every y_{k-s}^{k-1} there exists a permutation $\tau_{y_{k-s}^{k-1}}$ of $\{1, ..., M\}$ such that,

$$\forall x \in \{1, ..., M\}, \quad \gamma_{0,(1)}^{(x)} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x)} e^{iy_{k-\ell}} = \gamma_{0,(2)}^{(\tau_{y_{k-s}^{k-1}}(x))} + \sum_{\ell=1}^{s} \gamma_{\ell,(2)}^{(\tau_{y_{k-s}^{k-1}}(x))} e^{iy_{k-\ell}}.$$

Since the set of permutations of $\{1, ..., M\}$ is finite, there exists a positive Lebesgue measure subset of $\mathbb{T}^s$ on which the permutation is the same permutation τ. From this, we deduce that

$$\forall x \in \{1, ..., M\}, \ \forall j \in \{0, ..., s\}, \quad \gamma_{j,(1)}^{(x)} = \gamma_{j,(2)}^{(\tau(x))}$$

and that, for Lebesgue almost every y_{k-s}^{k+1}, the following holds true

$$\forall x \in \{1, ..., M\}, \bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) = \bar{\mathbb{P}}_{\theta_2}(X_k = \tau(x) | y_{k-s}^{k-1}).$$

Let us now discuss the identifiability of the other components of θ_1 and θ_2. If $\bar{\mathbb{P}}_{\theta_1}^{Y} = \bar{\mathbb{P}}_{\theta_2}^{Y}$ then

$$\bar{p}_{\theta_1}(Y_k = y_k, Y_{k+1} = y_{k+1} | Y_{k-s}^{k-1} = y_{k-s}^{k-1}) = \bar{p}_{\theta_2}(Y_k = y_k, Y_{k+1} = y_{k+1} | Y_{k-s}^{k-1} = y_{k-s}^{k-1}) \ \ \bar{\mathbb{P}}_{\theta_1}^{Y_{k-s}^{k+1}} - a.e.\ y_{k-s}^{k+1}$$

and thus, for Lebesgue almost every y_{k-s}^{k+1}, we have

$$\begin{aligned}
&\sum_{x,x'=1}^{M} \bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) p_{1,\theta_1}(x' | x, y_k) f_{\gamma_{0,(1)}^{(x)} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x)} e^{iy_{k-\ell}}}(y_k) f_{\gamma_{0,(1)}^{(x')} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x')} e^{iy_{k-\ell+1}}}(y_{k+1}) \\
= &\sum_{x,x'=1}^{M} \bar{\mathbb{P}}_{\theta_2}(X_k = x | y_{k-s}^{k-1}) p_{1,\theta_2}(x' | x, y_k) f_{\gamma_{0,(2)}^{(x)} + \sum_{\ell=1}^{s} \gamma_{\ell,(2)}^{(x)} e^{iy_{k-\ell}}}(y_k) f_{\gamma_{0,(2)}^{(x')} + \sum_{\ell=1}^{s} \gamma_{\ell,(2)}^{(x')} e^{iy_{k-\ell+1}}}(y_{k+1}).
\end{aligned}$$

This implies that, for almost every y_{k-s}^{k+1}, the quantity

$$\sum_{x,x'} \bar{\mathbb{P}}_{\theta_1}(X_k = x | y_{k-s}^{k-1}) (p_{1,\theta_1}(x' | x, y_k) - p_{1,\theta_2}(\tau(x') | \tau(x), y_k)) f_{\gamma_{0,(1)}^{(x)} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x)} e^{iy_{k-\ell}}}(y_k) f_{\gamma_{0,(1)}^{(x')} + \sum_{\ell=1}^{s} \gamma_{\ell,(1)}^{(x')} e^{iy_{k-\ell+1}}}(y_{k+1})$$

is null and so (again using the identifiability of von Mises distribution)

$$\forall x, x', \quad p_{1,\theta_1}(x' | x, y) = p_{1,\theta_2}(\tau(x') | \tau(x), y) \text{ for } m-a.e.\ y.$$

Now, due to the special form of $p_{1,\theta}$ specified in (14), we get

$$\forall x, x', \quad m-a.e.\ y, \quad \frac{q_{x,x',(1)} \left| \exp\left(\tilde{\lambda}_{x,x',(1)} e^{-iy}\right) \right|}{\sum_{x''=1}^{M} q_{x,x'',(1)} \left| \exp\left(\tilde{\lambda}_{x,x'',(1)} e^{-iy}\right) \right|} = \frac{q_{\tau(x),\tau(x'),(2)} \left| \exp\left(\tilde{\lambda}_{\tau(x),\tau(x'),(2)} e^{-iy}\right) \right|}{\sum_{x''=1}^{M} q_{\tau(x),x'',(2)} \left| \exp\left(\tilde{\lambda}_{\tau(x),x'',(2)} e^{-iy}\right) \right|}. \tag{44}$$

Let $x \in \{1,...,M\}$ be fixed. Applying (44) a first time with $x' = x$ and a second time with any x', we get

$$\forall x', \quad \text{for } m-a.e.\ y, \quad \frac{q_{x,x',(1)}\left|\exp\left(\tilde{\lambda}_{x,x',(1)}e^{-iy}\right)\right|}{q_{x,x,(1)}\left|\exp\left(\tilde{\lambda}_{x,x,(1)}e^{-iy}\right)\right|} = \frac{q_{\tau(x),\tau(x'),(2)}\left|\exp\left(\tilde{\lambda}_{\tau(x),\tau(x'),(2)}e^{-iy}\right)\right|}{q_{\tau(x),\tau(x),(2)}\left|\exp\left(\tilde{\lambda}_{\tau(x),\tau(x),(2)}e^{-iy}\right)\right|}$$

and so

$$\forall x', \quad \frac{q_{x,x',(1)}}{q_{x,x,(1)}} = \frac{q_{\tau(x),\tau(x'),(2)}}{q_{\tau(x),\tau(x),(2)}} \tag{45}$$

and

$$\forall x', \quad \tilde{\lambda}_{x,x',(1)} - \tilde{\lambda}_{x,x,(1)} = \tilde{\lambda}_{\tau(x),\tau(x'),(2)} - \tilde{\lambda}_{\tau(x),\tau(x),(2)}. \tag{46}$$

Now, since $\sum_{x'} q_{x,x',(1)} = 1 = \sum_{x'} q_{\tau(x),\tau(x'),(2)}$, due to (45), it comes $q_{x,x,(1)} = q_{\tau(x),\tau(x),(2)}$ and so

$$\forall x' \in E, \quad q_{x,x',(1)} = q_{\tau(x),\tau(x'),(2)}.$$

If θ_1 and θ_2 are in Θ', since $\tilde{\lambda}_{x,x,(1)} = 0 = \tilde{\lambda}_{\tau(x),\tau(x),(2)}$, due to (46), we conclude that

$$\forall x' \in E, \quad \tilde{\lambda}_{x,x',(1)} = \tilde{\lambda}_{\tau(x),\tau(x'),(2)}.$$

If θ_1 and θ_2 are in Θ'', since $\sum_{x'} \tilde{\lambda}_{x,x',(1)} = 0 = \sum_{x'} \tilde{\lambda}_{\tau(x),\tau(x'),(2)}$, due to (46), we get $\tilde{\lambda}_{x,x,(1)} = \tilde{\lambda}_{\tau(x),\tau(x),(2)}$ and, applying again (46), we conclude that

$$\forall x' \in E, \quad \tilde{\lambda}_{x,x',(1)} = \tilde{\lambda}_{\tau(x),\tau(x'),(2)}.$$

□

C Identifiability for the gaussian model: proof of Proposition 9

Using similar arguments as in Appendix B, existing results on the identifiability of mixture of Gaussian distributions (see [?]) we obtain that if $\bar{\mathbb{P}}^Y_{\theta_1} = \bar{\mathbb{P}}^Y_{\theta_2}$, then for all $x \in \{1,2\}$ and $y \in \mathbb{R}$

$$\left(\beta^{(x)}_{0,(1)}, \beta^{(x)}_{1,(1)}, ..., \beta^{(x)}_{r,(1)}, \sigma^{(x)}_{(1)}\right) = \left(\beta^{(x)}_{0,(2)}, \beta^{(x)}_{1,(2)}, ..., \beta^{(x)}_{r,(2)}, \sigma^{(x)}_{(2)}\right)$$

and

$$p_{1,\theta_1}(x|x,y) = \pi^{(x)}_{-,(1)} + \frac{1-\pi^{(x)}_{-,(1)} - \pi^{(x)}_{+,(1)}}{1+exp\left(\lambda^{(x)}_{0,(1)} + \lambda^{(x)}_{1,(1)}y\right)} = \pi^{(x)}_{-,(2)} + \frac{1-\pi^{(x)}_{-,(2)} - \pi^{(x)}_{+,(2)}}{1+exp\left(\lambda^{(x)}_{0,(2)} + \lambda^{(x)}_{1,(2)}y\right)} = p_{1,\theta_2}(x|x,y) \tag{47}$$

where the regimes have been labeled such that the permutation τ is the identity.

If θ_1 and θ_2 are in Θ' then $\lambda^{(x)}_{1,(i)} \neq 0$ for $i \in \{1,2\}$ and looking at the asymptotic behavior of the terms which appear in (47) when $y \to \pm\infty$ permits to show that $\pi^{(x)}_{-,(1)} = \pi^{(x)}_{-,(2)}$, $\pi^{(x)}_{+,(1)} = \pi^{(x)}_{+,(2)}$. We can then easily deduce that $\lambda^{(x)}_{0,(1)} = \lambda^{(x)}_{0,(2)}$ and $\lambda^{(x)}_{1,(1)} = \lambda^{(x)}_{1,(2)}$ and thus that $\theta_1 = \theta_2$.

If θ_1 and θ_2 are in Θ'', then we directly obtain that $\pi^{(x)}_{-,(1)} = \pi^{(x)}_{-,(2)} = \pi^{(x)}_{+,(1)} = \pi^{(x)}_{-,(1)} = \pi_0$ and then that $\theta_1 = \theta_2$. □

D Identifiability for the Rainfall model: proof of Proposition 19

Assume that $\bar{\mathbb{P}}^Y_{\theta_1} = \bar{\mathbb{P}}^Y_{\theta_2}$. First, we use the fact that

$$\bar{p}_{\theta_1}(Y_k = y_k | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) = \bar{p}_{\theta_2}(Y_k = y_k | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \ \text{ for } \ \bar{\mathbb{P}}^{Y^k_{k-s}}_{\theta_1} - a.e.\ y^k_{k-s} \tag{48}$$

to prove that

$$(\pi^{(x)}_{i,(1)}, \alpha^{(x)}_{i,(1)}, \beta^{(x)}_{i,(1)})_{i,x} = (\pi^{(x)}_{i,(2)}, \alpha^{(x)}_{i,(2)}, \beta^{(x)}_{i,(2)})_{i,x}.$$

Using (48) on the set $\{r^{(i)}_k > 0,\ \forall i \in \{1, ..., \ell\}\}$, we conclude that there exists a permutation τ of $\{1, ..., M\}$ such that, for every $i \in \{1, ..., \ell\}$ and every $x \in \{1, ..., M\}$, we have

$$(\alpha^{(x)}_{i,(1)}, \beta^{(x)}_{i,(1)}) = (\alpha^{(\tau(x))}_{i,(2)}, \beta^{(\tau(x))}_{i,(2)}) \tag{49}$$

and

$$\bar{\mathbb{P}}_{\theta_1}(X_k = x | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \prod_{i=1}^{\ell} \pi^{(x)}_{i,(1)} = \bar{\mathbb{P}}_{\theta_2}(X_k = x | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \prod_{i=1}^{\ell} \pi^{(\tau(x))}_{i,(2)}.$$

Now, for every $J \subseteq \{1, ..., \ell\}$, we use (48) on the set $\{r^{(j)}_k > 0,\ \forall j \in J,\ r^{(i)}_k = 0,\ \forall i \notin J\}$. Due to (49) and since θ_1 satisfies (25), we obtain

$$\bar{\mathbb{P}}_{\theta_1}(X_k = x | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \prod_{j \in J} \pi^{(x)}_{j,(1)} \prod_{i \notin J} (1 - \pi^{(x)}_{i,(1)}) = \bar{\mathbb{P}}_{\theta_2}(X_k = x | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \prod_{j \in J} \pi^{(\tau(x))}_{j,(2)} \prod_{i \notin J} (1 - \pi^{(\tau(x))}_{i,(2)}).$$

From which, we conclude

$$\forall i \in \{1, ..., \ell\},\ \forall x \in \{1, ..., M\},\quad \pi^{(x)}_{i,(1)} = \pi^{(\tau(x))}_{i,(2)}. \tag{50}$$

Now it remains to prove that $(q_{x,x',(1)}, \mu_{x,x',(1)}) = (q_{\tau(x),\tau(x'),(2)}, \mu_{\tau(x),\tau(x'),(2)})$. To this hand, as for the von Mises model (see Appendix B), we use the fact that

$$\bar{p}_{\theta_1}(Y_k = y_k, Y_{k+1} = y_{k+1} | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) = \bar{p}_{\theta_2}(Y_k = y_k, Y_{k+1} = y_{k+1} | Y^{k-1}_{k-s} = y^{k-1}_{k-s}) \text{ for } \bar{\mathbb{P}}^{Y^k_{k-s}}_{\theta_1} - a.e.\ y^k_{k-s} \tag{51}$$

and obtain that

$$\forall x, x',\quad p_{1,\theta_1}(x' | x, y_k) = p_{1,\theta_2}(\tau(x') | \tau(x), y_k) \text{ for } a.e. y_k.$$

This implies that

$$\frac{\tilde{q}_{x,x',(1)} \exp(-z'_{k-1} \tilde{\mu}_{x,x',(1)})}{\sum_{x"} \tilde{q}_{x,x",(1)} \exp(-z'_{k-1} \tilde{\mu}_{x,x',(1)})} = \frac{\tilde{q}_{\tau(x),\tau(x'),(2)} \exp(-z'_{k-1} \tilde{\mu}_{\tau(x),\tau(x'),(2)})}{\sum_{x"} \tilde{q}_{x,\tau(x"),(2)} \exp(-z'_{k-1} \tilde{\mu}_{x,\tau(x"),(2)})}, \tag{52}$$

with $\tilde{q}_{x,x',(j)} := q_{x,x',(j)} \exp(-\frac{1}{2}(\mu_{x,x',(j)})' \Sigma^{-1} \mu_{x,x',(j)})$ and $\tilde{\mu}_{x,x',(j)} := \Sigma^{-1} \mu_{x,x',(j)}$. From (52), we obtain that

$$\frac{\tilde{q}_{x,x',(1)} \exp(-z'_{k-1} \tilde{\mu}_{x,x',(1)})}{\tilde{q}_{x,x,(1)} \exp(-z'_{k-1} \tilde{\mu}_{x,x,(1)})} = \frac{\tilde{q}_{\tau(x),\tau(x'),(2)} \exp(-z'_{k-1} \tilde{\mu}_{\tau(x),\tau(x'),(2)})}{\tilde{q}_{\tau(x),\tau(x),(2)} \exp(-z'_{k-1} \tilde{\mu}_{\tau(x),\tau(x),(2)})},$$

and so that, for every $x, x' \in \{1, ..., M\}$,

$$\tilde{\mu}_{x,x',(1)} - \tilde{\mu}_{x,x,(1)} = \tilde{\mu}_{\tau(x),\tau(x'),(2)} - \tilde{\mu}_{\tau(x),\tau(x),(2)} \tag{53}$$

and

$$\frac{\tilde{q}_{x,x',(1)}}{\tilde{q}_{x,x,(1)}} = \frac{\tilde{q}_{\tau(x),\tau(x'),(2)}}{\tilde{q}_{\tau(x),\tau(x),(2)}}. \tag{54}$$

Finally, it comes from (53) that $\tilde{\mu}_{x,x',(1)} = \tilde{\mu}_{\tau(x),\tau(x'),(2)}$ (using $\sum_{x"} \tilde{\mu}_{x,x",(j)} = 0$) and so $\mu_{x,x',(1)} = \mu_{\tau(x),\tau(x'),(2)}$. So (54) becomes

$$\frac{q_{x,x',(1)}}{q_{x,x,(1)}} = \frac{q_{\tau(x),\tau(x'),(2)}}{q_{\tau(x),\tau(x),(2)}}$$

which implies that $q_{x,x',(1)} = q_{\tau(x),\tau(x'),(2)}$ (due to $\sum_{x"} q_{x,x",(j)} = 1$). $\square$

E Proof of Lemma 16

Let f be any probability pdf wrt $\mu = m_E \times m_K$. We have

$$\begin{aligned}[][Q_\theta^{*n}(f-h_\theta)](x_0,y_0) &= \int_{(E\times K)^n} \prod_{i=-n+1}^{0} q_\theta(x_i,y_i|x_{i-1})(f-h_\theta)(x_{-n},y_{-n})\, dm_E^{\otimes n}(x_{-n}^{-1})dm_K^{\otimes n}(y_{-n}^{-1}) \\ &= \int_{E^n\times K^{n-1}} \prod_{i=-n+1}^{0} q_\theta(x_i,y_i|x_{i-1})(F-h_{1,\theta})(x_{-n})\, dm_E^{\otimes n}(x_{-n}^{-1})dm_K^{\otimes(n-1)}(y_{-n+1}^{-1})\end{aligned}$$

with $F(x_{-n}) := \int_K f(x_{-n},y_{-n})\, dm_K(y_{-n})$. Now, since $q_\theta(x_i,y_i|x_{i-1}) = p_{1,\theta}(x_i|x_{i-1})p_{2,\theta}(y_i|x_i)$, we obtain that

$$[Q_\theta^{*n}(f-h_{1,\theta})](x_0,y_0) = p_{2,\theta}(y_0|x_0)\int_{E^n} \prod_{i=-n+1}^{0} p_{1,\theta}(x_i|x_{i-1})(F-h_{1,\theta})(x_{-n})\, dm_E^{\otimes n}(x_{-n}^{-1}).$$

Therefore

$$||Q_\theta^{*n}(f-h_\theta)||_{L^1(m_E\times m_K)} = ||Q_{1,\theta}^{*n}(F-h_{1,\theta})||_{L^1(m_E)}.$$

Now, let us assume that $p_{2,\theta} > 0$ and that $(X_k)_k$ is an aperiodic positive Harris recurrent Markov chain. We will use the notations of [**?**].

Since $(X_k)_k$ is positive, it is ψ-irreducible (with $\psi = \psi_0$). Due to the hypothesis on $p_{2,\theta}$, this implies the ψ-irreducibility of $(X_k,Y_k)_k$ (with $\psi = \psi_0 \times m_K$).

Moreover $(X_k,Y_k)_k$ is positive since it admits an invariant probability measure (due to the first point of this result).

The fact that $(X_k)_k$ is aperiodic means that, for every ν_M-small set C such that $\nu_M(C) > 0$ for $(X_k)_k$, the greatest common divisor of the set E_C defined as follows is equal to 1:

$$E_C := \{n \geq 1 \ : \ C \text{ is } \nu_n - \text{small with } \nu_n = \delta_n\nu_M \text{ and } \delta_n > 0\}.$$

Now, let C' be a ν'_M-small set for $(X_k,Y_k)_k$ with $\nu'_M(C') > 0$, then for every $(x_0,y_0) \in C'$ and every $(B,D) \in \mathcal{B}(E)\times\mathcal{B}(K)$, we have $Q_\theta^M \mathbb{1}_{B\times D}(x_0,y_0) \geq \nu'_M(B\times D)$. Moreover $Q_\theta^M \mathbb{1}_{B\times D}(x_0,y_0)$ is equal to

$$\int_{E^{M-1}} \left(\int_B \prod_{i=1}^{M} p_{1,\theta}(x_i|x_{i-1}) \left(\int_D p_{2,\theta}(y_M|x_M)\, dm_K(y_M) \right) dm_E(x_M) \right) dm_E^{\otimes(M-1)}(x_1^{M-1}).$$

Since $Q_\theta^M \mathbb{1}_{B\times D}(x_0,y_0)$ does not depend on y_0, we obtain

$$\forall (x_0,y_0) \in E\times K,\ \forall B \in \mathcal{B}(E),\quad Q_{1,\theta}^M \mathbb{1}_B(x_0) = Q_\theta^M \mathbb{1}_{B\times K}(x_0,y_0) \geq \nu'_M(B\times K)$$

and so $C := \{x \in E \ : \ \exists y \in K,\ (x,y) \in C'\}$ is ν_M-small with $\nu_M(B) = \nu'_M(B\times K)$ and $\nu_M(C) \geq \nu'_M(C') > 0$. Moreover $E_C = E_{C'}$. Indeed, if C' is ν'_n-small with $\nu'_n = \delta'_n\nu'_M$, then C is ν_n-small with $\nu_n(B) = \nu'_n(B\times K) = \delta_n\nu_M(B)$ with $\delta_n(x) = \int_K \delta'_n(x,y)\, dm_K(y)$; and conversely, if C is ν_n-small with $\nu_n = \delta_n\nu_M$, then C' is ν'_n-small with $\nu'_n(B\times D) = \delta'_n\nu'_M(B\times D)$ and with $\delta'_n(x,y) = \delta_n(x)p_{2,\theta}(y|x)$. Therefore $(X_k,Y_k)_k$ is also aperiodic.

Finally, the Harris recurrence property of $(X_k,Y_k)_k$ follows from the Harris-recurrence of $(X_k)_k$ and from $p_{2,\theta} > 0$. □