Asymptotic properties of functional maximum-likelihood ARH parameter estimators

M.D. Ruiz-Medina and R. Salmerón

Department of Statistics and Operation Research, University of Granada
Department of Quantitative Methods for the Economy and Enterprise
mruiz@ugr.es
romansg@ugr.es
http://www.ugr.es/ mruiz/

Résumé. Dans cet article, nous nous intéressons au problème du calcul de la variance asymptotique des estimateurs du maximum de vraisemblance des paramètres de projection de processus autorégressifs hilbertiens. Ces estimateurs obtenus à partir d'observations fonctionnelles gaussiennes incomplètes ont été étudiés dans Ruiz-Medina et Salmerón (2010). Notre approche consiste à estimer ces variances à l'aide d'une version fonctionnelle de l'algorithme SEM (Supplemented Expectation Maximization) par Meng et Rubin (1991). La mise en oeuvre de l'algorithme est basée sur l'extraction des fonctions propres et valeurs propres de l'opérateur d'autocorrélation et de son adjoint. Les effets de l'ordre de troncature du spectre et de la discrétisation des fonctions sont illustrés par des simulations.

1 Introduction

Functional Statistics (see, for example, Bosq, 2000 ; Bosq and Blanke, 2007 ; Ferraty and Vieu, 2006 ; Ramsay and Silverman, 2005) provides a suitable framework to analyze large dimensional data sets. Complex biological and artificial systems can then be studied from this perspective. In particular, in Biomedicine, Bioinformatics and image processing (see, for instance, Germain *et al.*, 1999 ; Haoudi and Bensmail, 2006 ; Hyndman and Ullah, 2006 ; Leng and Müller, 2006 ; Monk, 2003 ; Song *et al.*, 2007, among others), the most extended projection estimation methodology has been Functional Principal Component Analysis (FPCA), implemented from the spectral decomposition of the empirical covariance operator, which is usually computed by application of the method of moments. However, the ML (Maximum Likelihood) estimation methodology has not been considered in these applications. In Ruiz-Medina and Salmerón (2010), ML projection estimators are computed from the application of the forward and backward Kalman recursion, combined with EM algorithm (see, for example, Hartley, 1958 ; Dempster, Laird and Rubin, 1977), in terms of the eigenfunction bases of the autocorrelation operator and its adjoint. This projection methodology is most suitable than FPCA in the case of ML estimation. The motivation of our paper lies on this fact, since, for the application of the ML projection estimation methodology, information on its performance is obviously needed. Specifically, the asymptotic variance of the ML projection estimators must be computed. The

sensitivity of the ML projection estimates to the truncation order and discretization level must be analyzed. Theses two issues will be addressed in this paper.

Within the current literature on functional projection estimation, we will cite the papers by Abramovich and Angelini (2006), and Abramovich, Sapatinas and Vidakovic (2004) on the derivation of wavelet-based projection estimators in the framework of FANOVA models. Kernel-based non-parametric projection estimators are studied in the context of functional regression models in Cardot *et al.* (2003). Also, we refer to Ramsay and Silverman (2005) on FPCA-based projection estimators, in the context of functional linear models. The asymptotic properties of functional moment estimators and FPCA-based projection estimators for ARH processes are studied in Bosq (2000, 2008), Bosq and Blanke (2007) and Mas (1999 ; 2007), among others. Wavelet-based autoregressive projection estimators are considered in Antoniadis and Sapatinas (2003). A comparative study of autoregressive estimators and non-parametric kernel-based estimators (e.g. local autoregressive estimators) is developed in Besse, Cardot and Stephenson (2000). In all the cited references, ML estimation and, in particular, ML projection estimators have not been studied. In particular, as commented before, in the context of ARH processes, ML projection estimators, based on the spectral decomposition of the autocorrelation operator, have been derived in Ruiz-Medina and Salmerón (2010). From such a spectral decomposition, the autoregressive prediction had been previously addressed in Ruiz-Medina, Salmerón and Angulo (2007) and Salmerón and Ruiz-Medina (2009). Specifically, the right and left eigenfunction systems associated with the autocorrelation operator and its adjoint are considered. The projection of the autoregressive Hilbertian state equation on these bases leads to its diagonalization. The projection of the conditioned log-likelihood functional on these biorthogonal eigenfunction bases provides a very suitable computational expression, where recursive ML multivariate gaussian regression methods can be applied in a very simple form. While Functional-Principal-Component-based factorization, associated with the spectral decomposition of the covariance operator, provides a more complicated expression of the projected conditioned log-likelihood function (see Ruiz-Medina and Salmerón, 2010). The asymptotic properties of the ML projection estimators derived from the implementation of EM algorithm in combination with forward and backward recursion have not been derived yet. This is the aim of the present paper, to compute an approximation, at each finite-dimensional Hilbert space, defined by truncation, of the asymptotic variance of ML projection estimators. This issue is addressed here by implementation of supplemented EM (SEM) algorithm (see Meng and Rubin, 1991) combined with forward and backward Kalman recursion. Additionally, two important aspects must be taken into account in the implementation of the referred ML projection estimation methodology : the truncation order and the discretization level. A simulation study is here developed to analyze the effect of the truncation order and the discretization level on the quality of the ML projection estimates.

The outline of the paper is the following. In Section 2, the preliminary elements involved in the diagonalization of the autoregressive Hilbertian state equation, in terms of the right and left eigenfunction systems of the autocorrelation operator, are introduced. Forward Kalman filtering and backward Kalman smoothing equations are then derived. In Section 3, the functional version of SEM algorithm is implemented. A simulation study is devoted in Section 4 to obtain information on the performance of the ML projection estimation methodology depending on the truncation order and on the discretization level considered. Some final comments are provided in Section 5.

2 Preliminaries

Let H be a separable Hilbert space endowed with the norm $||\cdot||_H$ associated with the inner product $<\cdot,\cdot>_H$. Consider the autoregressive Hilbertian model of order one (ARH(1) model)

$$Z_t(\mathbf{x}) = \mathcal{A}[Z_{t-1}](\mathbf{x}) + \nu_t(\mathbf{x}), \quad \mathbf{x} \in D,\ t \in \mathbb{N}^\star, \tag{1}$$

where $D \subset \mathbb{R}^n$ is a bounded domain, $Z = (Z_t, t \in \mathbb{N})$ and $\nu = (\nu_t, t \in \mathbb{N})$ are Hilbert-valued random processes on the basic probability space $(\Omega, \mathcal{A}, P)$. Process ν is assumed to be strong Hilbertian white noise, that is, a sequence of independent and identically distributed Hilbert-valued random variables in H with

$$E(\|\nu_t\|_H^2) = \sigma_\nu^2 < \infty,$$

uncorrelated with the random initial condition $Z_0 \in H$. The autocorrelation operator $\mathcal{A}$ is a bounded operator defined on a dense domain in H, which is assumed to satisfy that there exists a $j \geq 1$, such that $\|\mathcal{A}^j\|_{\mathcal{L}} < 1$, with $\mathcal{L}$ denoting the space of linear bounded operators on H.

Remark 1 *The results derived below can be similarly formulated within the ARH(p) model framework considering its matrix AHR(1) formulation (see Salmerón and Ruiz-Medina, 2009, and Ruiz-Medina and Salmerón, 2010).*

2.1 Projection of ARH(1) equation

The derivation of the diagonal expression of the ARH(1) state equation, from its projection on the eigenfunction systems associated with the autocorrelation operator $\mathcal{A}$ and its adjoint $\mathcal{A}^*$, is now provided. From now on, * stands for the adjoint, or equivalently, for transposition, in the finite-dimensional case.

Specifically, the spectral diagonal version of the functional equation (1) is derived in terms of the respective left and right eigenfunction systems $\{\psi_i,\ i \in \mathbb{N}\}$ and $\{\phi_i,\ i \in \mathbb{N}\}$ satisfying

$$\begin{aligned} \mathcal{A}\psi_i &= \lambda_i\psi_i, \quad i \in \mathbb{N}, \\ \mathcal{A}^*\phi_i &= \lambda_i\phi_i, \quad i \in \mathbb{N}. \end{aligned} \tag{2}$$

Here, $\{\lambda_i,\ i \in \mathbb{N}\}$ denotes the sequence of eigenvalues defining the pure point spectrum of $\mathcal{A}$ and $\mathcal{A}^*$ (see, for example, Dautray & Lions, 1992). The systems $\{\psi_i,\ i \in \mathbb{N}\}$ and $\{\phi_i,\ i \in \mathbb{N}\}$ are dual Riesz bases, that is, they are bases of H and its dual H^* satisfying

$$\langle \phi_i, \psi_j \rangle_H = \delta_{i,j}, \quad i,j \in \mathbb{N}.$$

Equivalently,

$$\Phi^*\Psi = I, \tag{3}$$

with I representing the identity operator, and Φ and Ψ being projection operators on the corresponding systems $\{\phi_i :\ i \in \mathbb{N}\}$ and $\{\psi_i :\ i \in \mathbb{N}\}$. From (2) and (3), operator $\mathcal{A}$ admits the spectral factorization

$$\mathcal{A} = \Psi\Lambda\Phi^*,$$

where Λ denotes a diagonal operator defined by the sequence of eigenvalues $\{\lambda_i,\ i \in \mathbb{N}\}$. Applying the projection operator Φ^* to both sizes of equation (1), the following diagonal version is obtained

$$\Phi^* Z_t = \Lambda \Phi^* Z_{t-1} + \Phi^* \nu_t.$$

Equivalently, for all $j \in \mathbb{N}$,

$$a_j(t) = \lambda_j a_j(t-1) + \nu_j(t), \quad j \in \mathbb{N}, \tag{4}$$

where

$$\begin{aligned} a_j(t) &= \langle Z_t(\cdot), \phi_j(\cdot) \rangle_H, \quad t \geq 0,\ j \in \mathbb{N}, \\ \nu_j(t) &= \langle \nu_t(\cdot),\ \phi_j(\cdot) \rangle_H, \quad t \geq 0,\ j \in \mathbb{N}. \end{aligned}$$

2.2 Forward Kalman recursion

First, forward Kalman filtering is implemented from a truncated version of equation (4), that is, from the projection of equation (4) on the $M-$finite-dimensional Hilbert dual spaces H_M and H_M^*, generated by the first M left and right eigenfunctions $\{\psi_1, \ldots, \psi_M\}$ and $\{\phi_1, \ldots, \phi_M\}$. Thus, the following finite dimensional diagonal equation is obtained :

$$\mathbf{a}(t) = \mathbf{\Lambda a}(t-1) + \boldsymbol{\nu}(t), \tag{5}$$

where, at each time $t \in \mathbb{N}$,

$$\begin{aligned} \mathbf{a}(t) &= (a_1(t), \ldots, a_M(t))^*, \\ \boldsymbol{\nu}(t) &= (\nu_1(t), \ldots, \nu_M(t))^*, \end{aligned}$$

and, as before, $\mathbf{\Lambda}$ denotes a diagonal $M \times M$ matrix with entries the eigenvalues $\lambda_i,\ i = 1, \ldots, M$. For implementation of Kalman filtering, we consider the functional observation equation

$$Y_t = Z_t + \varepsilon_t, \tag{6}$$

where $\varepsilon = (\varepsilon_t, t \in \mathbb{N})$ represents a zero-mean strong Hilbertian white noise, which is assumed to be uncorrelated with Z. Given the observations of process $Y = (Y_t, t \in \mathbb{N})$ up to time t, the estimators $\widehat{\mathbf{a}}(t|t) = E(\mathbf{a}(t)|Y_t, \ldots, Y_1)$ and $\widehat{\mathbf{a}}(t|t-1) = E(\mathbf{a}(t)|Y_{t-1}, \ldots, Y_1)$ of the temporal random Fourier coefficients $\mathbf{a}(t) = (a_j(t),\ j = 1, \ldots, M)$ will be computed from the following equations

$$\widehat{\mathbf{a}}(t|t) = \widehat{\mathbf{a}}(t|t-1) + \mathbf{K}_t \left(Y_t - \mathbf{\Psi} \widehat{\mathbf{a}}(t|t-1) \right),$$

$$\widehat{\mathbf{a}}(t|t-1) = \mathbf{\Lambda} \widehat{\mathbf{a}}(t-1|t-1).$$

Here $\mathbf{K}_t$ is the gain operator, that is, the filter defining the smoothing effect on the spatial functional data, given by

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{\Psi}^* \left(\mathbf{R}_\varepsilon + \mathbf{\Psi} \mathbf{P}_{t|t-1} \mathbf{\Psi}^* \right)^{-1},$$

in terms of the covariance operator $\mathbf{R}_\varepsilon$ of the observation noise, and the conditional second-order moment

$$\mathbf{P}_{t|t-1} = \mathrm{var}\left(\mathbf{a}(t)|Y_{t-1}, \ldots, Y_1\right) = \mathbf{\Lambda} \mathbf{P}_{t-1|t-1} \mathbf{\Lambda} + \mathbf{Q}_t,$$

denoting by $\mathbf{P}_{t|t} = E\left((\mathbf{a}(t) - \hat{\mathbf{a}}(t|t))(\mathbf{a}(t) - \hat{\mathbf{a}}(t|t))^*\right)$ and being $\mathbf{Q}_t = \mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M$ the covariance operator of the random projections of ν_t on the right eigenfunction system. The functional mean-square error is then approximated by

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{\Psi} \mathbf{P}_{t|t-1}.$$

The initial values considered are

$$\begin{aligned} \widehat{\mathbf{a}}(0|0) &= \mathbf{0}, \\ \mathbf{P}_{0|0} &= \mathbf{E}\left(\mathbf{a}(0)\mathbf{a}(0)^*\right), \end{aligned}$$

with $\mathbf{a}(0) = (a_1(0), \ldots, a_M(0))^*$ being the vector constituted by the truncated sequence of Fourier coefficients of the random initial condition, i.e.,

$$Z_0(\cdot) \sim \sum_{i=1}^{M} a_i(0)\psi_i(\cdot).$$

2.3 Backward Kalman smoothing

The backward Kalman smoothing is implemented from the above forward recursion as follows :

$$\begin{aligned} E\left(\mathbf{a}(t-1)|Y_{t_1}, \ldots, Y_{t_T}\right) &= \widehat{\mathbf{a}}_{t-1|t-1} + \left(\mathbf{P}_{t-1|t-1}\widehat{\mathbf{\Lambda}^*}(\mathbf{P}_{t|t-1})^{-1}\right) \\ &\quad \cdot \left(E\left(\mathbf{a}(t)|Y_{t_1}, \ldots, Y_{t_T}\right) - \widehat{\mathbf{\Lambda}}\widehat{\mathbf{a}}_{t-1|t-1}\right), \\ \mathrm{Var}\left(\mathbf{a}(t-1)|Y_{t_1}, \ldots, Y_{t_T}\right) &= \mathbf{P}_{t-1|t-1} + \left(\mathbf{P}_{t-1|t-1}\widehat{\mathbf{\Lambda}}^*(\mathbf{P}_{t|t-1})^{-1}\right) \\ &\quad \cdot \left(\mathrm{var}\left(\mathbf{a}(t)|Y_{t_1}, \ldots, Y_{t_T}\right) - \mathbf{P}_{t|t-1}\right) \\ &\quad \cdot \left(\mathbf{P}_{t-1|t-1}\widehat{\mathbf{\Lambda}^*}(\mathbf{P}_{t|t-1})^{-1}\right)^*, \end{aligned}$$

for $t = T, \ldots, 1$, where, as before, * stands transposition in the finite-dimensional case.

3 Implementation of SEM algorithm

We first describe the implementation of the EM algorithm. The selection problem related to choosing a suitable initial input values must be previously solved. We refer to Ruiz-Medina and Salmerón (2009), where this problem is widely discussed. In particular, moment functional estimators (see Bosq, 2000) are considered as initial values to begin with the first iteration of forward and backward Kalman filtering followed by EM iteration. Secondly we show the implementation of the SEM algorithm for missing functional data (see Meng and Rubin, 1991).

3.1 EM iteration

This section provides the implementation of the EM algorithm for ML parameter estimation of ARH processes.

To implement the EM algorithm we consider the *incomplete observed functional data* vector (noisy data) $Y^* = (Y_1, \ldots, Y_{t_T})$, which must be related with the complete functional data vector $(Z^* = (\Phi_M^* Z) = (a(t_1), \ldots, a(t_T)); (\Phi_M^* \varepsilon) = (\varepsilon(t_1), \ldots, \varepsilon(t_T)))$, in terms of a suitable transformation F, with $t_1, \ldots, t_T$ being the times where process Y is observed.

Under the Gaussian distribution assumption for the innovation and observation noise sequences, the EM algorithm is implemented, from the computation of the conditional expectation of the finite-dimensional log-likelihood function. First, the log-likelihood of the projected complete functional data has the following expression under stationary in time of processes ν and ε :

$$
\begin{aligned}
& C + \log f_{Z_0} - \frac{T}{2} \log |\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M| - \frac{T}{2} \log |\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M| \\
& -\frac{1}{2} \sum_{i=1}^{T} (\mathbf{a}(t_i) - \mathbf{\Lambda} \mathbf{a}(t_i - 1))^* \left(\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M\right)^{-1} (\mathbf{a}(t_i) - \mathbf{\Lambda} \mathbf{a}(t_i - 1)) \\
& -\frac{1}{2} \sum_{i=1}^{T} (\boldsymbol{\varepsilon}(t_i))^* \left(\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M\right)^{-1} (\boldsymbol{\varepsilon}(t_i)) \,.
\end{aligned}
$$

Considering that the effect of the random initial condition is negligible at the times $t_1, \ldots, t_T$, from standard results on quadratic forms, the conditional expectation of the *complete data*-based log-likelihood given the *incomplete data* is computed as

$$
\begin{aligned}
& C - \frac{T}{2} \log |\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M| - \frac{T}{2} \log |\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M| \\
& -\frac{1}{2} \mathrm{tr} \left\{ \left(\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M\right)^{-1} (C_Z - \mathbf{\Lambda} B_Z^* - B_Z \mathbf{\Lambda} + \mathbf{\Lambda} A_Z \mathbf{\Lambda}) \right\} \\
& -\frac{1}{2} \mathrm{tr} \left\{ \left(\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M\right)^{-1} C_\varepsilon \right\}, \qquad (7)
\end{aligned}
$$

where tr denotes the trace, and

$$
\begin{aligned}
C_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i) \mathbf{a}(t_i)^* | Y_{t_1}, \ldots, Y_{t_T}\right), \\
B_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i) \mathbf{a}(t_i - 1)^* | Y_{t_1}, \ldots, Y_{t_T}\right), \\
A_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i - 1) \mathbf{a}(t_i - 1)^* | Y_{t_1}, \ldots, Y_{t_T}\right), \\
C_\varepsilon &= \sum_{i=1}^{T} E\left(\mathbf{\Phi}_M^* \varepsilon_{t_i} (\mathbf{\Phi}_M^* \varepsilon_{t_i})^* | Y_{t_1}, \ldots, Y_{t_T}\right).
\end{aligned}
$$

Differentiating (7) with respect to $\mathbf{\Lambda}$, $\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M$, and $\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M$, the maximum likelihood estimates of the model parameters computed in the $k-$iteration of the Maximization step

(M-step) are given by

$$\begin{aligned}
\widehat{\mathbf{\Lambda}}^{(k)} &= \text{diag}[B_Z]^{(k-1)}[(\text{diag}[A_Z])^{-1}]^{(k-1)} \\
\mathbf{\Phi}^*_{M\times M}\widehat{R_\nu \mathbf{\Phi}}_{M\times M}{}^{(k)} &= \frac{1}{T}\sum_{i=1}^{T} E\left[\left(\mathbf{a}(t_i)-\widehat{\mathbf{\Lambda}}^{(k)}\mathbf{a}(t_i-1)\right)\right. \\
&\quad \left.\cdot\left(\mathbf{a}(t_i)-\widehat{\mathbf{\Lambda}}^{(k)}\mathbf{a}(t_i-1)\right)^T/Y_1,\ldots,Y_T\right] \\
&= \frac{1}{T}\left[C_Z^{(k-1)}-B_Z^{(k-1)}\widehat{\mathbf{\Lambda}}^{(k)}-\widehat{\mathbf{\Lambda}}^{(k)}[B_Z^T]^{(k-1)}\right. \\
&\quad \left.+\widehat{\mathbf{\Lambda}}^{(k)}A_Z^{(k-1)}\widehat{\mathbf{\Lambda}}^{(k)}\right], \\
\mathbf{\Phi}^*_{M\times M}\widehat{R_\varepsilon \mathbf{\Phi}}_{M\times M} &= C_\varepsilon^{(k-1)}/T,
\end{aligned}$$

where diag$[A]$ denotes the diagonal of matrix A, and where $A_Z^{(k-1)}$, $B_Z^{(k-1)}$ and $C_Z^{(k-1)}$ are computed from the forward Kalman filtering, and the backward Kalman smoothing, with input values the functional parameter estimates obtained in the $k-1-$iteration of the EM algorithm.

Remark 2 *Note that, from the EM algorithm, the estimators $\widehat{\mathbf{\Lambda}}$ and $\widehat{\boldsymbol{Q_t}}$ are computed. However, Φ and Ψ are not, in general, known. For the estimation of $\mathcal{A}$ and R_ν, we proceed as follows : At the first iteration, the initial functional values considered for the eigenfunction systems $\{\phi_i,\ i\in\mathbb{N}\}$ and $\{\psi_i,\ i\in\mathbb{N}\}$, defining projection operators $\widehat{\Phi}_0$ and $\widehat{\Psi}_0$, are constructed from the spectral decomposition of the functional moment estimator*

$$\widehat{\mathcal{A}} = \widehat{R}_{Z_1Z_0}\cdot\widehat{R}_{Z_0}^{-1},$$

and its adjoint, where

$$\begin{aligned}
\widehat{R}_{Z_1Z_0} &= \frac{1}{T-1}\left[\sum_{i=1}^{T-1}Y_i\otimes Y_{i+1}\right]-\frac{1}{T-1}\left[\sum_{i=1}^{T-1}Y_i\right]\otimes\frac{1}{T-1}\left[\sum_{i=1}^{T-1}Y_{i+1}\right] \\
\widehat{R}_{Z_0} &= \frac{1}{T}\left[\sum_{i=1}^{T}Y_i\otimes Y_i\right]-\frac{1}{T}\left[\sum_{i=1}^{T}Y_i\right]\otimes\frac{1}{T}\left[\sum_{i=1}^{T}Y_i\right]-\widehat{\sigma}_\varepsilon^2 I, \qquad (8)
\end{aligned}$$

with Y_i, $i=1,\ldots,T$, being the observed functional incomplete data, defined as in equation (6). Estimator $\widehat{\sigma}_\varepsilon^2$ can be computed from the nugget effect of the functional variogram (see, for example, Ruiz-Medina, Salmerón and Angulo, 2007). In the rest of iterations, projection operators $\widehat{\Phi}_k$ and $\widehat{\Psi}_k$ are derived from the spectral decomposition of $\widehat{\mathcal{A}}^{(k-1)}$ computed in the previous iteration of the EM algorithm.

3.2 Finite-dimensional approximation of the SEM algorithm

The functional SEM algorithm provides the observed-functional-data asymptotic covariance operator of the infinite-dimensional ML parameter estimator, from the computation of

the complete-functional-data asymptotic covariance operator, and the convergence rate operator, i.e., the operator providing information about the rate of convergence of the EM algorithm in the Hilbert-valued random variable context. Note that, in this context, each iteration of the EM algorithm can be expressed as

$$\boldsymbol{\Theta}^{(k+1)} = F(\boldsymbol{\Theta}^{(k)}), \quad k \in \mathbb{N},$$

where F denotes the functional which implicitly describes the EM algorithm for ML estimation of the functional parameter vector $\boldsymbol{\Theta}$. Here, the limit $\boldsymbol{\Theta}^*$, in the bounded linear operator norm, of the functional sequence $\{\boldsymbol{\Theta}^{(k)},\ k \in \mathbb{N}\}$ is defined as

$$\boldsymbol{\Theta}^* = F(\boldsymbol{\Theta}^*),$$

in the case where F is continuous with respect to the operator norm considered. The convergence rate operator is then defined as the Fréchet derivative of the functional F at point $\boldsymbol{\Theta}^*$.

In this section, a finite-dimensional approximation of the functional SEM algorithm is implemented, in terms of the recursive projection estimators obtained in the previous section from the EM algorithm. Specifically, the complete-functional-data asymptotic covariance operator, I_{CFD}^{-1}, is approximated from the expression

$$I_{CFD}^{-1} = L_T^{-1} \cdot \left[\sum_{t=1}^{T} L(t-1)^* \left(\Phi_M^* R_\nu \Phi_M \right) L(t-1) \right] \cdot L_T^{-1},$$

with $L_T = \sum_{t=1}^{T} L(t-1)^* L(t-1)$, $L(t-1) = \text{diag}\left(a_1(t-1), \ldots, a_M(t-1)\right)_{M\times M}$ and R_ν being the covariance operator of ν. In practice, $L(t-1)$ is replaced by $\widehat{L(t-1)} = \text{diag}\left(E\left(\mathbf{a}(t-1)|Y_{t_1}, \ldots, Y_{t_T}\right)\right)_{M\times M}$, obtained as the output of the corresponding forward and backward Kalman recursion iteration, and R_ν is replaced by $\widehat{R}_\nu$, computed as the output of the corresponding iteration of the EM algorithm.

The finite-dimensional approximation of the convergence rate operator is computed from the rate matrix, $DM = (r(i,j))_{i,j=1,\ldots,M}$, which is calculated as follows : First run the forward and backward Kalman filter and EM algorithm to obtain $\widehat{\Lambda}^{(k)}$, at iteration k, and then use the following algorithm, for $i = 1, \ldots, M$:

Step 1 : Compute $\widetilde{\Lambda}_{(i,i)}^{(k)}$, with $i \in \{1, \ldots, M\}$, by replacing in the original matrix Λ the value of the (i,i) element by the (i,i) element of the matrix $\widehat{\Lambda}^{(k)}$. That is,

$$\widetilde{\Lambda}_{(i,i)}^{(k)} = \begin{cases} \lambda_j, & j = 1, \ldots, i-1, i+1, \ldots, M \\ \widehat{\lambda}_j^{(k)}, & j = i, \end{cases}$$

where $\widehat{\lambda}_j^{(k)}$ is the (j,j) element of $\widehat{\Lambda}^{(k)}$.

Step 2 : Run the Kalman filter and EM algorithm, with input parameter values given by the elements of $\widetilde{\Lambda}_{(i,i)}^{(k)}$, to obtain $\widetilde{\Lambda}_{(i,i)}^{(k+1)}$.

Step 3 : Obtain, for $l = 1, \ldots, M$, and for $\widetilde{\lambda}_{(i,i)}^{(k+1)}(l,l)$, the (l,l) element of $\widetilde{\Lambda}_{(i,i)}^{(k+1)}$, the ratio

$$r(i,l) = \frac{\widetilde{\lambda}_{(i,i)}^{(k+1)}(l,l) - \lambda_l}{\widehat{\lambda}_i^{(k)} - \lambda_i}. \tag{9}$$

After that, run the forward and backward Kalman filter and EM algorithm, with initial input parameter values given by the elements of $\widehat{\Lambda}^{(k)}$, to obtain $\widehat{\Lambda}^{(k+1)}$, and repeat steps 1, 2 and 3.

Finally, the observed-functional-data asymptotic covariance operator of the computed $\widehat{\Lambda}$ estimator is approximated by

$$V = I_{CFD}^{-1} \left(I - DM\right)^{-1}, \tag{10}$$

with I being the identity matrix. Note that the diagonal elements of DM matrix provide information on the convergence rates to the elements of Λ, in the EM approximation.

The finite-dimensional estimation of the asymptotic covariance operator of the estimator $\widehat{\Phi_M^* R_\nu \Phi_M}$ can be computed, in a similar way to the above algorithm, from the forward and backward Kalman filter and smoothing iteration, as well as from the corresponding EM iteration, considering the finite-dimensional estimation of the corresponding complete-functional-data covariance operator and DM matrix. Similar computations must be done for the finite-dimensional approximation of the asymptotic cross covariance operator between $\widehat{\Phi_M^* R_\nu \Phi_M}$ and $\widehat{\Lambda}$.

4 Simulation study

Since in the projection estimation methodology proposed, the truncation order is selected according to the percentage of the trace norm of the autocorrelation operator explained, a good approximation of the theoretical trace of such an operator must be previously derived. In the simulation study developed in this section, this fact is illustrated. The goal of this section is then to investigate the fixed-domain asymptotic properties of the estimates obtained for the eigenvalues of the autocorrelation operator, as well as for the covariance operator of the functional innovation process.

We consider the case where the ARH(1) process Z has autocorrelation operator defined in terms of the integral operator K with kernel

$$k(z;\theta) = \frac{1}{\theta} \exp\left\{-\frac{||z||}{\theta}\right\},$$

where $||z||$ denotes the distance between spatial locations, and θ is the scale parameter involved in the definition of the exponential kernel.

The random initial condition is assumed to be a Hilbert-valued Gaussian random variable with integral covariance operator defined in terms of a factorizable kernel $r_{Y_0} = t_{Y_0} * t_{Y_0}$, where $*$ denotes convolution, and with

$$t_{Y_0}(z;\gamma) = \frac{1}{\sqrt{2 \cdot \pi \cdot \gamma}} \cdot \exp\left\{-\frac{||z||^2}{2 \cdot \gamma}\right\}.$$

Covariance factorization is derived in Ruiz-Medina, Angulo and Anh (2003) considering the scale of fractional Sobolev spaces. In this example, we consider such a scale on $\mathbb{R}^2$. Indeed, the underlying nuclear Hilbert structure, defined by Bessel potentials and their inverses, is considered for the choice of the norm of the Hilbert space H where process Z, as well as the innovation and observation processes are suitably defined as second-order Hilbert-valued

Functional SEM algorithm

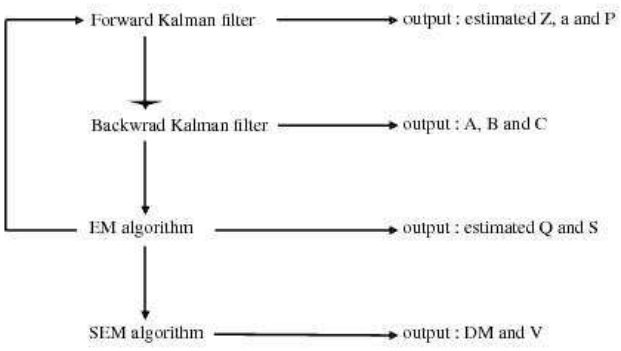

FIG. 1 – *Main steps in the implementation of the ML projection estimation method*

processes. Specifically, the innovation process is assumed to be a functional Gaussian process with covariance operator $R_\nu = \sigma_\nu^2 I$, with I being the identity operator, and

$$\sigma_\nu^2 = \frac{0.1}{N\sqrt{2 \cdot \pi}}.$$

The observation noise is also assumed to be a Gaussian spatiotemporal white noise with variance σ_ε^2. The space H must then be endowed with the norm generated by a Bessel potential of order $\alpha > 1$ (see Embedding Theorems between fractional Besov spaces in, for example, Triebel, 1978). Regarding the spectral factorization of the autocorrelation operator in this setting, note that dual Riesz bases of H and H^* can then be constructed from an orthonormal basis of $L^2(\mathbb{R}^2)$ transformed by the corresponding Bessel potential and the adjoint of its inverse (see Ruiz-Medina, Salmerón and Angulo, 2007, and Salmerón and Ruiz-Medina, 2009, for such a construction). The quality of the projection estimates is computed below in terms of the Hilbert-Schmidt norm.
In the results displayed below, we consider the case where $\theta = 5$ and $\gamma = 1$.

The steps followed in the implementation of the ML projection estimation methodology, from the generated data, are described in Figure 1. Specifically, the first step consists of the implementation of forward Kalman filtering, getting an approximation $\widehat{Z}$ of the functional values of Z, in terms of the computation of $\widehat{\mathbf{a}}(t|t)$ and $\mathbf{P}_{t|t}$. The last two terms are entries in the backward Kalman filter. The application of backward Kalman filter allows the computation of matrices A_Z, B_Z and C_Z, i.e., the matrices involved in the estimation of functional parameters Q and Λ by EM algorithm. These estimates will be used as initial values in a new iteration of the forward Kalman filtering. Finally, the values obtained from the EM algorithm will also be used to calculate V and DM through the SEM algorithm. This loop will be repeated until convergence

Different truncation orders ($M-$values) (e.g., corresponding to 50%, 60%, 70%, 80%, and 90% of the trace norm of the empirical autocorrelation operator) are considered in the simulation study developed below. Table 1 provides the value of M corresponding to different percentages of the trace norm of the autocorrelation operator, when the discretization step size changes.

Regular grid	16×16	21×21	26×26
50%	3	5	7
60%	5	8	11
70%	8	13	19
80%	15	25	37
90%	37	62	93
N	256	441	676

TAB. 1 – *Truncation orders, M, for each spatial regular grid*

Regular grid	AVD
16×16	0.005088
21×21	0.004028
26×26	0.003437

TAB. 2 – *Mean absolute difference between empirical and theoretical spectral tails*

Table 2 displays the mean absolute value of the differences (AVD) between the smallest $N-16$ theoretical and estimated eigenvalues, for each spatial regular grid considered. That is

$$\mathrm{AVD} = \frac{1}{\mathrm{N}-16} \sum_{\mathrm{i}=16}^{\mathrm{N}} |\lambda_{\mathrm{i}} - \widehat{\lambda}_{\mathrm{i}}|.$$

The AVD is considered to measure the distance between the empirical and the theoretical spectral tails of the autocorrelation operator. Since consistency of the empirical eigenvalues computed leads to the convergence of the empirical spectral tail to the theoretical one, the AVD is used as a criterion for selection of a convenient discretization step size. Furthermore, the local regularity properties of the autocorrelation operator are reflected in the decay velocity of its spectrum. Therefore, we consider the last $N-16$ smallest eigenvalues, where the spectral decay velocity is captured to compute, in terms of the AVD, the most significative differences between the empirical and theoretical autocorrelation local regularity properties. As expected, in Table 2, it is observed that the mean absolute difference decreases when density of the spatial grid considered increases, i.e., when discretization step size goes to zero. Also, Figure 3 illustrates the fact that differences between the empirical and theoretical spectral tails of the autocorrelation operator disappear when the discretization step size goes to zero, i.e., the empirical local regularity properties of the autocorrelation operator tends to the theoretical ones. On the other hand, in Figure 9, differences between the theoretical values and EM estimates of the eigenvalues are displayed, for $N = 121$, $N = 256$ and $N = 441$ spatial locations, and different truncation orders. Additionally to the commented effect of the discretization step size on the empirical eigenvalues, initializing the EM algorithm, it can also be seen that, for a fixed N, the quality of the EM estimates of the eigenvalues is improved when M increases, as expected (see light blue line). Note that, in all the cases, differences are very small (of order of 10^{-17}).

Figures 4-6 display the theoretical values and the functional estimates of the process of interest Z, at times $t = 1, 10, 20$, and considering the truncation order corresponding to explai-

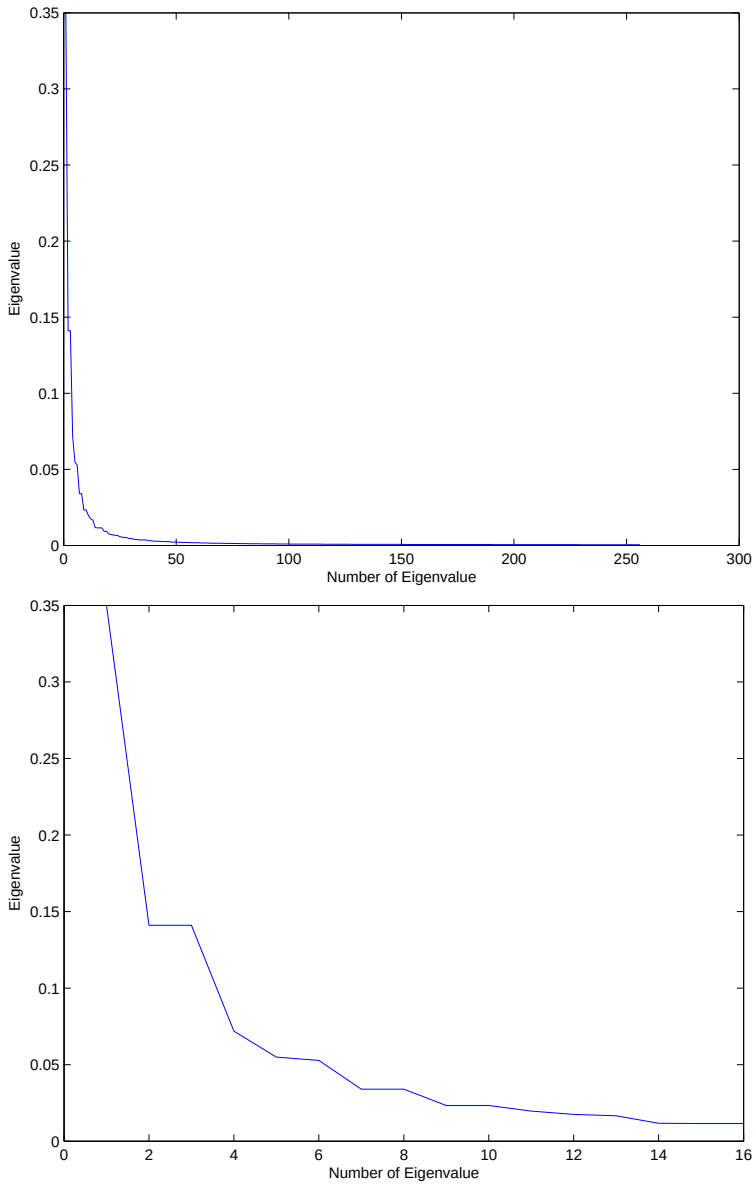

FIG. 2 – *Decay velocity of the theoretical eigenvalues of the autocorrelation exponential-type kernel*

ning a 90% of the trace norm of the autocorrelation operator. These results are again displayed for a regular spatial grid with $N = 121$, $N = 256$ and $N = 441$ points. As before, it can be appreciated that the quality of the estimates increases when the discretization step size decreases.

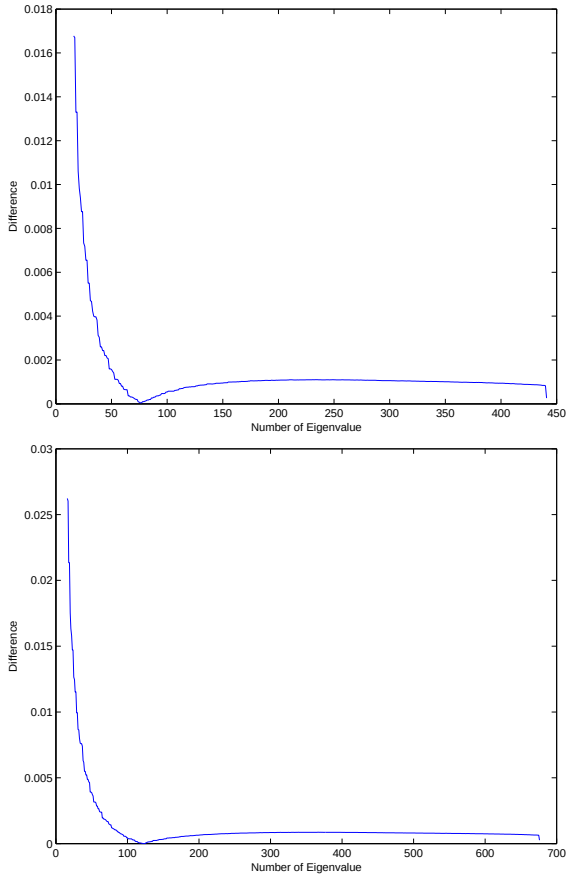

FIG. 3 – *Absolute value differences between the last* $N - 16$ *theoretical and empirical eigenvalues for* $N = 441$ *and* $N = 676$ *spatial locations (top to bottom)*

The functional quadratic errors (F.Q.E.) associated with the functional estimation of ARH(1) process Z (after applying forward and backward Kalman recursion combined with EM) are computed as

$$\frac{1}{N}\sum_{i=1}^{\sqrt{N}}\sum_{j=1}^{\sqrt{N}}\left[Z_t(i,j)-\widehat{Z}_t(i,j)\right]^2.$$

Figure 7 displays the F.Q.E. for the three spatial functional data discretization levels considered ($N = 121$, $N = 256$ and $N = 441$). As commented in Figures 4-6, the quality of the estimates is improved when the discretization step size goes to zero. Thus, when N increases, F.Q.E. are mainly produced by the truncation error (see also Figures 4-6), while, for small N, truncation and discretization errors are involved in the F.Q.E., and less quality of the estimates is observed, for each truncation order, in comparison with large values of N.

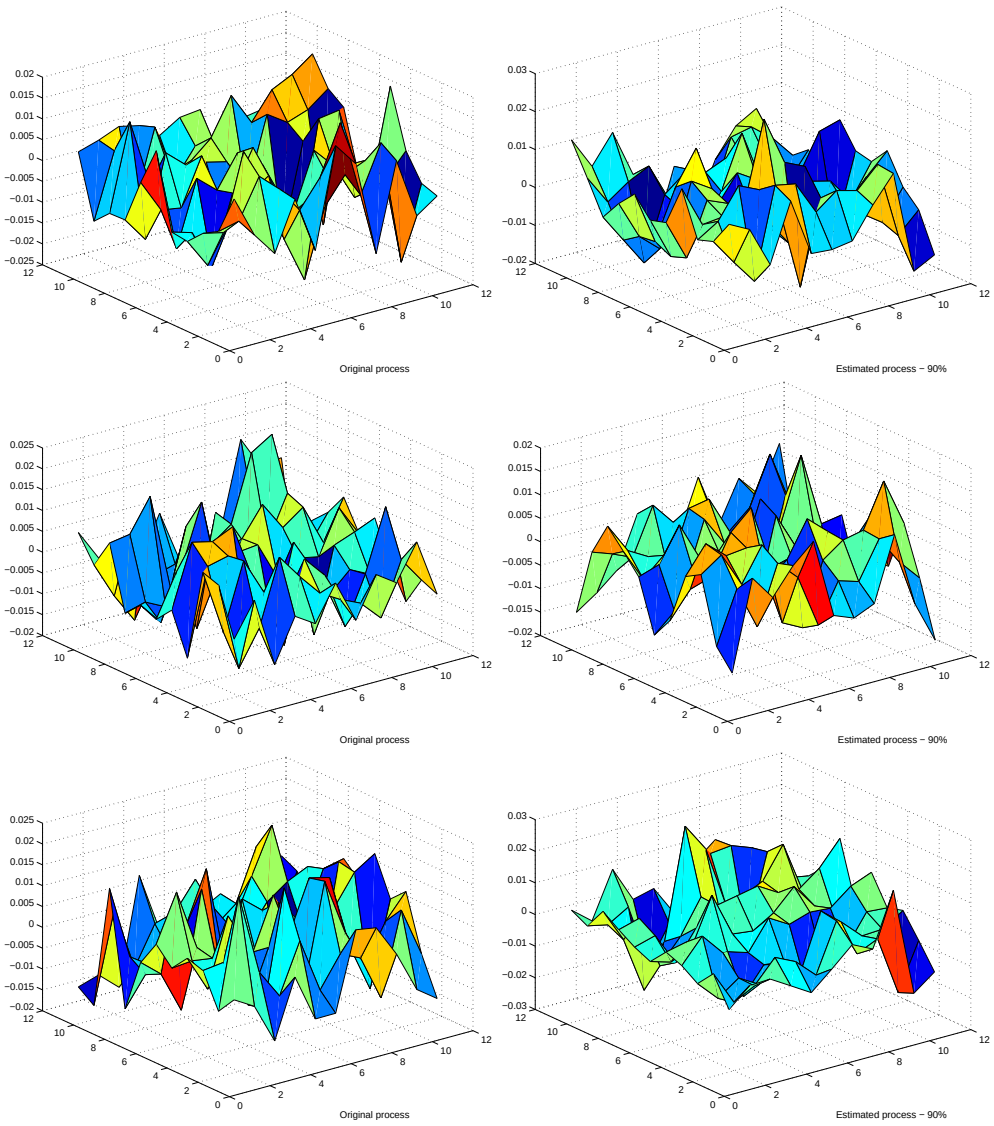

FIG. 4 – *Original (left) and estimated process Z (right) at times $t = 1, 10, 20$ (top to bottom), for $N = 121$, and truncation order 90%*

Figure 8 shows the estimated variance $\widehat{\sigma}^2_\nu$, for the truncation orders considered, and for $N = 121$, $N = 256$ and $N = 441$ jointly with the real value (black squares). As expected, for each value of N, the quality of the estimates increases when M increases (purple line).

To complete this study, the distance between the theoretical and estimated operators Λ and $\Phi^* R_\nu \Phi$ is displayed in Figure 10, for each truncation order and regular grid, at 6 iterations.

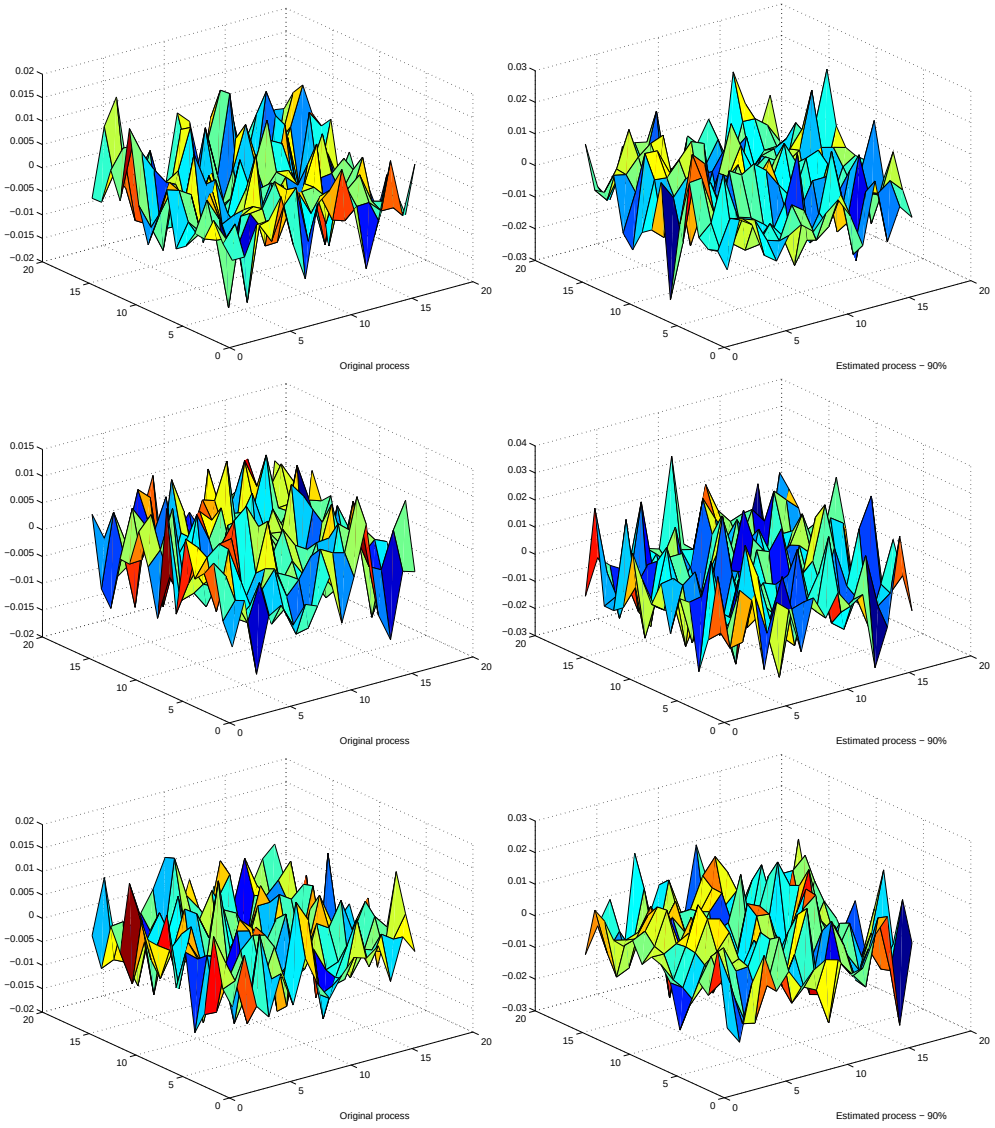

FIG. 5 – *Original (left) and estimated process Z (right) at times $t = 1, 10, 20$ (top to bottom), for $N = 256$, and truncation order 90%*

That is, the following norms

$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(Q(i,j)-\widehat{Q}(i,j)\right)^2, \quad \frac{\frac{1}{M^2}\sum\limits_{i,j}\left(\Lambda(i,j)-\widehat{\Lambda}(i,j)\right)^2}{(1/M)\sum_i|\Lambda(i,i)-\widehat{\Lambda}(i,i)|}$$

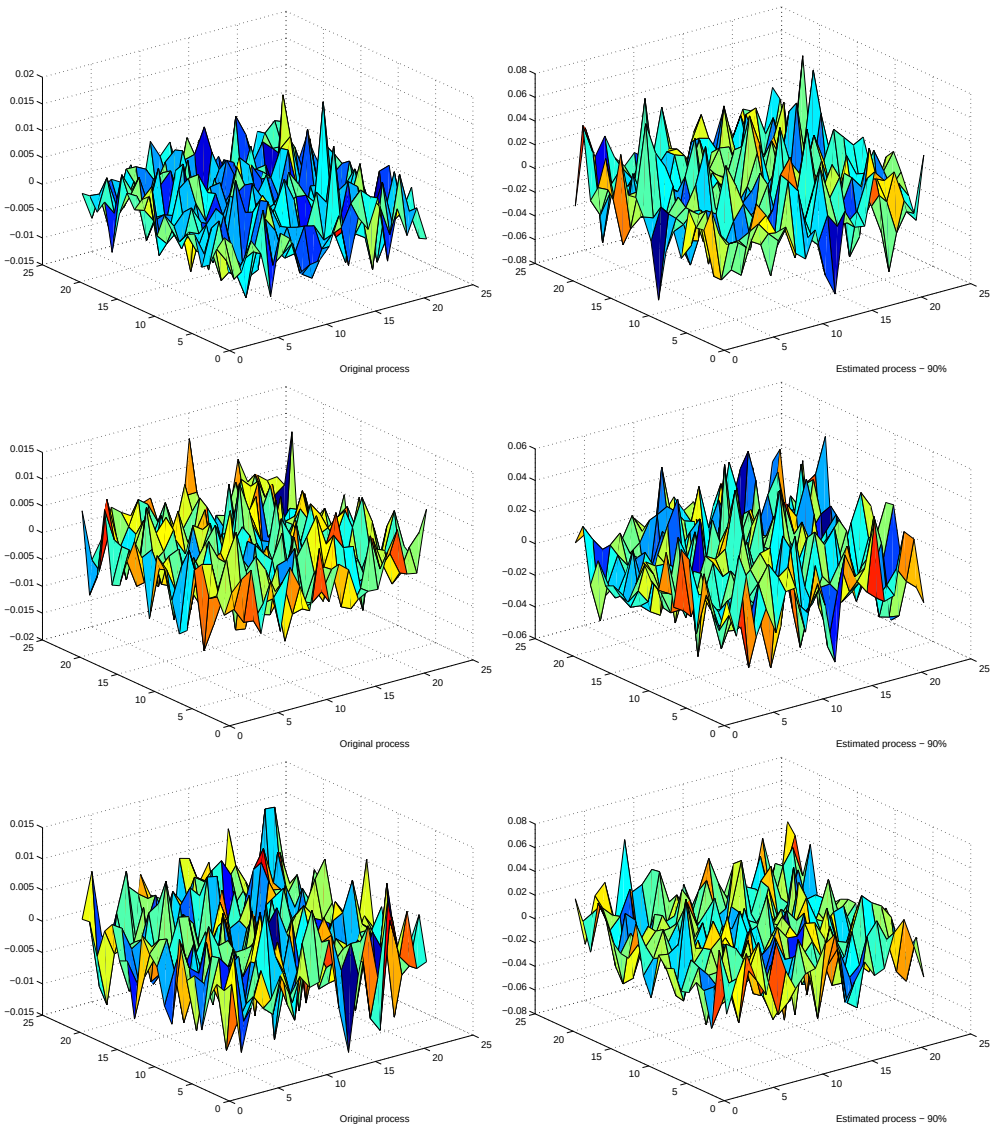

FIG. 6 – *Original (left) and estimated process Z (right) at times $t = 1, 10, 20$ (top to bottom), for $N = 441$, and truncation order 90%*

are computed, where $Q = \Phi^* R_\nu \Phi$. These norms are approximately of order $10^{-17} - 10^{-16}$, for Λ, and of order 10^{-8}, for $\Phi^* R_\nu \Phi$. The results displayed reflect the commented dependence on the discretization step size and truncation order of the performance of the projection estimation methodology proposed.

Finally, Figure 11 shows, after four iterations, the ratio of convergence in the estimation

of $\widehat{\Lambda}$ and $\widehat{\Phi^* R_\nu \Phi}$, for each discretization level ($N = 121$, $N = 256$ and $N = 441$), and truncation order. Faster convergence is observed for $N = 441$ than for $N = 121$. The averages of the diagonal elements of DM matrix computed with the SEM algorithm are also displayed in Figure 12. The diagonal of DM provides information on the ratio of convergence of the EM estimate sequence of Λ. From expression (9), these diagonal values must be close to one for the convergence of the estimate sequence. In Figure 12, it can be appreciated that, after 6 iterations, for all discretization step sizes, convergence is achieved by the EM estimates of the eigenvalues of the autocorrelation operator.

5 Final Comments

The recursive functional ML projection estimation methodology developed here is based on the biorthogonal spectral decomposition of the autocorrelation operator. The left and right autocorrelation eigenfunction systems provide a more suitable computational framework for implementation of recursive functional ML estimation algorithms, from incomplete functional data. In particular, Functional Principal Component Analysis of ARH(1) processes is more suitable for moment-based estimation (see Bosq, 2000). Finally, we remark that the asymptotic normality of the incomplete-functional-data based ML projection estimators of the ARH(1)-infinite-dimensional parameters is studied in Ruiz-Medina (2010).

Acknowledgments. This work has been supported in part by projects MTM2008-03903 and MTM2009-13393 of the DGI, MEC, and P06-FQM-02271 and P09-FQM-5052 of the Andalousian CICE, Spain.

References

Abramovich, F. and Angelini, C. (2006). Testing in mixed effects FANOVA models. *Journal of Statistical Planning and Inference* **136**, 4326-4348.

Abramovich, F., Antoniadis, A., Sapatinas, T and Vidakovic, B. (2004). Optimal testing in functional analysis of variance models. *Int. J. Wavelets Multiresolution Inform. Process.* **2**, 323-349.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis* **87**, 133-158.

Besse, P., Cardot, H. and Stephenson, D.B. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**, 673-687.

Bosq, D. (2000). *Linear processes in function spaces*, Springer-Verlag.

Bosq, D. (2008). Nonparametric statistics for stochastic processes, estimation and prediction. *Lectures Notes in Statistics* **110**, Springer-Verlag, New York.

Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions*, Wiley Series in Probability and Statistics, John Wiley & Sons.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* **30**, 241-255.

Dautray, R. and Lions, J.L. (1992). *Mathematical analysis and numerical methods for science and technology 3, Spectral Theory and Applications*, Springer-Verlag, Berlin.

Dempster, A.P., Laird, N.M. and Rubin D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1, 1-38.

Ferraty, F. and Vieu, P. (2006). *Nonparameric functional data analysis*, Springer.

Germain, F., Doisy, A., Ronot, X. and Tracqui, P. (1999). Characterization of cell deformation and migration using a parametric estimation of image motion. *IEEE Transactions on Biomedical Engineering* **46**, 584-600.

Haoudi, A. and Bensmail, H. (2006). Bioinformatics and data mining in proteomics. *Expert Review of Proteomics*, **3**, 333–343.

Hartley, H.O. (1958). Maximum Likelihood Estimation from Incomplete Data. *Biometrics*, **14**, 2, 174-194.

Hyndman, R.J. and Ullah, M.S. (2006). Robust forecasting of mortality and fertility rates : A functional data approach. *Computational Statistics & Data Analysis* **51**, 4942-4956.

Leng, X. and Müller, H.G. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatics* **22**, 68–76.

Mas, A. (1999) . Normalittè asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). *C. R. Acad. Sci. Paris* **329** Série I 899-902.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *J. Multivariate Analysis* **98,** 1231-1261.

Meng, X.L. and Rubin, D.B. (1991). Using EM to Obtain Asymptotic Variance-Covariance Matrices : The SEM Algorithm. *Journal of the American Statistical Association* **86**, 899-909.

Monk, N.A.M. (2003). Unravelling Nature's networks. *Biochemical Society Transactions* **31**, 1457-1561.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis*, Springer, New York.

Ruiz-Medina, M.D. (2010). Asymptotic normality of functional maximum-likelihood ARH parameter estimators. (Submitted).

Ruiz-Medina, M.D., Angulo, J.M. and Anh, V.V. (2003) Fractional generalized random fields on bounded domains. *Stochastics Analysis and Applications*, **21**, 465-492.

Ruiz-Medina, M.D., Salmerón, R. and Angulo, J.M. (2007). Kalman filtering from POP-based diagonalization of ARH(1). *Computational Statistics & Data Analysis.* **51**, 4994-5008.

Ruiz-Medina, M.D. and Salmerón, R. (2009). Functional maximum-likelihood estimation of ARH(p) models. Stochastic Environmental Research and Risk Assessment. doi : 10.1007/s00477-009-0306-2.

Salmerón, R. and Ruiz-Medina, M.D. (2009). Multispectral decomposition of functional autoregressive models. *Stochastic Environmental Research and Risk Assessment* **23**, 289-297.

Song, J.J., Lee, H.J., Morris, J.S. and Kangd, S. (2007). Clustering of time-course gene expression data using fuctional data analysis. *Computational Biology and Chemistry* **31**, 265-274.

Triebel, H. (1978) *Interpolation theory, function spaces, differential operators*. North-Holland Publishing Co, New York.

Summary

This paper addresses the problem of computation of the asymptotic variance of the maximum likelihood projection parameter estimators derived in Ruiz-Medina and Salmerón (2010) for autoregressive Hilbertian (ARH) processes, from Gaussian incomplete functional data. A functional version of the SEM (Supplemented Expectation Maximization) algorithm, derived by Meng and Rubin (1991), is implemented for approximation of such variances. The implementation is performed in terms of the eigenfunction systems and eigenvalues diagonalizing the autocorrelation operator and its adjoint. A simulation study is devoted to investigate the effect of the truncation order, and the functional data discretization level.

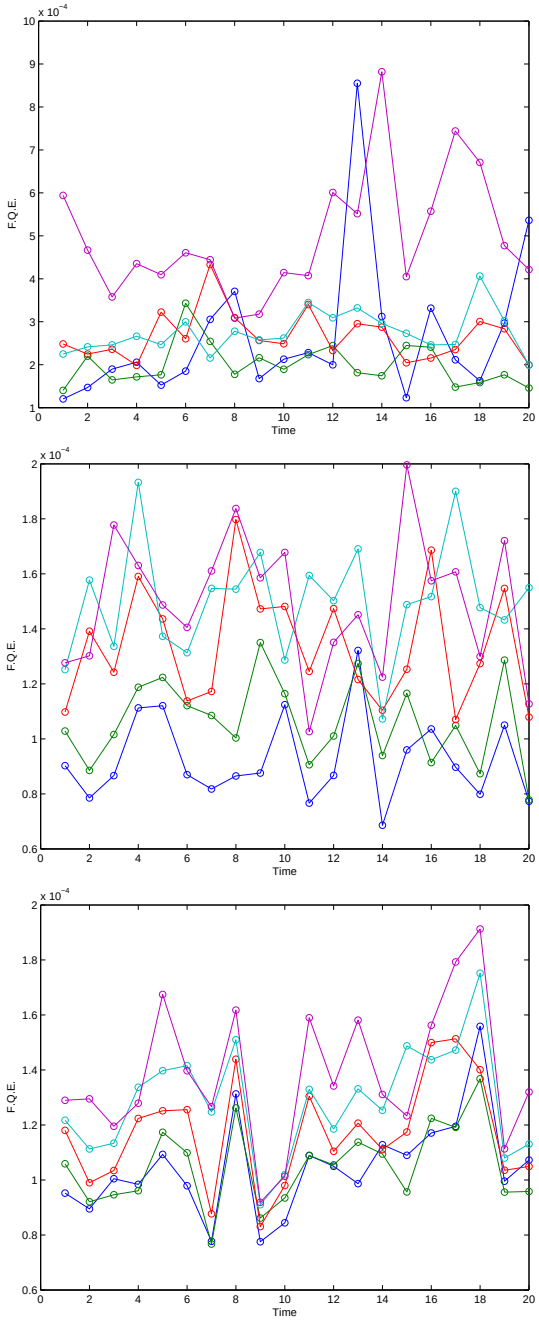

FIG. 7 – *F.Q.E. in the functional estimation of Z for $N = 121$, $N = 256$ and $N = 441$ spatial locations (top to bottom). The purple line represents the F.Q.E. for truncation order corresponding to 50% of the trace norm of the empirical autocorrelation operator, green line for 60%, red line for 70%, light blue for 80%, and blue line for 90%*

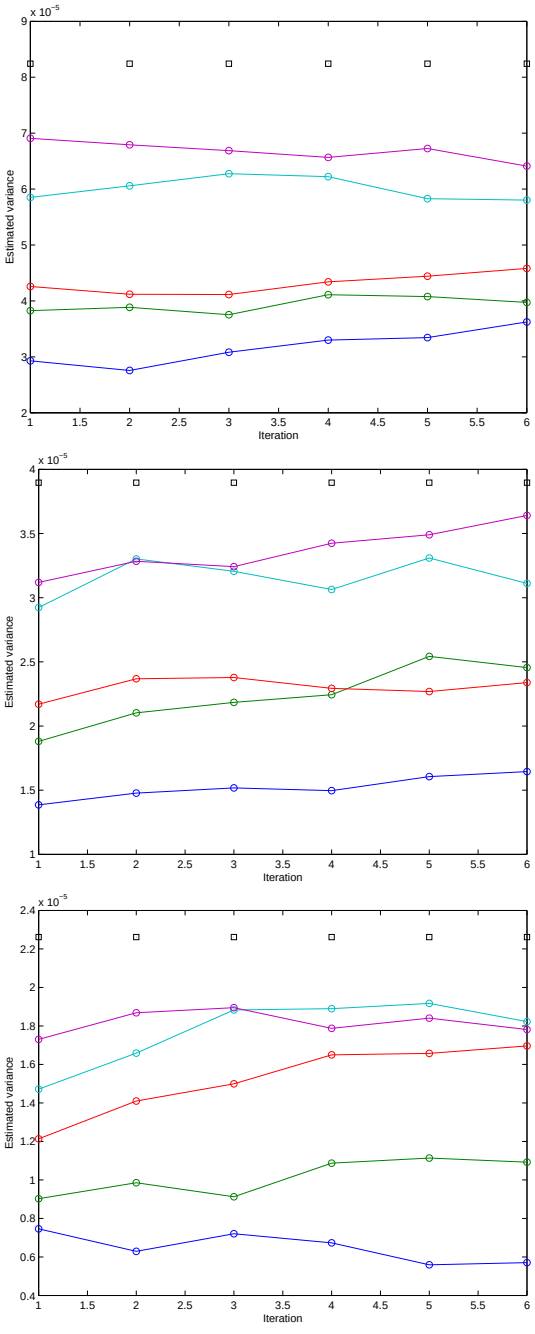

FIG. 8 – *Estimated variance* $\widehat{\sigma}_{\nu}^2$ *for* $N = 121$, $N = 256$ *and* $N = 441$ *spatial locations (top to bottom). The blue line represents the estimate obtained for truncation order corresponding to 50% of the trace norm of the empirical autocorrelation operator, green line for 60%, red line for 70%, light blue for 80%, and purple line for 90%*

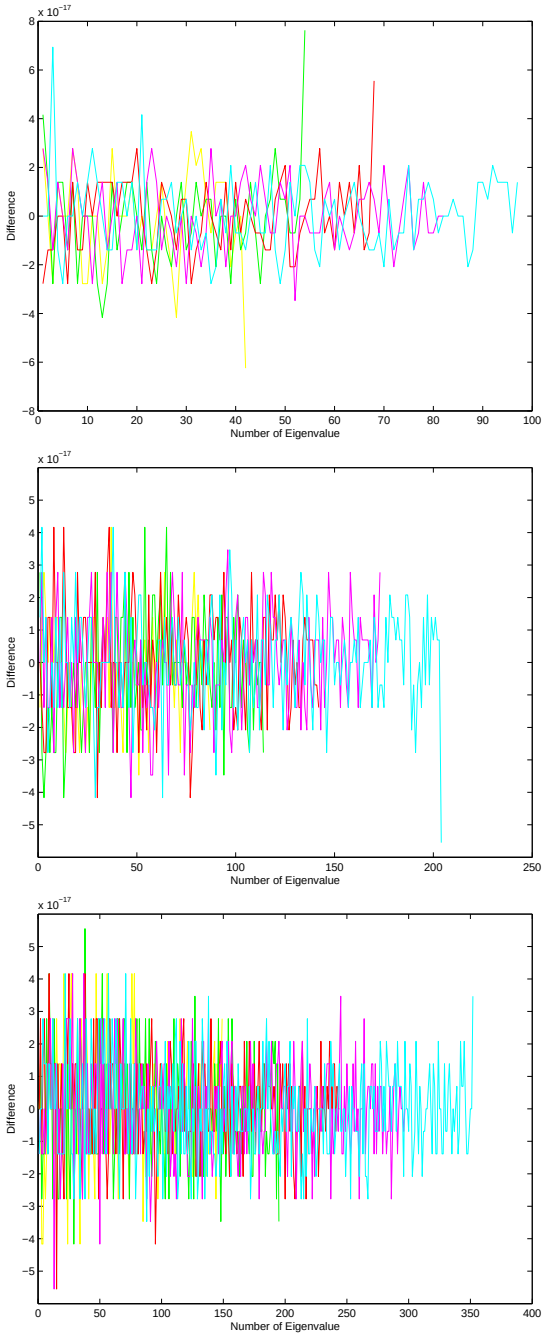

FIG. 9 – *Differences between theoretical values and EM estimates of the eigenvalues for* $N = 121$, $N = 256$ *and* $N = 441$ *spatial locations (top to bottom). Yellow line for truncation order corresponding to 50% , green line for 60% , red line for 70% , purple line for 80%, and light blue line for 90%.*

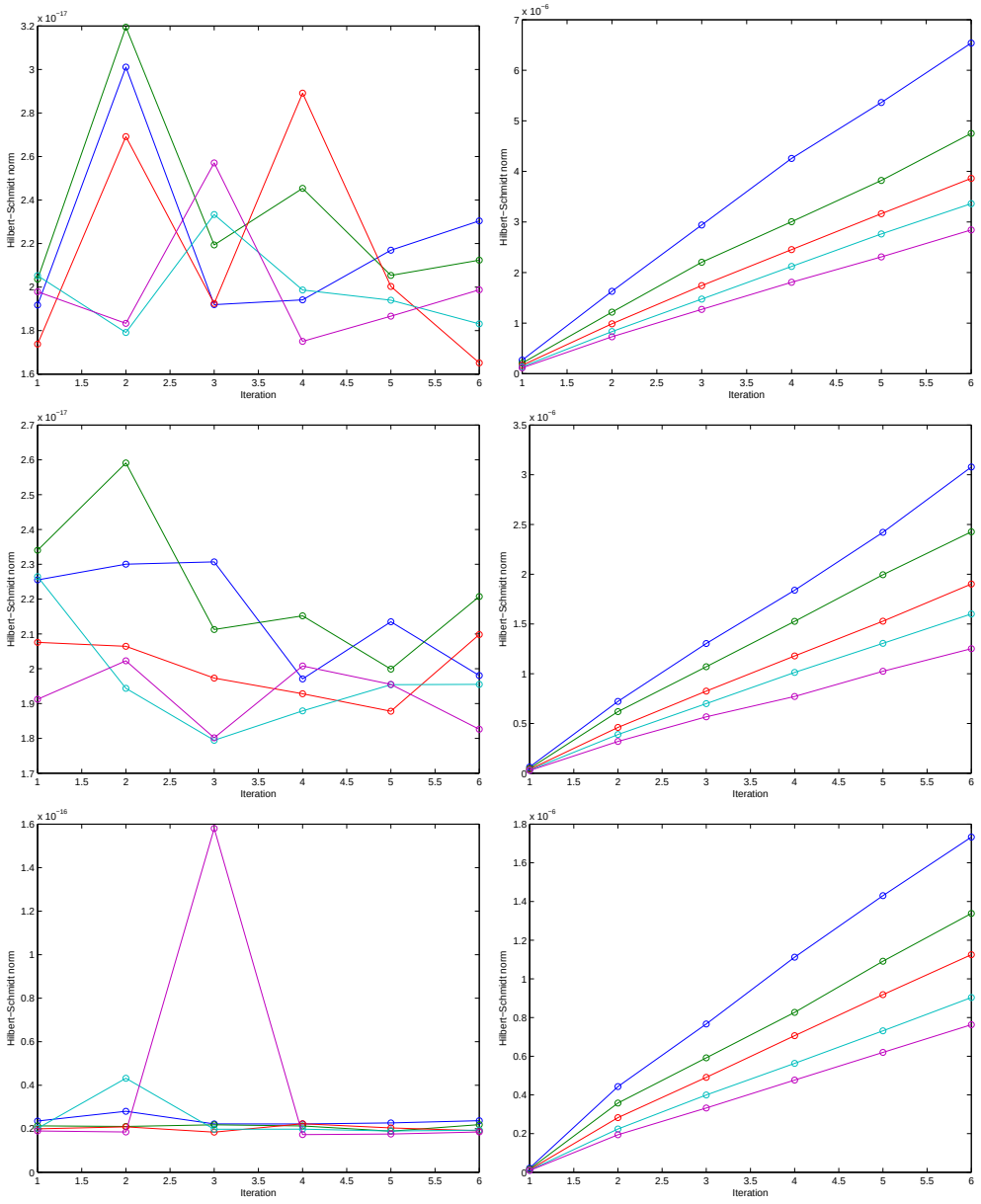

FIG. 10 – *The relative Hilbert-Schmidt norm of the error matrix in the estimation of* Λ*, and the Hilbert-Schmidt norm of the error matrix in the estimation of* $\Phi^* R_\nu \Phi$ *(left to right), for* $N = 121$*,* $N = 256$ *and* $N = 441$ *spatial locations (top to bottom). The blue line is associated with the truncation order corresponding to 50% of the trace norm of the empirical autocorrelation operator, green line for 60%, red line for 70%, light blue for 80%, and purple line for 90%*

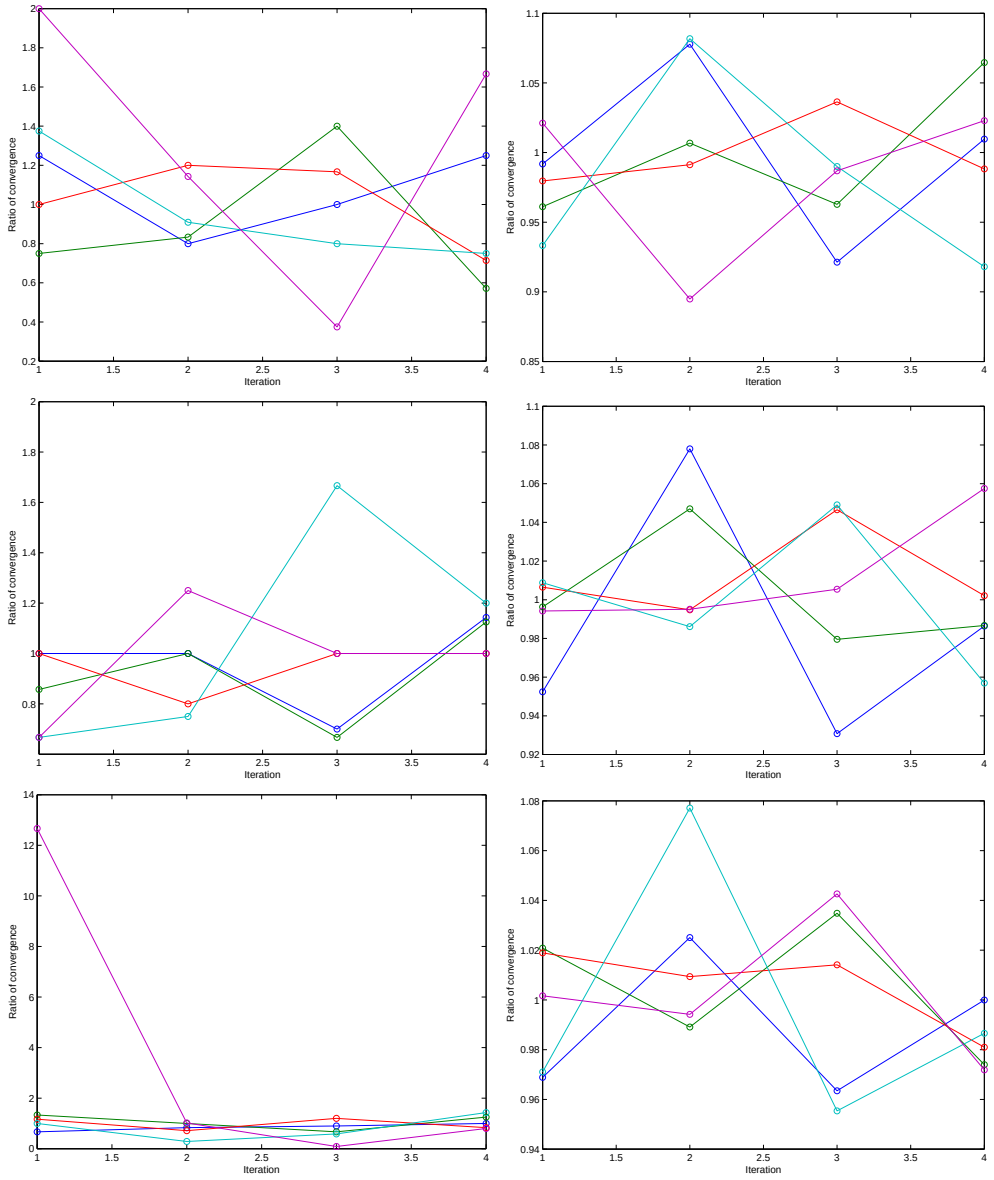

FIG. 11 – *Ratio of convergence in the functional estimation of* Λ *and* $\Phi^* R_\nu \Phi$ *(left to right), for* $N = 121$, $N = 256$ *and* $N = 441$ *spatial locations (top to bottom). The blue line represents the ratio of convergence for truncation order corresponding to 50% of the trace norm of the empirical autocorrelation operator, green line for 60%, red line for 70%, light blue for 80%, and purple line for 90%*

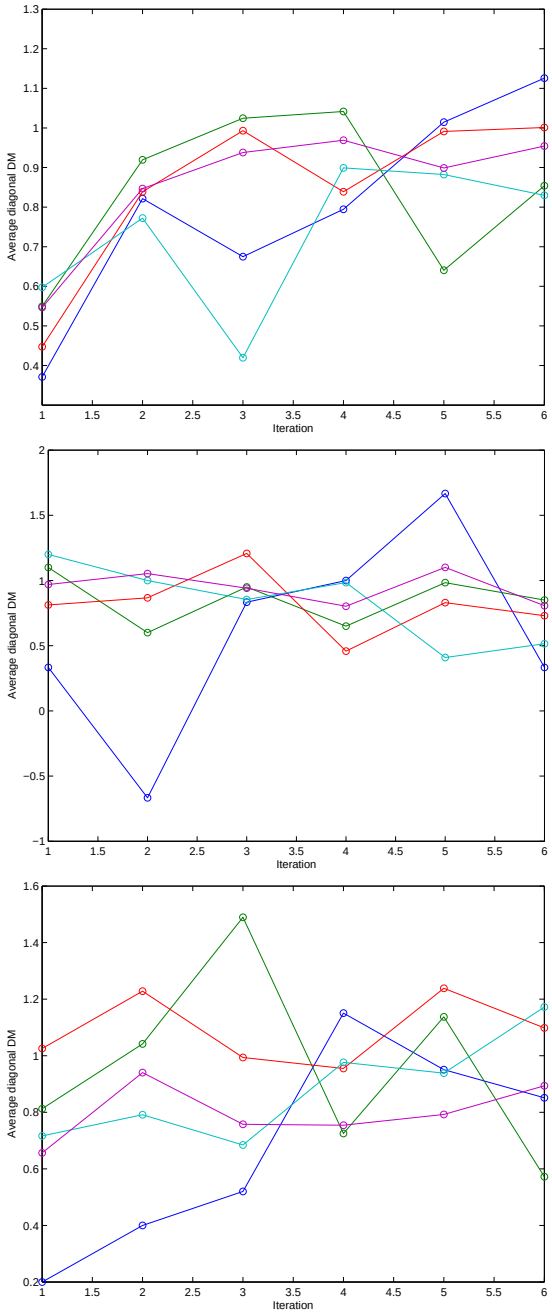

FIG. 12 – *Average of the diagonal elements of DM matrix, for $N = 121$, $N = 256$ and $N = 441$ spatial locations (top to bottom). The blue line represents the diagonal average for truncation order corresponding to 50%, green line for 60%, red line for 70%, light blue for 80%, and purple line for 90%*