Prediction of functional ARMA processes with an application to traffic data

Johannes Klepsch* Claudia Klüppelberg* Taoran Wei*

March 7, 2016

Abstract

We study the functional ARMA(p, q) and a corresponding approximating vector model, based on functional PCA. We investigate the structure of the multivariate vector process and derive conditions for the existence of a stationary solution to both the functional and the vector model equation. We then use the stationary vector process to predict the functional process, and compare the resulting predictor to the functional best linear predictor proposed by [3]. We derive bounds for the error due to dimension reduction. We conclude by applying functional ARMA processes for the modelling and prediction of highway traffic data.

AMS 2010 Subject Classifications: primary: 62M20 secondary: 60G25
Keywords: FPCA, functional ARMA process, functional principal component analysis, FTSA, functional time series analysis, prediction, projection, traffic data

1 Introduction

A *macroscopic highway traffic model* involves velocity, flow (number of vehicles passing a reference point per unit of time) and density (number of vehicles on a given road segment). The relation among these three variables can be depicted in diagrams of "velocity-flow relation" and "flow-density relation". The diagram of "flow-density relation" is also called *fundamental diagram of traffic flow* and can be used to determine the capacity of a road system and give guidance for inflow regulations or speed limits. Figures 1 and 2 depict the "velocity-flow relation" and "flow-density relation" for traffic data provided by the Autobahndirektion Südbayern. At a critical traffic density the state of flow will change from stable to unstable. In Figure 2, the critical traffic density for traffic on highway A92 in southern Bavaria is depicted.

In this paper we develop a statistical model for traffic data and apply it to the above data. As can be seen from Figure 4 and 5 the data show a certain pattern over the day,

*Center for Mathematical Sciences, Technical University of Munich, 85748 Garching, Boltzmannstrasse 3, Germany, e-mail: j.klepsch@tum.de , cklu@tum.de , taoran.wei@tum.de, http://www.statistics.tum.de

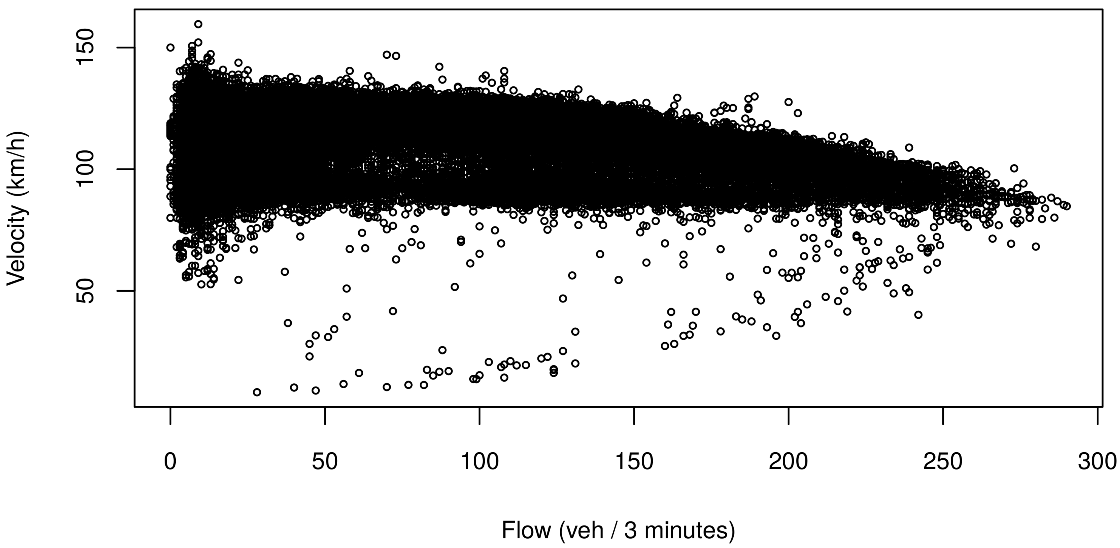

Figure 1: Velocity-flow relation on highway A92 in Southern Bavaria. Depicted are average velocities per 3 min versus number of vehicles within these 3 min during the period 01/01/2014 0:00 to 30/06/2014 23:59.

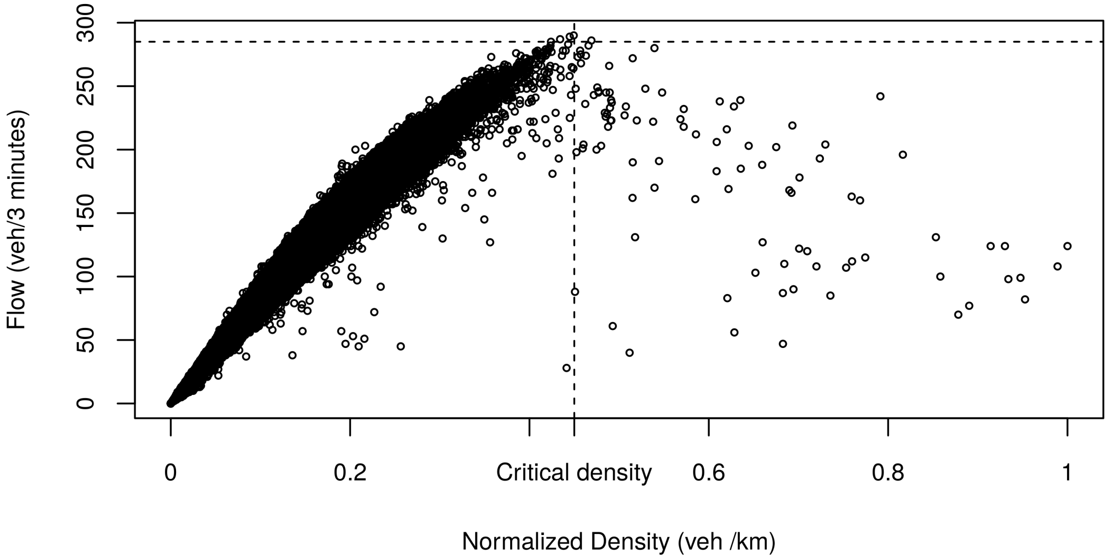

Figure 2: Flow-density relation for the data from Figure 1.

which we want to capture using tools from functional data analysis. The basic idea of functional data analysis is to represent the very high-dimensional raw data by a random function $X(\cdot)$, which in our case describes the traffic velocity over a day. In this paper we do not focus on the transformation of discrete to functional data. For theoretical considerations we work with data in functional form. For a sound introduction on the transformation from vector observations to functions, we refer to [15]. We want to assess temporal dependence between different days; i.e., our goal is a realistic time series model for functional data, which captures the day to day dependence. We hope that our analysis may support short term traffic regulation realised in real-time by electronic devices during the day, which could benefit from a more precise day-to-day prediction.

From a statistical point of view we are mainly interested in the prediction of a functional ARMA(p, q) process for arbitrary orders p, q. In scalar and multivariate time series analysis there exist several prediction methods which can be easily implemented like the Durbin-Levinson and the innovations algorithms (e.g see [6]). For functional time series, [3] has proposed the *functional best linear predictor* for a general linear process. However, implementation of the predictor is in general not feasible in practice, because explicit formulas of the predictor can not be derived. The class of functional AR(p) models is an exception. Functional autoregressive model of finite order are well studied (e.g. [3], Chapter 3) and allow for a elaborate prediction theory. Two well known approaches are presented in [3], Chapter 8 and [14]. The AR(1) model has also been applied for the prediction of traffic data in [2].

In [1] a prediction algorithm is proposed, which combines the idea of functional principal component analysis (FPCA) and functional time series analysis. The basic idea is to reduce the infinite-dimensional functional data by FPCA to finite-dimensional vector data. Thus, the prediction of the functional time series is transformed to the prediction of a multivariate time series. In [1] this algorithm is used to predict linear functional time series, with a focus on the functional AR(1) process for which bounds for the prediction error are derived.

In this paper we focus on functional ARMA(p, q) processes. In a first step we obtain a multivariate vector process by projection of the functional process X on the linear span of the d most important eigenfunctions of the covariance operator of X. We derive conditions such that the projected process follows a vector ARMA(p, q). If these assumptions do not hold, we show that the projected process can at least be approximated by a vector ARMA(p, q) process and assess the quality of the approximation. We then present conditions such that both functional and multivariate vector process have a unique stationary solution. This opens the way for prediction of functional ARMA(p, q) processes and we discuss relevant methods. The prediction algorithm of [1] can be applied, and makes sense under stationarity of the functional and the vector ARMA(p, q) process. We derive bounds for the prediction error based on the multivariate vector process in comparison to the functional best linear predictor derived by [3].

An extended simulation study can be found in [17], Chapter 5. It shows in particular that approximating the projection of functional ARMA processes by vector ARMA processes is reasonable. This is seen by comparing the model fit based on AIC and BIC criteria. The simulation study also yields a more detailed assessment of the quality of the functional predictor obtained by an extension of the algorithm [1] for different linear processes.

Our paper is organised as follows. In Section 2 we introduce the necessary Hilbert space theory and notation, which we use throughout. Here we present the Karhunen-Loéve Theorem and describe the FPCA based on the functional covariance operator. We also introduce the CVP method, which is used for truncation of the functional data. In Section 3 we turn to functional time series models with special emphasis on ARMA(p,q) processes. Section 3.1 is devoted to stationarity conditions for the functional ARMA(p,q) model. In Section 3.2 we study the multivariate vector process obtained by projection of the functional process on the linear span of the d most important eigenfunctions of the covariance operator of X. We investigate its stationarity and prove that the multivariate vector ARMA process approximates the functional ARMA process in a natural way. Section 4 investigates the prediction algorithm for functional ARMA(p,q) processes invoking the multivariate vector process and compares it to the functional best linear predictor. Finally, in Section 5 we apply our results to a traffic dataset of velocity measurements from 01/01/2014 to 30/06/2014 (obtained from the Autobahndirektion Südbayern) on a highway in Southern Bavaria, Germany.

2 Methodology

We summarize some concepts we shall use throughout. For details and more background see e.g. the monographs [3], [11] and [13]. Let $H = L^2([0,1])$ be the real separable Hilbert space of square integrable functions $x : [0,1] \to \mathbb{R}$ with norm $\|x\| = (\int_0^1 x^2(s)ds)^{1/2}$ generated by the inner product

$$\langle x, y\rangle := \int_0^1 x(t)y(t)dt, \quad x, y \in L^2([0,1]). \tag{2.1}$$

We shall often use Parseval's equality, which ensures that for a countable orthonormal basis $(e_i)_{i\in\mathbb{N}}$,

$$\langle x, y\rangle = \sum_{i=1}^{\infty} \langle x, e_i\rangle\langle e_i, y\rangle, \quad x, y \in H.$$

We denote by $\mathcal{L}$ the space of bounded linear operators acting on H. If not stated differently, we take the standard operator norm defined for a bounded operator $\Psi \in \mathcal{L}$ by $\|\Psi\|_{\mathcal{L}} := \sup_{\|x\|\le 1} \|\Psi(x)\|$.

A bounded linear operator Ψ is a *Hilbert-Schmidt* operator, if it is compact and for every orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of H,

$$\sum_{i=1}^{\infty} \|\Psi(e_i)\|^2 < \infty.$$

We denote by $\mathcal{S}$ the space of Hilbert-Schmidt operators acting on H, which is again a separable Hilbert space equipped with the following inner product and corresponding

Hilbert-Schmidt norm,

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} := \sum_{i=1}^{\infty} \langle \Psi_1(e_i), \Psi_2(e_i) \rangle \quad \text{and} \quad \|\Psi\|_{\mathcal{S}} := \sqrt{\langle \Psi, \Psi \rangle_{\mathcal{S}}} = \sqrt{\sum_{i=1}^{\infty} \|\Psi(e_i)\|^2} < \infty. \tag{2.2}$$

If Ψ is a Hilbert-Schmidt operator, then

$$\|\Psi\|_{\mathcal{L}} \le \|\Psi\|_{\mathcal{S}}. \tag{2.3}$$

Let $\mathcal{B}_H$ be the Borel σ-algebra of subsets of H. All random functions are defined on some probability space $(\Omega, \mathcal{A}, P)$ and are $\mathcal{A} - \mathcal{B}_H$-measurable. Then the space of square integrable random functions $L^2_H = L^2_H(\Omega, \mathcal{A}, P)$ is a Hilbert space with inner product $E\langle X, Y \rangle = E\int_0^1 X(s)Y(s)ds$ for $X, Y \in L^2_H$. We call such X an *H-valued random function*. For $X \in L^2_H$ the *functional mean* of X is defined as

$$\mu(t) := E[X(t)], \quad t \in [0,1]. \tag{2.4}$$

W.l.o.g. we will assume throughout that $\mu \equiv 0$.

Definition 2.1. The *covariance operator* C_X of X acts on H and is defined as

$$C_X : x \mapsto E\left[\langle X, x \rangle X\right], \quad x \in H. \tag{2.5}$$

More precisely,

$$C_X(x)(t) = E\left[\int_0^1 X(s)x(s)\mathrm{d}s\, X(t)\right] = \int_0^1 E\left[X(t)X(s)\right] x(s)ds \tag{2.6}$$

□

C_X is a symmetric, non-negative definite Hilbert-Schmidt operator with spectral representation

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j, \quad x \in H,$$

for eigenpairs $(\lambda_j, \nu_j)_{j\in\mathbb{N}}$, where $(\nu_j)_{j\in\mathbb{N}}$ is an orthonormal basis of H and $(\lambda_j)_{j\in\mathbb{N}}$ is a sequence of positive real numbers such that $\sum_{j=1}^{\infty} \lambda_j < \infty$. When considering eigendecompositions, we will assume that the λ_j are ordered decreasingly, hence $\lambda_i \ge \lambda_k$ for $i < k$. Every $X \in L^2_H$ can be represented as a linear combination of the eigenfunctions $(\nu_i)_{i\in\mathbb{Z}}$ of the covariance operator C_X. This is known as the *Karhunen-Loéve representation.*

Theorem 2.2 (Karhunen-Loéve Theorem)**.** *Suppose $X \in L^2_H$ with $EX = 0$, then*

$$X = \sum_{i=1}^{\infty} \langle X, \nu_i \rangle \nu_i, \tag{2.7}$$

where $(\nu_i)_{i\in\mathbb{Z}}$ are the orthonormal eigenfunctions of the covariance operator C defined in (2.5). The scalar products $(\langle X, \nu_i \rangle)_{i\in\mathbb{Z}}$ have mean-zero, variance λ_i and are uncorrelated; i.e., for all $i, j \in \mathbb{Z}$, $i \ne j$,

$$E\langle X, \nu_i \rangle = 0, \quad E(\langle X, \nu_i \rangle \langle X, \nu_j \rangle) = 0 \quad \textit{and} \quad E\langle X, \nu_i \rangle^2 = \lambda_i, \tag{2.8}$$

where $(\lambda_i)_{i\in\mathbb{Z}}$ are the eigenvalues of the covariance operator C_X.

The scalar products $(\langle X, \nu_i \rangle)_{i\in\mathbb{Z}}$ defined in (2.7) are called the *scores* of X. By the last equation in (2.8), we have

$$\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} E \langle X, \nu_j \rangle^2 = E\|X\|^2 < \infty, \quad X \in L_H^2. \tag{2.9}$$

Combining (2.8) and (2.9), every λ_j represents some proportion of the total variability of X.

Remark 2.3. [The CVP method] For any integer $d \in \mathbb{N}$, we consider the largest d eigenvalues $\lambda_1, \dots, \lambda_d$ of C_X. The *cumulative percentage of total variance* CPV(d) is defined as

$$\mathrm{CPV}(d) := \sum_{j=1}^{d} \lambda_j \Big/ \sum_{j=1}^{\infty} \lambda_j. \tag{2.10}$$

If we choose $d \in \mathbb{N}$ such that the $\mathrm{CPV}(d)$ exceeds a predetermined high value, then $\lambda_1, \dots, \lambda_d$ or the corresponding $\nu_1, \dots, \nu_d$ explain most of the variability of X. In this context $\nu_1, \dots, \nu_d$ are called the *functional principal components* (FPC's). If we project the H-valued random function X on the finite dimensional subspace of H, spanned by the d most important eigenfunctions $\nu_1, \dots, \nu_d$ of the covariance operator of X, which we denote by $\mathrm{sp}\{\nu_1, \dots, \nu_d\}$

$$X_d := \sum_{i=1}^{d} \langle X, \nu_i \rangle \nu_i, \tag{2.11}$$

then it contains most of the variability of X. □

3 Functional ARMA processes

In this section, we first introduce the functional ARMA(p, q) equations and derive sufficient conditions for the equations to have a stationary and causal solution, which we present. We then project the functional linear process on a finite dimensional subspace of H. We approximate this finite dimensional process by a suitable vector ARMA(p, q) process, and give conditions for the stationarity of this multivariate approximation. We also give conditions on the functional ARMA model such that the projection of the functional process on a finite dimensional space still follows an ARMA structure.

We start by defining functional white noise, which will be needed throughout the paper.

Definition 3.1. [[3], Definition 3.1] Let $(\varepsilon_n)_{n\in\mathbb{Z}}$ be a sequence of H-valued random variables.
(i) $(\varepsilon_n)_{n\in\mathbb{Z}}$ is *H-white noise* (WN) if for all $n \in \mathbb{Z}$, $E\varepsilon_n = 0$, $0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$, $C_{\varepsilon_n} = C_\varepsilon$, and if $C_{\varepsilon_n, \varepsilon_m}(\cdot) := E[\langle \varepsilon_m, \cdot \rangle \varepsilon_n] = 0$ for all $n \neq m$.
(ii) $(\varepsilon_n)_{n\in\mathbb{Z}}$ is *H-strong white noise* (SWN), if for all $n \in \mathbb{Z}$, $E\varepsilon_n = 0$, $0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$ and $(\varepsilon_n)_{n\in\mathbb{Z}}$ is i.i.d. □

We assume throughout that $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN with zero mean and $E\|\varepsilon_n^2\| = \sigma_\varepsilon^2 < \infty$. When SWN is required, this will be specified.

3.1 Stationary functional ARMA processes

Formally we can define a functional ARMA process of arbitrary order.

Definition 3.2. Let $(\varepsilon_n)_{n\in\mathbb{Z}}$ be WN as in Definition 3.1(i). Let furthermore $\phi_1,\dots,\phi_p$, $\theta_1,\dots,\theta_q \in \mathcal{L}$. Then a solution of

$$X_n = \sum_{i=1}^{p} \phi_i(X_{n-i}) + \sum_{j=1}^{q} \theta_j(\varepsilon_{n-j}) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{3.1}$$

is called a *functional* ARMA(p,q) *process.* □

We will derive conditions such that (3.1) has a stationary solution. We begin with the functional ARMA$(1,q)$ process, and need the following assumption.

Assumption 3.3. *There exists some $j_0 \in \mathbb{N}$ such that $\|\phi^{j_0}\|_{\mathcal{L}} < 1$.*

Theorem 3.4. *Under Assumption 3.3 there exists a unique stationary and causal solution to (3.1) for $p = 1$ given by*

$$\begin{aligned}
X_n &= \varepsilon_n + (\phi+\theta_1)\varepsilon_{n-1} + (\phi^2+\phi\theta_1+\theta_2)\varepsilon_{n-2} \\
&\quad + \cdots + (\phi^{q-1}+\phi^{q-2}\theta_1+\cdots+\theta_{q-1})\varepsilon_{n-(q-1)} \\
&\quad + \sum_{j=q}^{\infty} \phi^{j-q}(\phi^q+\phi^{q-1}\theta_1+\cdots+\theta_q)\varepsilon_{n-j} \\
&= \sum_{j=0}^{q-1}\Big(\sum_{k=0}^{j}\phi^{j-k}\theta_k\Big)\varepsilon_{t-j} + \sum_{j=q}^{\infty}\phi^{j-q}\Big(\sum_{k=0}^{q}\phi^{q-k}\theta_k\Big)\varepsilon_{t-j},
\end{aligned} \tag{3.2}$$

where $\phi^0 = I$ denotes the identity operator in H. Furthermore, the series in (3.2) converges almost surely and in L^2_H.

For the proof we need the following lemma.

Lemma 3.5 ([3], Lemma 3.1)**.** *For every $\phi \in \mathcal{L}$ the following two conditions are equivalent:*

(i) There exists an integer j_0 such that $\|\phi^{j_0}\|_{\mathcal{L}} < 1$.

(ii) There exist $a > 0$ and $0 < b < 1$ such that for every $j \ge 0$, $\|\phi^j\|_{\mathcal{L}} < ab^j$.

Proof of Theorem 3.4. We follow the lines of the proof of Prop. 3.1.1 of [6] and Theorem 3.1 in [3]. First we prove the mean square convergence of the series in (3.2). Take $m' > m \ge q$ and consider the truncated series

$$\begin{aligned}
X_n^{(m)} &:= \varepsilon_n + (\phi+\theta_1)\varepsilon_{n-1} + (\phi^2+\phi\theta_1+\theta_2)\varepsilon_{n-2} \\
&\quad + \cdots + (\phi^{q-1}+\phi^{q-2}\theta_1+\cdots+\theta_{q-1})\varepsilon_{n-(q-1)} \\
&\quad + \sum_{j=q}^{m} \phi^{j-q}(\phi^q+\phi^{q-1}\theta_1+\cdots+\theta_q)\varepsilon_{n-j}.
\end{aligned} \tag{3.3}$$

Define

$$\beta(\phi,\theta) := \phi^q + \phi^{q-1}\theta_1 + \cdots + \phi\theta_{q-1}\theta_q \in \mathcal{L}.$$

Then for all $m' > m \geq q$, using that $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN,

$$\begin{aligned}
E\|X_n^{(m')} - X_n^{(m)}\|^2 &= E\Big\| \sum_{j=m}^{m'} \phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j})\Big\|^2 \\
&= \sum_{j=m}^{m'}\sum_{k=m}^{m'} E\left\langle \phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j}), \phi^{k-q}(\beta(\phi,\theta)(\varepsilon_{n-k})\right\rangle \\
&= \sum_{j=m}^{m'} E\left\|\phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j})\right\|^2 \\
&\leq E\|\varepsilon_0\|^2 \sum_{j=m}^{m'} \left\|\phi^{j-q}\beta(\phi,\theta)\right\|_{\mathcal{L}}^2 \\
&\leq \sigma_\varepsilon^2 \sum_{j=m}^{m'} \left\|\phi^{j-q}\right\|_{\mathcal{L}}^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2.
\end{aligned}$$

Since $\phi \in \mathcal{L}$ satisfies Lemma 3.5(i), and equivalently (ii), we obtain

$$\sum_{j=0}^{\infty} \|\phi^j\|_{\mathcal{L}}^2 < \sum_{j=0}^{\infty} a^2 b^{2j} = \frac{a^2}{1-b^2} < \infty. \tag{3.4}$$

Using (3.4) we get

$$\sum_{j=m}^{m'} \left\|\phi^{j-q}\right\|_{\mathcal{L}}^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \ \sigma_\varepsilon^2 \leq \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \ \sigma_\varepsilon^2 a^2 \sum_{j=m}^{m'} b^{2(j-q)} \to 0, \quad \text{as } m, m' \to \infty.$$

By the Cauchy criterion, the series in (3.2) converges in mean square. To prove almost sure convergence we verify that

$$\sum_{j=1}^{\infty} \left\|\phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j})\right\| < \infty \quad \text{a.s.}$$

Since

$$E\Big(\sum_{j=1}^{\infty} \left\|\phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j})\right\|\Big)^2 \leq \Big(\sum_{j=1}^{\infty} \left\|\phi^{j-q}\right\|_{\mathcal{L}} \|\beta(\phi,\theta)\|_{\mathcal{L}}\Big)^2 E\|\varepsilon_0\|^2 = \sigma_\varepsilon^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \sum_{j=1}^{\infty} \left\|\phi^{j-q}\right\|_{\mathcal{L}}^2,$$

then by (3.4), we have

$$\sigma_\varepsilon^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \Big(\sum_{j=1}^{\infty} \|\phi^{j-q}\|_{\mathcal{L}}^2\Big)^2 = \sigma_\varepsilon^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \Big(\sum_{j=1}^{\infty} ab^{j-q}\Big)^2 = \sigma_\varepsilon^2 \|\beta(\phi,\theta)\|_{\mathcal{L}}^2 \frac{a^2}{(1-b)^2} < \infty.$$

Hence

$$E\Big(\sum_{j=1}^{\infty} \left\|\phi^{j-q}\beta(\phi,\theta)(\varepsilon_{n-j})\right\|\Big)^2 < \infty.$$

Thus we obtain a.s. convergence of the series in (3.2). Note that (3.2) is stationary, since its second order stucture only depends on $(\varepsilon_n)_{n\in\mathbb{Z}}$, which is WN.

In order to prove that (3.2) is a solution of (3.1) with $p = 1$, we plug (3.2) into (3.1), and obtain for $n \in \mathbb{Z}$,

$$X_n - \phi(X_{n-1}) = \sum_{j=0}^{q-1} (\sum_{k=0}^{j} \phi^{j-k}\theta_k)\varepsilon_{n-j} + \sum_{j=q}^{\infty} \phi^{j-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-j}$$

$$- \phi\Big(\sum_{j=0}^{q-1} (\sum_{k=0}^{j} \phi^{j-k}\theta_k)\varepsilon_{n-1-j} + \sum_{j=q}^{\infty} \phi^{j-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-1-j} \Big). \tag{3.5}$$

Now notice that the second term of the right-hand side can be written as

$$\begin{aligned}
&\phi\Big(\sum_{j=0}^{q-1} (\sum_{k=0}^{j} \phi^{j-k}\theta_k)\varepsilon_{n-1-j} + \sum_{j=q}^{\infty} \phi^{j-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-1-j} \Big)\\
&\quad = \sum_{j=0}^{q-1} (\sum_{k=0}^{j} \phi^{j+1-k}\theta_k)\varepsilon_{n-1-j} + \sum_{j=q}^{\infty} \phi^{j+1-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-1-j}\\
&\quad = \sum_{j'=1}^{q} (\sum_{k=0}^{j'-1} \phi^{j'-k}\theta_k)\varepsilon_{n-j'} + \sum_{j'=q+1}^{\infty} \phi^{j'-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-j'}\\
&\quad = \sum_{j'=1}^{q} (\sum_{k=0}^{j'} \phi^{j'-k}\theta_k - \phi^{j'-j'}\theta_{j'})\varepsilon_{n-j'} + \sum_{j'=q+1}^{\infty} \phi^{j'-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-j'}\\
&\quad = \sum_{j'=1}^{q} (\sum_{k=0}^{j'} \phi^{j'-k}\theta_k)\varepsilon_{n-j'} + \sum_{j'=q+1}^{\infty} \phi^{j'-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-j'} - \sum_{j'=1}^{q} \theta_{j'}\varepsilon_{n-j'}.
\end{aligned}$$

Hence, comparing the sums in (3.5), the only remaining terms are

$$\begin{aligned}
X_n - \phi(X_{n-1}) &= \varepsilon_n - \sum_{k=0}^{q} \phi^{q-k}\theta_k\varepsilon_{n-q} + \sum_{j'=1}^{q} \theta_{j'}\varepsilon_{n-j'} + \sum_{k=0}^{q} \phi^{q-k}\theta_k\varepsilon_{n-q}\\
&= \varepsilon_n + \sum_{j'=1}^{q} \theta_{j'}\varepsilon_{n-j'}, \quad n \in \mathbb{Z},
\end{aligned}$$

which shows that (3.2) is a solution of equation (3.1) with $p = 1$. Finally we prove the uniqueness of the solution. Assume that there is another stationary solution X'_n of (3.1). Iteration gives (cf. [16], eq. (4)) for all $r > q$,

$$\begin{aligned}
X'_n = &\sum_{j=0}^{q-1} (\sum_{k=0}^{j} \phi^{j-k}\theta_k)\varepsilon_{n-j} + \sum_{j=q}^{r-1} \phi^{j-q} (\sum_{k=0}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-j}\\
&+ \sum_{j=0}^{q-1} \phi^{r+j-q} (\sum_{k=j+1}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-(r+j)} + \phi^r X'_{n-r}
\end{aligned}$$

Therefore, with $X^{(r)}$ as in (3.3), for $r > q$,

$$\begin{aligned}
E\big\|X'_n - X_n^{(r)}\big\|^2 &= E\Big\| \sum_{j=0}^{q-1} \phi^{r+j-q} (\sum_{k=j+1}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-(r+j)} + \phi^r X'_{n-r} \Big\|^2\\
&\le 2E\Big\| \sum_{j=0}^{q-1} \phi^{r+j-q} (\sum_{k=j+1}^{q} \phi^{q-k}\theta_k)\varepsilon_{n-(r+j)} \Big\|^2 + 2\ E\,\|\phi^r(X'_{n-r})\|^2
\end{aligned}$$

Since $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN, and using the linearity of the operators

$$E\big\|X'_n - X_n^{(r)}\big\|^2 \le 2\|\phi^{r-q}\|_{\mathcal{L}}^2 \Big(\sum_{j=0}^{q-1} \|\phi^j\|_{\mathcal{L}} \Big)^2 \Big(\sum_{k=j+1}^{q} \|\phi^{q-k}\theta_k\|_{\mathcal{L}} \Big)^2 E\|\varepsilon_{n-(r+j)}\|^2 + 2\|\phi^r\|_{\mathcal{L}}^2 E\|(X'_{n-r})\|^2$$

By stationarity of $(X_n)_{n\in\mathbb{Z}}$ and $(\varepsilon_n)_{n\in\mathbb{Z}}$, the boundedness of ϕ and θ_j, $j = 1,\dots,q$ and by condition (ii) of Lemma 3.5,

$$E\left\|X'_n - X_n^{(r)}\right\|^2 \to 0, \quad r \to \infty.$$

Thus X'_n is equal in L^2_H to the limit of $X_n^{(r)}$, hence to X_n, which proves uniqueness. □

Remark 3.6. Spangenberg [16] derived a strictly stationary, not necessarily causal solution of a functional ARMA(p,q) equation in Banach spaces under minimal conditions. He thus extended known results considerably. □

For a functional ARMA(p,q) process we use the state space representation

$$\underbrace{\begin{pmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_{n-p+1} \end{pmatrix}}_{Y_n} = \underbrace{\begin{pmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ I & & & 0 \\ & \ddots & & \vdots \\ & & I & 0 \end{pmatrix}}_{\widetilde{\phi}} \underbrace{\begin{pmatrix} X_{n-1} \\ X_{n-2} \\ \vdots \\ X_{n-p} \end{pmatrix}}_{Y_{n-1}} + \sum_{j=0}^{q} \underbrace{\begin{pmatrix} \theta_j & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}}_{\widetilde{\theta}_j} \underbrace{\begin{pmatrix} \varepsilon_{n-j} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\delta_{n-j}}, \tag{3.6}$$

where $\theta_0 = I$, and I and 0 in (3.6) denote the indentity and zero operators, respectively. We summarize this as

$$Y_n = \widetilde{\phi}\,(Y_{n-1}) + \sum_{j=0}^{q} \widetilde{\theta}_j(\delta_{n-1}), \quad n \in \mathbb{Z}. \tag{3.7}$$

Since the X_n and ε_n take values in H, Y_n and δ_n take values in the product Hilbert space $H_p := (L^2([0,1]))^p$ with inner product and norm given by

$$\langle x,y\rangle_p := \sum_{j=1}^{p} \langle x_j, y_j\rangle, \quad \text{and} \quad \|x\|_p := \sqrt{\langle x,y\rangle_p}. \tag{3.8}$$

We denote by $\mathcal{L}(H^p)$ the space of bounded linear operators acting on H^p, the operator norm of $\widetilde{\phi} \in \mathcal{L}(H^p)$ is defined by $\|\widetilde{\phi}\|_{\mathcal{L}} := \sup_{\|x\|_p \le 1} \|\widetilde{\phi}(x)\|_p$.
$(\delta_n)_{n\in\mathbb{Z}}$ is WN in H_p. Let P_1 be the projection of H^p on H defined as

$$P_1(x_1,\dots,x_n) = x_1, \quad (x_1,\dots,x_n) \in H^p. \tag{3.9}$$

Assumption 3.7. *There exists some $j_0 \in \mathbb{N}$ such that $\widetilde{\phi}$ as in* (3.6) *satisfies $\|\widetilde{\phi}^{j_0}\|_{\mathcal{L}} < 1$.*

Since the proof of Theorem 3.4 works in arbitrary Hilbert spaces, using the state space representation of a functional ARMA(p,q) in H as a functional ARMA$(1,q)$ in H^p, we get the following theorem as a consequence of Theorem 3.4.

Theorem 3.8. *Under Assumption 3.7 there exists a unique stationary and causal solution to the functional* ARMA(p,q) *equations (3.1). The solution can be written as $X_n = P_1 Y_n$, where Y_n is the solution to the state space equation* (3.7)*, given by*

$$Y_n = \delta_n + (\widetilde{\phi} + \widetilde{\theta}_1)\delta_{n-1} + (\widetilde{\phi}^2 + \widetilde{\phi}\,\widetilde{\theta}_1 + \widetilde{\theta}_2)\delta_{n-2}$$

$$
\begin{aligned}
&+\cdots+(\widetilde{\phi}^{q-1}+\widetilde{\phi}^{q-2}\widetilde{\theta}_1+\cdots+\widetilde{\theta}_{q-1})\delta_{n-(q-1)}\\
&+\sum_{j=q}^{\infty}\widetilde{\phi}^{j-q}(\widetilde{\phi}^{q}+\widetilde{\phi}^{q-1}\widetilde{\theta}_1+\cdots+\widetilde{\theta}_j)\delta_{n-j},\\
=\ &\sum_{j=0}^{q-1}(\sum_{k=0}^{j}\widetilde{\phi}^{j-k}\widetilde{\theta}_k)\delta_{t-j}+\sum_{j=q}^{\infty}\widetilde{\phi}^{j-q}(\sum_{k=0}^{q}\widetilde{\phi}^{q-k}\widetilde{\theta}_k)\delta_{t-j} \qquad (3.10)
\end{aligned}
$$

where $\widetilde{\phi}^0$ denotes the identity operator in H^p and Y_n, δ_n, $\widetilde{\phi}$ and $\widetilde{\theta}_j$ are defined in (3.7). Furthermore, the series converges almost surely and in L^2_H.

3.2 The multivariate vector ARMA(p,q) process

We project the functional ARMA(p,q) process on a finite dimensional subspace of H, spanned by the d most important eigenfunctions $\nu_1,\dots,\nu_d$ of the covariance operator of X, which we denote by $\mathrm{sp}\{\nu_1,\dots,\nu_d\}$. With the CPV-method from Remark 2.3 we choose $d\in\mathbb{N}$ such that most of the variability of the stationary functional time series variables can be explained by $\nu_1,\dots,\nu_d$. Recalling the concept of functional principal components of (2.11) we consider the projection on $\mathrm{sp}\{\nu_1,\dots,\nu_d\}$

$$X_{n,d}=P_{\mathrm{sp}\{\nu_1,\dots,\nu_d\}}X_n=\sum_{i=1}^{d}\langle X_n,\nu_i\rangle\nu_i. \qquad (3.11)$$

In what follows, we are interested in

$$\mathbf{X}_n := (\langle X_n,\nu_1\rangle,\dots,\langle X_n,\nu_d\rangle)^\top. \qquad (3.12)$$

Due to its finite dimensionality $\mathbf{X}_n$ is isomorph to $X_{n,d}$.

Remark 3.9. We will in the following assume that the eigenfunctions of the covariance operators are known. In practice, this is of course not the case, and the eigenfunctions that show up in the following have to replaced by their empirical counterpart. Our results remain unchanged, except that we need stronger assumptions on the innovation process $(\varepsilon_n)_{n\in\mathbb{Z}}$ to ensure consistency of the estimators. For details on the estimation of covariance operators and their eigenelements in the case of dependent data we refer to [10]. □

A first result concerns the projection of the WN $(\varepsilon_n)_{n\in\mathbb{Z}}$ on $\mathrm{sp}\{\nu_1,\dots,\nu_d\}$, which we will need throughout.

Lemma 3.10. *Let $(e_i)_{i\in\mathbb{Z}}$ be an arbitrary orthonormal basis of H. For $d\in\mathbb{N}$, we define the d-dimensional vector process*

$$\mathbf{Z}_n := (\langle\varepsilon_n,e_1\rangle,\dots,\langle\varepsilon_n,e_d\rangle)^\top,\quad n\in\mathbb{Z}. \qquad (3.13)$$

(i) If $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN *as in Definition 3.1(i), then $(\mathbf{Z}_n)_{n\in\mathbb{N}}$ is d-dimensional* WN.
(ii) If $(\varepsilon_n)_{n\in\mathbb{Z}}$ is SWN *as in Definition 3.1(ii), then $(\mathbf{Z}_n)_{n\in\mathbb{N}}$ is d-dimensional* SWN.

As in Section 3.1 we start with the functional ARMA$(1,q)$ process for $q \in \mathbb{N}$ and are interested in the dynamics of $(X_{n,d})_{n\in\mathbb{Z}}$ of (3.11) for fixed $d \in \mathbb{N}$. For every $l \in \mathbb{Z}$, using the model equation (3.1) with $p = 1$, we get

$$\langle X_n, \nu_l \rangle = \langle \phi(X_{n-1}), \nu_l \rangle + \sum_{j=0}^{q} \langle \theta_j(\varepsilon_{n-j}), \nu_l \rangle, \quad l \in \mathbb{Z}. \tag{3.14}$$

For every l we expand $\langle \phi(X_{n-1}), \nu_l \rangle$, using that $(\nu_l)_{l\in\mathbb{Z}}$ is a ONB of H

$$\langle \phi(X_{n-1}), \nu_l \rangle = \Big\langle \phi\Big(\sum_{l'=1}^{\infty} \langle X_{n-1}, \nu_{l'} \rangle \nu_{l'} \Big), \nu_l \Big\rangle = \sum_{l'=1}^{\infty} \langle \phi(\nu_{l'}), \nu_l \rangle \langle X_{n-1}, \nu_{l'} \rangle,$$

and $\langle \theta_j(\varepsilon_{n-j}), \nu_l \rangle$ for $j = 1, \dots, q$ as

$$\langle \theta_j(\varepsilon_{n-j}), \nu_l \rangle = \Big\langle \theta_j\Big(\sum_{l'=1}^{\infty} \langle \varepsilon_{n-j}, \nu_{l'} \rangle \nu_{l'} \Big), \nu_l \Big\rangle = \sum_{l'=1}^{\infty} \langle \theta_j(\nu_{l'}), \nu_l \rangle \langle \varepsilon_{n-j}, \nu_{l'} \rangle.$$

In order to derive the d-dimensional vector process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, for notational ease, we restrict a precise presentation to the ARMA$(1,1)$ model. The presentation of the ARMA$(1,q)$ model is an obvious extension.

For $q = 1$, with $\theta_0 = I$ and $\theta_1 = \theta$, in matrix form (3.14) is given by

$$\left(\begin{array}{c} \langle X_n, \nu_1 \rangle \\ \vdots \\ \langle X_n, \nu_d \rangle \\ \hline \langle X_n, \nu_{d+1} \rangle \\ \vdots \end{array}\right) = \left[\begin{array}{ccc|cc} \langle \phi(\nu_1), \nu_1 \rangle & \dots & \langle \phi(\nu_d), \nu_1 \rangle & \langle \phi(\nu_{d+1}), \nu_1 \rangle & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots \\ \langle \phi(\nu_1), \nu_d \rangle & \dots & \langle \phi(\nu_d), \nu_d \rangle & \langle \phi(\nu_{d+1}), \nu_d \rangle & \dots \\ \hline \langle \phi(\nu_1), \nu_{d+1} \rangle & \dots & \langle \phi(\nu_d), \nu_{d+1} \rangle & \langle \phi(\nu_{d+1}), \nu_{d+1} \rangle & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right] \left(\begin{array}{c} \langle X_{n-1}, \nu_1 \rangle \\ \vdots \\ \langle X_{n-1}, \nu_d \rangle \\ \hline \langle X_{n-1}, \nu_{d+1} \rangle \\ \vdots \end{array}\right)$$
$$+ \left(\begin{array}{c} \langle \varepsilon_n, \nu_1 \rangle \\ \vdots \\ \langle \varepsilon_n, \nu_d \rangle \\ \hline \langle \varepsilon_n, \nu_{d+1} \rangle \\ \vdots \end{array}\right) + \left[\begin{array}{ccc|cc} \langle \theta(\nu_1), \nu_1 \rangle & \dots & \langle \theta(\nu_d), \nu_1 \rangle & \langle \theta(\nu_{d+1}), \nu_1 \rangle & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots \\ \langle \theta(\nu_1), \nu_d \rangle & \dots & \langle \theta(\nu_d), \nu_d \rangle & \langle \theta(\nu_{d+1}), \nu_d \rangle & \dots \\ \hline \langle \theta(\nu_1), \nu_{d+1} \rangle & \dots & \langle \theta(\nu_d), \nu_{d+1} \rangle & \langle \theta(\nu_{d+1}), \nu_{d+1} \rangle & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right] \left(\begin{array}{c} \langle \varepsilon_{n-1}, \nu_1 \rangle \\ \vdots \\ \langle \varepsilon_{n-1}, \nu_d \rangle \\ \hline \langle \varepsilon_{n-1}, \nu_{d+1} \rangle \\ \vdots \end{array}\right) \tag{3.15}$$

We simplify the notation in (3.15) by summarizing vectors and matrices to

$$\left(\begin{array}{c} \mathbf{X}_n \\ \hline \mathbf{X}_n^\infty \end{array}\right) = \left[\begin{array}{c|c} \mathbf{\Phi} & \mathbf{\Phi}^\infty \\ \hline \vdots & \vdots \end{array}\right] \left(\begin{array}{c} \mathbf{X}_{n-1} \\ \hline \mathbf{X}_{n-1}^\infty \end{array}\right) + \left(\begin{array}{c} \mathbf{E}_n \\ \hline \mathbf{E}_n^\infty \end{array}\right) + \left[\begin{array}{c|c} \mathbf{\Theta} & \mathbf{\Theta}^\infty \\ \hline \vdots & \vdots \end{array}\right] \left(\begin{array}{c} \mathbf{E}_{n-1} \\ \hline \mathbf{E}_{n-1}^\infty \end{array}\right), \tag{3.16}$$

where

$$\begin{aligned} \mathbf{E}_n &:= (\langle \varepsilon_n, \nu_1 \rangle, \dots, \langle \varepsilon_n, \nu_d \rangle)^\top, \\ \mathbf{X}_n^\infty &:= (\langle X_n, \nu_{d+1} \rangle, \dots)^\top, \\ \mathbf{E}_n^\infty &:= (\langle \varepsilon_n, \nu_{d+1} \rangle, \dots)^\top. \end{aligned}$$

The operators $\mathbf{\Phi}$ and $\mathbf{\Theta}$ in (3.16) are $d \times d$ matrices with entries $\langle \phi(\nu_{l'}), \nu_l \rangle$ and $\langle \theta(\nu_{l'}), \nu_l \rangle$ in the l-th row and l'-th column, respectively. $\mathbf{\Phi}^\infty$ and $\mathbf{\Theta}^\infty$ are $d \times \infty$ matrices with ll'-th entries $\langle \phi(\nu_{l'+d}), \nu_l \rangle$ and $\langle \theta(\nu_{l'+d}), \nu_l \rangle$, respectively.

By (3.16), we write the d-dimensional vector equation

$$\mathbf{X}_n = \mathbf{\Phi}\mathbf{X}_{n-1} + \mathbf{E}_n + \mathbf{\Theta}\mathbf{E}_{n-1} + \mathbf{\Delta}_{n-1}, \quad n \in \mathbb{Z}, \tag{3.17}$$

where

$$\mathbf{\Delta}_{n-1} := \mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1} + \mathbf{\Theta}^\infty \mathbf{E}^\infty_{n-1}. \tag{3.18}$$

By Lemma 3.10 $(\mathbf{E}_n)_{n\in\mathbb{Z}}$ is d-dimensional WN. Note that $\mathbf{\Delta}_{n-1}$ in (3.18) is a d-dimensional vector with l-th component

$$(\mathbf{\Delta}_{n-1})_l = \sum_{l'=d+1}^{\infty} \langle \phi(\nu_{l'}), \nu_l \rangle \langle X_{n-1}, \nu_{l'} \rangle + \sum_{l'=d+1}^{\infty} \langle \theta(\nu_{l'}), \nu_l \rangle \langle \varepsilon_{n-1}, \nu_{l'} \rangle. \tag{3.19}$$

Thus, the "error term" $\mathbf{\Delta}_{n-1}$ depends on X_{n-1}, and the vector process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ in (3.17) is in general not a vector ARMA$(1,1)$ process with innovations $(\mathbf{E}_n)_{n\in\mathbb{Z}}$. However, we can use a vector ARMA model as an approximation to $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, where we can make $\mathbf{\Delta}_{n-1}$ arbitrarily small by increasing the dimension d.

Lemma 3.11. *Let $\|\cdot\|_2$ denote the Euclidean norm in $\mathbb{R}^d$, and let the d-dimensional vector $\mathbf{\Delta}_{n-1}$ be defined as in (3.18). Then $E\|\mathbf{\Delta}_{n-1}\|_2^2$ is bounded and tends to 0 as $d \to \infty$.*

Proof. Using (3.18) we obtain

$$E\|\mathbf{\Delta}_{n-1}\|_2^2 = E\|\mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1} + \mathbf{\Theta}^\infty \mathbf{E}^\infty_{n-1}\|_2^2 \le 2\left(E\|\mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1}\|_2^2 + E\|\mathbf{\Theta}^\infty \mathbf{E}^\infty_{n-1}\|_2^2\right). \tag{3.20}$$

We estimate the two parts $E\|\mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1}\|_2^2$ and $E\|\mathbf{\Theta}^\infty \mathbf{E}^\infty_{n-1}\|_2^2$ separately. By (3.19), we have (using Parseval's identity in the third line),

$$\begin{aligned}
E\|\mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1}\|_2^2 &= E\Big[\sum_{l=1}^{d}\Big(\sum_{l'=d+1}^{\infty} \langle \phi(\nu_{l'})\langle X_{n-1}, \nu_{l'}\rangle, \nu_l \rangle\Big)^2\Big] \\
&\le E\Big[\sum_{l=1}^{\infty}\Big\langle \sum_{l'=d+1}^{\infty} \phi(\nu_{l'})\langle X_{n-1}, \nu_{l'}\rangle, \nu_l \Big\rangle^2\Big] \\
&= E\Big\| \sum_{l'=d+1}^{\infty} \langle X_{n-1}, \nu_{l'}\rangle \phi(\nu_{l'})\Big\|^2 \\
&= E\Big\langle \sum_{l=d+1}^{\infty} \langle X_{n-1}, \nu_l\rangle \phi(\nu_l), \sum_{l'=d+1}^{\infty} \langle X_{n-1}, \nu_{l'}\rangle \phi(\nu_{l'})\Big\rangle.
\end{aligned}$$

By the Karhunen-Loéve Theorem (Theorem 2.2) the scores $(\langle X_{n-1,l}, \nu_l\rangle)_{l\in\mathbb{Z}}$ are uncorrelated. Thus,

$$E\|\mathbf{\Phi}^\infty \mathbf{X}^\infty_{n-1}\|_2^2 \le E\left[\sum_{l'=d+1}^{\infty} \langle X_{n-1}, \nu_{l'}\rangle^2 \|\phi(\nu_{l'})\|^2\right] = \sum_{l'=d+1}^{\infty} E\left(\langle X_{n-1}, \nu_{l'}\rangle\right)^2 \|\phi(\nu_{l'})\|^2.$$

Since by (2.8) we have $E\langle X_{n-1},\nu_{l'}\rangle^2 = \lambda_{l'}$, we get

$$\sum_{l'=d+1}^{\infty} E\left(\langle X_{n-1},\nu_{l'}\rangle\right)^2 \|\phi(\nu_{l'})\|^2 = \sum_{l'=d+1}^{\infty} \lambda_{l'}\|\phi\|_{\mathcal{L}}^2\|\nu_{l'}\|^2 \le \|\phi\|_{\mathcal{L}}^2 \sum_{l'=d+1}^{\infty} \lambda_{l'}. \tag{3.21}$$

The bound for $E\|\mathbf{\Theta}^\infty\mathbf{E}_{n-1}^\infty\|_2^2$ can be obtained in exactly the same way, and we get

$$\begin{aligned} E\|\mathbf{\Theta}^\infty\mathbf{E}_{n-1}^\infty\|_2^2 &\le \sum_{l'=d+1}^{\infty} E\langle\varepsilon_{n-1},\nu_{l'}\rangle^2 \|\theta(\nu_{l'})\|^2 \le \|\theta\|_{\mathcal{L}}^2 \sum_{l'=d+1}^{\infty} E\langle\langle\varepsilon_{n-1},\nu_{l'}\rangle\varepsilon_{n-1},\nu_{l'}\rangle \\ &= \|\theta\|_{\mathcal{L}}^2 \sum_{l'=d+1}^{\infty}\langle C_\varepsilon(\nu_{l'}),\nu_{l'}\rangle, \end{aligned} \tag{3.22}$$

where C_ε is the covariance operator of the WN. As a covariance operator, it has finite nuclear operator norm $\|C_\varepsilon\|_{\mathcal{N}} := \sum_{l'=1}^{\infty}\langle C_\varepsilon(\nu_{l'}),\nu_{l'}\rangle < \infty$. Hence, $\sum_{l'=d+1}^{\infty}\langle C_\varepsilon(\nu_{l'}),\nu_{l'}\rangle \to 0$ for $d\to\infty$. Combining (3.20), (3.21) and (3.22) we find that $E\|\mathbf{\Delta}_{n-1}\|_2^2$ is bounded and tends to 0 as $d\to\infty$. □

The proof of bounding $E\|\mathbf{\Delta}_{n-1}\|_2^2$ is analogous in the ARMA$(1,q)$ case. We now summarize our findings in the case of a functional ARMA$(1,q)$ process.

Theorem 3.12. *Consider a functional* ARMA$(1,q)$ *process such that Assumption 3.3 holds. For $d\in\mathbb{N}$ the vector process $\mathbf{X}_n := (\langle X_n,\nu_1\rangle,\dots,\langle X_n,\nu_d\rangle)^\top$ has the representation*

$$\mathbf{X}_n = \mathbf{\Phi}\mathbf{X}_{n-1} + \mathbf{E}_n + \sum_{j=1}^{q}\mathbf{\Theta}_q\mathbf{E}_{n-1} + \mathbf{\Delta}_{n-1}, \quad n\in\mathbb{Z}, \tag{3.23}$$

where

$$\mathbf{\Delta}_{n-1} := \mathbf{\Phi}^\infty\mathbf{X}_{n-1}^\infty + \sum_{j=1}^{q}\mathbf{\Theta}_j^\infty\mathbf{E}_{n-j}$$

and all quantities are defined analogously to (3.12), (3.17), *and* (3.18). *Define*

$$\check{\mathbf{X}}_n = \mathbf{\Phi}\check{\mathbf{X}}_{n-1} + \mathbf{E}_n + \sum_{j=1}^{q}\mathbf{\Theta}_j\mathbf{E}_{n-j}, \quad n\in\mathbb{Z}. \tag{3.24}$$

Then both the functional ARMA$(1,q)$ *process* $(X_n)_{n\in\mathbb{Z}}$ *in* (3.1) *and the d-dimensional vector process* $(\check{\mathbf{X}}_n)_{n\in\mathbb{Z}}$ *in* (3.24) *have a unique stationary and causal solution.*

Moreover $E\|\mathbf{\Delta}_{n-1}\|_2^2$ is bounded and tends to 0 as $d\to\infty$.

Proof. Recall that the $d\times d$ matrix $\mathbf{\Phi}$ of the vector process (3.24) (see (3.15) and (3.16)) has representation

$$\mathbf{\Phi} = \begin{pmatrix} \langle\phi(\nu_1),\nu_1\rangle & \dots & \langle\phi(\nu_d),\nu_1\rangle \\ \vdots & \ddots & \vdots \\ \langle\phi(\nu_1),\nu_d\rangle & \dots & \langle\phi(\nu_d),\nu_d\rangle \end{pmatrix}.$$

In order to show that (3.24) has a stationary solution, by Theorem 11.3.1 of [6], it suffices to prove that every eigenvalue λ_k of $\mathbf{\Phi}$ with corresponding eigenvector $\mathbf{a}_k = (\mathbf{a}_{k,1},\dots,\mathbf{a}_{k,d})$, $k=1,\dots,d$ of $\mathbf{\Phi}$ satisfies $|\lambda_k|<1$. Note that $|\lambda_k|<1$ is equivalent to $|\lambda_k^{j_0}|<1$, for all

$j_0 \in \mathbb{N}$. Let $a_k = \mathbf{a}_{k,1}\nu_1 + \cdots + \mathbf{a}_{k,d}\nu_d$, and $\|a_k\| = \sum_{l=1}^d \langle a_{k,\nu_l}\rangle^2 = \sum_{l=1}^d \mathbf{a}_{k,l}^2 = \|\mathbf{a}_k\|_2 = 1$ for all $1 \le k \le d$. With the orthogonality of $\nu_1, \ldots, \nu_d$, $\|\Phi \mathbf{a}_k\|_2^2 = \sum_{l=1}^d \big(\sum_{l'=1}^d \langle \phi\nu_{l'}, \nu_l\rangle \mathbf{a}_{k,l}\big)^2$ and, defining $A_d = \{\nu_1, \ldots, \nu_d\}$, we calculate

$$
\begin{aligned}
\|P_{A_d}\phi P_{A_d} a_k\|^2 &= \|\sum_{l=1}^d \langle \phi P_{A_d} a_k, \nu_l\rangle \nu_l\|^2 \\
&= \sum_{l=1}^d \Big\langle \phi(\sum_{l'=1}^d \mathbf{a}_{k,l'}\nu_{l'}), \nu_l\Big\rangle^2 \|\nu_l\|^2 \\
&= \sum_{l=1}^d \Big(\sum_{l'=1}^d \mathbf{a}_{k,l'} \langle \phi\nu_{l'}, \nu_l\rangle\Big)^2 = \|\Phi \mathbf{a}_k\|_2^2
\end{aligned}
$$

Hence, for j_0 as in Assumption 3.3,

$$
|\lambda_k^{j_0}| = \|\lambda_k^{j_0}\mathbf{a}_k\|_2 = \|\Phi^{j_0} a_k\|_2 = \left\|\big(P_{A_d}\phi P_{A_d}\big)^{j_0}(a_k)\right\| \le \left\|\big(P_{A_d}\phi P_{A_d}\big)^{j_0}\right\|_{\mathcal{L}} \|a_k\| \le \|\phi^{j_0}\|_{\mathcal{L}} < 1,
$$

which finishes the proof. □

In order to extend approximation (3.24) of a functional ARMA$(1,q)$ process to a functional ARMA(p,q) process we use again the state space representation (3.7) given by

$$
Y_n = \widetilde{\phi}(Y_{n-1}) + \sum_{j=0}^q \widetilde{\theta}_j(\delta_{n-j}), \quad n \in \mathbb{Z},
$$

where $\widetilde{\theta}_0 = I$, Y_n, $\widetilde{\phi}$, $\widetilde{\theta}$ and δ_n are defined as in Theorem 3.8 and take values in $H_p = (L^2([0,1]))^p$; see (3.8).

Theorem 3.13. *Consider the functional* ARMA(p,q) *process as defined in* (3.1) *such that Assumption 3.7 holds. Then for* $d \in \mathbb{N}$ *the vector process*

$$
\mathbf{X}_n := (\langle X_n, \nu_1\rangle, \ldots, \langle X_n, \nu_d\rangle)^\top \tag{3.25}
$$

has representation

$$
\mathbf{X}_n = \sum_{i=1}^p \mathbf{\Phi}_i \mathbf{X}_{n-i} + \mathbf{E}_n + \sum_{j=1}^q \mathbf{\Theta}_q \mathbf{E}_{n-j} + \mathbf{\Delta}_{n-1}, \quad n \in \mathbb{Z}, \tag{3.26}
$$

where

$$
\mathbf{\Delta}_{n-1} := \sum_{i=1}^p \mathbf{\Phi}_i^\infty \mathbf{X}_{n-i}^\infty + \sum_{j=1}^q \mathbf{\Theta}_j^\infty \mathbf{E}_{n-j}
$$

and all quantities are defined analogously to (3.12), (3.17), *and* (3.18). *Define*

$$
\check{\mathbf{X}}_n = \sum_{i=1}^p \mathbf{\Phi}_i \check{\mathbf{X}}_{n-i} + \mathbf{E}_n + \sum_{j=1}^q \mathbf{\Theta}_q \mathbf{E}_{n-1}, \quad n \in \mathbb{Z}. \tag{3.27}
$$

Then both the functional ARMA(p,q) *process* $(X_n)_{n\in\mathbb{Z}}$ *in* (3.1) *and the d-dimensional vector process* $(\check{\mathbf{X}}_n)_{n\in\mathbb{Z}}$ *in* (3.27) *have a unique stationary and causal solution.*

Moreover $E\|\mathbf{\Delta}_{n-1}\|_2^2$ *is bounded and tends to 0 as* $d \to \infty$.

We are now interested in conditions for $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ actually following a vector ARMA(p,q) model. A trivial condition is that the projection of ϕ_i and θ_j on $A_d^\perp$, the orthogonal complement of $A_d = \mathrm{sp}\{\nu_1,\dots,\nu_d\}$, satisfies

$$P_{A_d^\perp}\phi_i P_{A_d^\perp} \quad = \quad P_{A_d^\perp}\theta_j P_{A_d^\perp} = 0 \tag{3.28}$$

for all $i = 1,\dots,p$ and $j = 1,\dots,q$. Then the vector process $\check{\mathbf{X}}_n \equiv \mathbf{X}_n$ for all $n \in \mathbb{Z}$.

However, as we show next, assumptions on the moving average parameters are actually not required. We start with a well known result that characterises vector MA processes.

Lemma 3.14 ([6], Proposition 3.2.1)**.** *If $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ is a stationary vector process with autocovariance function $\mathbf{C}_{X_h,X_0} = E[\mathbf{X}_h\mathbf{X}_0^\top]$ with $\mathbf{C}_{X_q,X_0} \neq 0$ and $\mathbf{C}_{X_h,X_0} = 0$ for $|h| > q$, then $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ is a vector* MA(q)*.*

Proposition 3.15. *Denote again by $A_d := \mathrm{sp}\{\nu_1,\dots,\nu_\mathrm{d}\}$, and by $A_d^\perp$ its orthogonal complement. If $P_{A_d^\perp}\phi_i P_{A_d^\perp} = 0$ for all $i = 1,\dots,p$, then the d-dimensional process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ in (3.26) is a vector* ARMA(p,q) *process.*

Proof. Since ϕ_i for $i = 1,\dots,p$ only acts on A_d, from (3.26) we get

$$\begin{aligned}\mathbf{X}_n &= \sum_{i=1}^p \mathbf{\Phi}_i\mathbf{X}_{n-i} + \mathbf{E}_n + \sum_{j=1}^q \mathbf{\Theta}_j\mathbf{E}_{n-j} + \mathbf{\Delta}_{n-1}\\ &= \sum_{i=1}^p \mathbf{\Phi}_i\mathbf{X}_{n-i} + \mathbf{E}_n + \sum_{j=1}^q \mathbf{\Theta}_j\mathbf{E}_{n-j} + \sum_{j=1}^q \mathbf{\Theta}_j^\infty\mathbf{E}_{n-j}^\infty, \quad n \in \mathbb{Z}.\end{aligned}$$

Hence, in order to show that $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ follows an ARMA(p,q) process, we have to show that

$$\mathbf{Y}_n := \mathbf{E}_n + \sum_{j=1}^q \mathbf{\Theta}_j\mathbf{E}_{n-j} + \sum_{j=1}^q \mathbf{\Theta}_j^\infty\mathbf{E}_{n-j}^\infty, \quad n \in \mathbb{Z},$$

follows an vector MA(q) model. According to Lemma 3.14, it is sufficient to show that $(\mathbf{Y}_n)_{n\in\mathbb{Z}}$ is stationary and has an appropriate autocovariance structure. Defining (with $\theta_0 = I$)

$$Y_n := \sum_{j=0}^q \theta_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z},$$

observe that $\mathbf{Y}_n = (\langle Y_n,\nu_1\rangle,\dots,\langle Y_n,\nu_d\rangle)$ is isomorph to $P_{A_d}Y_n = \sum_{j=1}^d\langle Y_n,\nu_j\rangle\nu_j$ for all $n \in \mathbb{Z}$. Hence, stationarity of $(\mathbf{Y}_n)_{n\in\mathbb{Z}}$ immediately follows from the stationarity of $(Y_n)_{n\in\mathbb{Z}}$. Furthermore,

$$E[P_{A_d}Y_h\langle P_{A_d}Y_0,\cdot\rangle] = P_{A_d}E[Y_h\langle Y_0,\cdot\rangle]P_{A_d} = P_{A_d}C_{Y_h,Y_0}P_{A_d}.$$

But since $(Y_n)_{n\in\mathbb{Z}}$ is a functional MA(q) process, $C_{Y_h,Y_0} = 0$ for $|h| > q$. Due to the relation between $P_{A_d}Y_n$ and $\mathbf{Y}_n$, we also have $\mathbf{C}_{Y_h,Y_0} = 0$ for $|h| > q$ and, hence, $(\mathbf{Y}_n)_{n\in\mathbb{Z}}$ is a vector MA(q). □

4 Prediction of functional ARMA process

We derive the best linear predictor of a functional ARMA(p,q) process $(X_n)_{n\in\mathbb{Z}}$ based on $\mathbf{X}_1,\dots,\mathbf{X}_n$, defined as in (3.26). We then compare the vector best linear predictor to the functional best linear predictor based on $X_1,\dots,X_n$ and show that, under regularity conditions, the difference is bounded and tends to 0 as d tends to infinity.

4.1 Prediction based on the vector process

In finite dimensions, the concept of a best linear predictor is well studied. For a d-dimensional stationary time series $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ we denote the "matrix linear span" of $\mathbf{X}_1,\dots,\mathbf{X}_n$ by

$$\mathbf{M}_1 := \Big\{ \sum_{i=1}^{n} \mathbf{A}_{ni}\mathbf{X}_i : \ \mathbf{A}_{ni} \text{ are real } d\times d \text{ matrices}, i=1,\dots,n \Big\}. \tag{4.1}$$

Then the *vector best linear predictor* $\hat{\mathbf{X}}_{n+1}$ of $\mathbf{X}_{n+1}$ based on $\mathbf{X}_1,\dots,\mathbf{X}_n$ is defined as the projection of $\mathbf{X}_{n+1}$ on $\mathbf{M}_1$; i.e.,

$$\widehat{\mathbf{X}}_{n+1} := P_{\mathbf{M}_1}\mathbf{X}_{n+1}. \tag{4.2}$$

Its properties are given by the projection theorem (e.g. Theorem 2.3.1 of [6]) and are summarized as follows.

Remark 4.1. Recall that $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$ and $\langle\ ,\ \rangle_{\mathbb{R}^d}$ the corresponding scalar product.
(i) $E\langle \mathbf{X}_{n+1}-\hat{\mathbf{X}}_{n+1},\mathbf{Y}\rangle_{\mathbb{R}^d} = \mathbf{0}$ for all $\mathbf{Y}\in\mathbf{M}_1$.
(ii) $\hat{\mathbf{X}}_{n+1}$ is the unique element in $\mathbf{M}_1$ such that $E\|\mathbf{X}_{n+1}-\hat{\mathbf{X}}_{n+1}\|_2^2 = \inf_{\mathbf{Y}\in\mathbf{M}_1} E\|\mathbf{Y}_{n+1}-\mathbf{Y}\|_2^2$.
(iii) $\mathbf{M}_1$ is a linear subspace of $\mathbb{R}^d$. □

In analogy to the prediction algorithm suggested in [1], a method for finding the best linear predictor of X_{n+1} based on $\mathbf{X}_1,\dots,\mathbf{X}_n$ is the following:

Algorithm[1]:

(1) Select the number d of FPC's by the CPV-method (Remark 2.3) such that most of the data variability can be explained by $\nu_1,\dots,\nu_d$. Compute the FPC scores $\langle X_k,\nu_l\rangle$ for $l=1,\dots,d$ and $k=1,\dots,n$ by projecting each X_k for $k=1,\dots,n$ on $\nu_1,\dots,\nu_d$. We summarize the scores in the vector

$$\mathbf{X}_k := (\langle X_k,\nu_1\rangle,\dots,\langle X_k,\nu_d\rangle), \quad k=1,\dots n. \tag{4.3}$$

(2) Now we consider the d-dimensional vectors $\mathbf{X}_1,\dots,\mathbf{X}_n$. Using (4.2), we compute the best vector linear predictor of $\mathbf{X}_{n+1}$ that we denote by

$$\widehat{\mathbf{X}}_{n+1} = (\widehat{\langle X_{n+1},\nu_1\rangle},\dots,\widehat{\langle X_{n+1},\nu_d\rangle})^\top.$$

[1]The first and the third step in the algorithm can be implemented in `R` with the package `fda`, and the second step can be achieved with the `R` package `mts`.

(3) Finally, we re-transform the best vector linear predictor $\widehat{\mathbf{X}}_{n+1}$ into a functional form $\widehat{X}_{n+1}$ by the truncated Karhunen-Loéve representation:

$$\widehat{X}_{n+1} := \widehat{\langle X_{n+1}, \nu_1\rangle}\nu_1 + \cdots + \widehat{\langle X_{n+1}, \nu_d\rangle}\nu_d = (\nu_1, \dots, \nu_d)\,\widehat{\mathbf{X}}_{n+1}. \tag{4.4}$$

In [1] the resulting predictor (3) is compared to the functional best linear predictor for functional AR(1) processes.

Our goal is to extend these results to functional ARMA(p,q) processes. However, when moving away from autoregressive models, the best linear predictor is no longer directly given by the process. Therefore, we start by recalling the notion of best linear predictors in Hilbert spaces.

4.2 Functional best linear predictor

We introduce to our setting the *functional best linear predictor* $\widehat{X}_{n+1}$ of X_{n+1} based on $X_1, \dots, X_n$ proposed in [5], whose notation we also use. It is the projection of X_{n+1} on a large enough subspace of L^2_H containing $X_1, \dots, X_n$. More formally, we use the concept of $\mathcal{L}$-*closed subspaces* as in Definition 1.1 in [3].

Definition 4.2. Recall that $\mathcal{L}$ denotes the space of bounded linear operators acting on H. We call G an $\mathcal{L}$-*closed subspace* (LCS) of L^2_H, if
(1) G is a Hilbertian subspace of L^2_H.
(2) If $X \in G$ and $g \in \mathcal{L}$, then $gX \in G$. □

We define

$$X^{(n)} := (X_n, \dots, X_1). \tag{4.5}$$

By Theorem 1.8 of [3] the LCS $G := G_{X^{(n)}}$ generated by $X^{(n)}$ is the closure of $G'_{X^{(n)}}$, where

$$G'_{X^{(n)}} := \left\{ g_n X^{(n)} : \; g_n \in \mathcal{L}(H^n, H) \right\}. \tag{4.6}$$

The *functional best linear predictor* $\hat{X}^G_{n+1}$ of X_{n+1} is defined as the projection of X_{n+1} on G, which we write as

$$\hat{X}^G_{n+1} := P_G X_{n+1} \in G. \tag{4.7}$$

Its properties are given by the projection theorem (e.g. Theorem 2.3.1 of [6]) and are summarized as follows.

Remark 4.3. (i) $E\langle X_{n+1} - \hat{X}^G_{n+1}, Y\rangle = 0$ for all $Y \in G$.
(ii) $\hat{X}^G_{n+1}$ is the unique element in G such that $E\|X_{n+1} - \hat{X}^G_{n+1}\|^2 = \inf_{Y\in G} E\|X_{n+1} - Y\|^2$.
(iii) The mean squared error of the functional best linear predictor $\hat{X}^G_{n+1}$ will be denoted by

$$\sigma^2_n := E\|X_{n+1} - \hat{X}^G_{n+1}\|^2. \tag{4.8}$$

□

Note that, since $G'_{X(n)}$ is not closed in general (cf. [5], Proposition 2.1), $\hat{X}^G_{n+1}$ is not necessarily of the form $\hat{X}^G_{n+1} = g_n X^{(n)}$ for some $g_n \in \mathcal{L}(H^n, H)$, where $\mathcal{L}(H^n, H)$ denotes the space of bounded linear operators from H^n to H. However, the following Proposition gives necessary and sufficient conditions for $\hat{X}^G_{n+1}$ to be represented in terms of bounded linear operators.

Proposition 4.4 (Proposition 2.2, [5]). *The following statements are equivalent:*
(i) There exists some $g_0 \in \mathcal{L}(H^n, H)$ such that $C_{X^{(n)},X_{n+1}} = g_0 C_{X^{(n)}}$.
(ii) $P_G X_{n+1} = g_0 X^{(n)}$ for some $g_0 \in \mathcal{L}(H^n, H)$.

We now formulate conditions for $\hat{X}^G_{n+1}$ to have the representation $\hat{X}^G_{n+1} = s_n X^{(n)}$ for some Hilbert-Schmidt operator s_n from H^n to H ($s_n \in \mathcal{S}(H^n, H)$).

Proposition 4.5. *The following statements are equivalent:*
(i) There exists some $s_0 \in \mathcal{S}(H^n, H)$ such that $C_{X^{(n)},X_{n+1}} = s_0 C_{X^{(n)}}$.
(ii) $P_G X_{n+1} = s_0 X^{(n)}$ for some $s_0 \in \mathcal{S}(H^n, H)$.

Proof. The proof is similar to the proof of Proposition 4.4. Assume there exists some $s_0 \in \mathcal{S}(H^n, H)$, such that $C_{X^{(n)},X_{n+1}} = s_0 C_{X^{(n)}}$.
Then, since $C_{X^{(n)},s_0 X^{(n)}} = E[s_0 X^{(n)} \langle X_{(n)}, \cdot \rangle] = s_0 C_{X^{(n)}}$, we have

$$C_{X^{(n)},X_{n+1}-s_0 X^{(n)}} = 0.$$

Therefore $X_{n+1} - s_0(X^{(n)}) \perp X^{(n)}$ and hence $X_{n+1} - s_0(X^{(n)}) \perp G$ which gives (ii).
For the reverse, note that (ii) implies

$$C_{X^{(n)},X_{n+1}-s_0 X^{(n)}} = C_{X^{(n)},X_{n+1}-P_G X_{n+1}} = 0.$$

Hence $C_{X^{(n)},X_{n+1}} = C_{X^{(n)},s_0 X^{(n)}} = s_0 C_{X^{(n)}}$, which finishes the proof. □

We will proceed with examples of processes where Proposition 4.4 or Proposition 4.5 applies.

Example 4.6. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary and invertible functional linear process, such that*

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \pi_j(X_{n-j}), \quad n \in \mathbb{Z},$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN *and $\pi_j \in \mathcal{L}$ with $\sum_{j=1}^{\infty} \|\pi_j\|_{\mathcal{L}} < \infty$. Then there exists some $l_0 \in \mathcal{L}(H^n, H)$ such that $C_{X^{(n)},X_{n+1}} = l_0 C_{X^{(n)}}$.*

Proof. By Lemma 2.1 in [4] it suffices to show that there exists some $\alpha > 0$ such that

$$\left\| C_{X^{(n)},X_{n+1}} \right\|_{\mathcal{L}} \leq \alpha \left\| C_{X^{(n)}} \right\|_{\mathcal{L}} \quad n \in \mathbb{Z}.$$

In the above equation, the norm used for the right-hand side is the operator norm on $\mathcal{L}(H^n, H)$, and on the left-hand side, it is the operator norm on $\mathcal{L}(H^n, H^n)$. To ease the representation we use the same notation being confident that this does not lead to misunderstandings. For $Y, Z \in L_H^2$, by repeatedly applying the Cauchy-Schwarz inequality, $\|C_{Y,Z}\|_{\mathcal{L}} \le \|C_Y\|_{\mathcal{L}}^{1/2} \|C_Z\|_{\mathcal{L}}^{1/2}$. Hence, we have

$$\begin{aligned}
\Big\| E\big[X_{n+1}\langle X^{(n)}, \cdot\rangle\big]\Big\|_{\mathcal{L}} &= \Big\| E\big[\sum_{j=1}^{\infty} \pi_j X_{n+1-j}\langle X^{(n)}, \cdot\rangle\big]\Big\|_{\mathcal{L}} \\
&= \Big\| \big(\pi_1, \dots, \pi_n\big) E\big[X^{(n)}\langle X^{(n)}, \cdot\rangle\big] + \sum_{j>n} \pi_j E\big[X_{n+1-j}\langle X^{(n)}, \cdot\rangle\big]\Big\|_{\mathcal{L}} \\
&\le \big\| \big(\pi_1, \dots, \pi_n\big)\big\|_{\mathcal{L}} \big\|C_{X^{(n)}}\big\|_{\mathcal{L}} + \sum_{j>n} \|\pi_j\|_{\mathcal{L}} \big\|C_{X^{(n)}}\big\|_{\mathcal{L}}^{1/2} \big\|C_{X_{n+1-j}}\big\|_{\mathcal{L}}^{1/2} \\
&\le \big\| \big(\pi_1, \dots, \pi_n\big)\big\|_{\mathcal{L}} \big\|C_{X^{(n)}}\big\|_{\mathcal{L}} + \Big(\sum_{j>n} \|\pi_j\|_{\mathcal{L}}\Big) \big\|C_{X^{(n)}}\big\|_{\mathcal{L}}^{1/2} \big\|C_{X_0}\big\|_{\mathcal{L}}^{1/2}.
\end{aligned}$$

By stationarity $C_{X_0} = C_{X_k}$ for all $k = 1, \dots, n$. But C_{X_0} is the projection of $C_{X^{(n)}} \in \mathcal{L}(H^n, H^n)$ on the first component in the following sense: with P_1 defined as in (3.9), $C_{X_0} = P_1 C_{X^{(n)}} P_1$. Hence, using Theorem 4.2.7 of [13], one can show that $\lambda_j(C_{X_0}) \le \lambda_j(C_{X^{(n)}})$ for $j \in \mathbb{N}$, where $\lambda_j(C_{X_0})$ and $\lambda_j(C_{X^{(n)}})$ denote the j-th eigenvalues of C_{X_0} and $C_{X^{(n)}}$, respectively. Since furthermore $\|C_{X_0}\|_{\mathcal{L}} = \lambda_1(C_{X_0})$ (e.g. [7], Theorem 4.9.8) and $\|C_{X_{(n)}}\|_{\mathcal{L}} = \lambda_1(C_{X^{(n)}})$ we get $\|C_{X_0}\|_{\mathcal{L}} \le \|C_{X_{(n)}}\|_{\mathcal{L}}$, and

$$\Big\| E\big[X_{n+1}\langle X^{(n)}, \cdot\rangle\big]\Big\| \le \Big(\big\|\big(\pi_1, \dots, \pi_n\big)\big\|_{\mathcal{L}} + \sum_{j>n} \|\pi_j\|_{\mathcal{L}}\Big) \big\|C_{X^{(n)}}\big\|_{\mathcal{L}}.$$

The invertibility of $(X_n)_{n\in\mathbb{Z}}$ assures the boundedness of $\|(\pi_1, \dots, \pi_n)\| + \sum_{j>n} \|\pi_j\|$. □

Note that an obvious special case of Example 4.6 is a functional autoregressive process of finite order. In this case we can also apply Proposition 4.5 in an obvious way.

Example 4.7. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary functional* AR(p) *process with representation*

$$X_n = \varepsilon_n + \sum_{j=1}^{p} \phi_j(X_{n-j}), \quad n \in \mathbb{Z},$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is WN *and $\phi_j \in \mathcal{S}$ are Hilbert Schmidt operators. Then for $n \ge p$ Proposition 4.5 applies, giving $P_G X_{n+1} = s_0 X^{(n)}$ for some $s_0 \in \mathcal{S}$.*

Proof. We immediately get

$$\begin{aligned}
C_{X^{(n)}, X_{n+1}}(\cdot) = E\big[X_{n+1}\langle X^{(n)}, \cdot\rangle\big] &= E\big[\sum_{j=1}^{p} \phi_j X_{n+1-j}\langle X^{(n)}, \cdot\rangle\big] \\
&= E\big[(\phi_1, \dots, \phi_p, \mathbf{0}, \dots, \mathbf{0}) X^{(n)}\langle X^{(n)}, \cdot\rangle\big] \\
&= \phi C_{X^{(n)}}(\cdot),
\end{aligned}$$

where $\phi = (\phi_1, \dots, \phi_p, \mathbf{0}, \dots, \mathbf{0}) \in \mathcal{L}(H^n, H)$. Now let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis of H. Then $(f_j)_{j\in\mathbb{N}}$ with $f_1 = (e_1, 0, \dots, 0)^\top, f_2 = (0, e_1, 0, \dots, 0)^\top, \dots, f_n = (0, \dots, 0, e_1)^\top$, $f_{n+1} = (e_2, 0, \dots, 0)^\top, f_{n+2} = (0, e_2, 0, \dots, 0)^\top, \dots, f_{2n} = (0, \dots, 0, e_2)^\top, f_{2n+1} = (e_3, 0, \dots, 0)^\top, \dots$ is an orthonormal basis of H^n and, by orthogonality of $(e_i)_{i\in\mathbb{N}}$, we get

$$\|\phi\|_{\mathcal{S}}^{1/2} = \sum_{j\in\mathbb{N}} \|\phi(f_j)\|^2 = \sum_{i\in\mathbb{N}} \sum_{j=1}^{p} \|\phi_j(e_i)\|^2 = \sum_{j=1}^{p} \sum_{i\in\mathbb{N}} \|\phi_j(e_i)\|^2 = \sum_{j=1}^{p} \|\phi_j\|_{\mathcal{L}}^2 < \infty,$$

since $\phi_j \in \mathcal{S}$ for every $j = 1, \dots, p$, which implies that $\phi \in \mathcal{S}(H^n, H)$. □

Example 4.8. *Let* $(X_n)_{n\in\mathbb{Z}}$ *be a stationary functional* MA(1) *process*

$$X_n = \varepsilon_n + \theta(\varepsilon_{n-1}) \quad n \in \mathbb{Z},$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ *is* WN, $\|\theta\|_{\mathcal{L}} < 1$, $\theta \in \mathcal{S}$ *and* θ *nilpotent, such that* $\|\theta^j\|_{\mathcal{L}} = 0$ *for* $j > j_0$ *for some* $j_0 \in \mathbb{N}$. *Then for* $n > j_0$ *Proposition 4.5 applies.*

Proof. Since $\|\theta\|_{\mathcal{L}} < 1$, $(X_n)_{n\in\mathbb{Z}}$ is invertible, and since θ is nilpotent, $(X_n)_{n\in\mathbb{Z}}$ can be represented as an autoregressive process of finite order, where the operators in the inverse representation are still Hilbert-Schmidt operators. Then the statement follows from the arguments of the proof of Example 4.7. □

Example 4.9. *Let* $(X_n)_{n\in\mathbb{Z}}$ *be a stationary functional* MA(1) *process*

$$X_n = \varepsilon_n + \theta(\varepsilon_{n-1}) \quad n \in \mathbb{Z},$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ *is* WN, *and denote by* C_ε *the covariance operator of* ε_0. *Assume that* $\|\theta\|_{\mathcal{L}} < 1$. *If* θ *and* C_ε *commute, then there exists an* $s_0 \in \mathcal{S}$ *such that* $C_{X_n,X_{n+1}} = s_0 C_{X_n}$.

Proof. Stationarity of $(X_n)_{n\in\mathbb{Z}}$ ensures that $C_{X_n,X_{n+1}} = C_{X_0,X_1}$. Let θ^* denote the adjoint operator of θ. Since $\theta C_\varepsilon = C_\varepsilon\theta$, we have that $C_{X_1,X_0} = C_{X_0,X_1}$ which implies $\theta C_\varepsilon = C_\varepsilon\theta^* = C_\varepsilon\theta$. Hence, $C_\varepsilon = C_{X_0} - \theta C_\varepsilon\theta^* = C_{X_0} - \theta C_\varepsilon\theta = C_{X_0} - \theta^2 C_\varepsilon$, and we get by iteration

$$C_{X_1,X_0} = \theta C_\varepsilon = \theta(C_{X_0} - \theta C_\varepsilon\theta) = \sum_{j\ge 0}(-\theta^2)^j(\theta C_{X_0}) = (Id + \theta^2)^{-1}\theta C_{X_0},$$

where $Id + \theta^2$ is invertible, since $\|\theta\|_{\mathcal{L}} < 1$. Furthermore, since the space $\mathcal{S}$ of Hilbert-Schmidt operators forms an ideal in the space of bounded linear (e.g. [8], Theorem VI.5.4.) operators and $\theta \in \mathcal{S}$, also $(Id + \theta^2)^{-1}\theta \in \mathcal{S}$. □

4.3 Bounds for the error of the vector predictor

We are now ready to derive bounds for the prediction error caused by the dimension reduction. More precisely, we want to compare the predictor

$$\widehat{X}_{n+1} := \sum_{j=1}^{d} \widehat{\langle X_k, \nu_j\rangle}\nu_j = (\nu_1, \dots, \nu_d)\,\widehat{\mathbf{X}}_{n+1}$$

as defined in (4.4), and based on the vector process, with the functional best linear predictor

$$\hat{X}^G_{n+1} = P_G X_{n+1},$$

as defined in (4.7). We first compare them on $\mathrm{sp}\{\nu_1, \dots, \nu_d\}$ where the vector representations are

$$\widehat{\mathbf{X}}_{n+1} = (\widehat{\langle X_{n+1}, \nu_1\rangle}, \dots, \widehat{\langle X_{n+1}, \nu_d\rangle})^\top \quad \text{and} \quad \widehat{\mathbf{X}}^G_{n+1} := \left(\left\langle \hat{X}^G_{n+1}, \nu_1\right\rangle, \dots, \left\langle \hat{X}^G_{n+1}, \nu_d\right\rangle\right)^\top. \tag{4.9}$$

We give Assumptions under which for $d \to \infty$ the mean squared distance between the vector best linear predictor $\hat{\mathbf{X}}_{n+1}$ and the vector $\hat{\mathbf{X}}^G_{n+1}$ becomes arbitrarily small.

For $l = 1, \dots, d$, the l-th component of $\widehat{\mathbf{X}}^G_{n+1}$ is given by

$$\left\langle \hat{X}^G_{n+1}, \nu_l \right\rangle = \left\langle \sum_{i=1}^n g_{ni}(X_i), \nu_l \right\rangle = \left\langle \sum_{i=1}^n \sum_{l'=1}^\infty \langle X_i, \nu_{l'} \rangle \, g_{ni}(\nu_{l'}), \nu_l \right\rangle = \sum_{i=1}^n \sum_{l'=1}^\infty \langle X_i, \nu_{l'} \rangle \langle g_{ni}(\nu_{l'}), \nu_l \rangle . \tag{4.10}$$

Using a vector representation, we can write

$$\begin{aligned}
\hat{\mathbf{X}}^G_{n+1} &= \sum_{i=1}^n \left[\begin{array}{ccc|cc} \langle g_{ni}(\nu_1), \nu_1 \rangle & \dots & \langle g_{ni}(\nu_d), \nu_1 \rangle & \langle g_{ni}(\nu_{d+1}), \nu_1 \rangle & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle g_{ni}(\nu_1), \nu_d \rangle & \dots & \langle g_{ni}(\nu_d), \nu_d \rangle & \langle g_{ni}(\nu_{d+1}), \nu_d \rangle & \dots \end{array} \right] \left(\begin{array}{c} \langle X_i, \nu_1 \rangle \\ \vdots \\ \langle X_i, \nu_d \rangle \\ \hline \langle X_i, \nu_{d+1} \rangle \\ \vdots \end{array} \right) \\
&=: \sum_{i=1}^n \mathbf{G}_{ni} \mathbf{X}_i + \sum_{i=1}^d \mathbf{G}^\infty_{ni} \mathbf{X}^\infty_i ,
\end{aligned} \tag{4.11}$$

where $\mathbf{G}_{ni}$ is a $d \times d$ matrix with ll'-th component $\langle g_{ni}(\nu_{l'}), \nu_l \rangle$ and $\mathbf{G}^\infty_{ni}$ is a $d \times \infty$ matrix with ll'-th component $\langle g_{ni}(\nu_{d+l'}), \nu_l \rangle$.

Moreover, for all $Y \in G$ there exist (possibly unbounded) linear operators $t_{n1}, \dots, t_{n,n}$ such that

$$Y = \sum_{i=1}^n t_{ni}(X_i). \tag{4.12}$$

Similar to (4.10), we project $Y \in G$ on $\nu_1, \dots, \nu_d$, which results in

$$\begin{aligned}
\mathbf{Y} &:= (\langle Y, \nu_1 \rangle, \dots, \langle Y, \nu_d \rangle)^\top \\
&= \left(\left\langle \sum_{i=1}^n t_{ni}(X_i), \nu_1 \right\rangle, \dots, \left\langle \sum_{i=1}^n t_{ni}(X_i), \nu_d \right\rangle \right)^\top \\
&:= \sum_{i=1}^n \mathbf{T}_{ni} \mathbf{X}_i + \sum_{i=1}^n \mathbf{T}^\infty_{ni} \mathbf{X}^\infty_i .
\end{aligned} \tag{4.13}$$

The $d \times d$ matrix $\mathbf{T}_{ni}$ and the $d \times \infty$ matrix $\mathbf{T}^\infty_{ni}$ in (4.12) are defined in the same way as $\mathbf{G}_{ni}$ and $\mathbf{G}^\infty_{ni}$ in (4.11). We denote by $\mathbf{M}$ the space of all $\mathbf{Y}$:

$$\mathbf{M} := \left\{ \mathbf{Y} = (\langle Y, \nu_1 \rangle, \dots, \langle Y, \nu_d \rangle)^\top : Y \in G \right\} \tag{4.14}$$

Observing that for all $\mathbf{Y}_1 \in \mathbf{M}_1$ there exist $d \times d$ matrices $\mathbf{A}_{n1}, \dots, \mathbf{A}_{nn}$ such that $\mathbf{Y}_1 = \sum_{i=1}^n \mathbf{A}_{ni} \mathbf{X}_i$, one can find operators t_{ni} such that $\mathbf{T}_{ni} = \mathbf{A}_{ni}$, and $\mathbf{T}^\infty_{ni} = \mathbf{0}$, which then gives $\mathbf{Y}_1 \in \mathbf{M}$. Hence $\mathbf{M}_1 \subseteq \mathbf{M}$.

Now that we have introduced the notation and the setting, we are ready to compute the mean squared distance $E \left\| \hat{\mathbf{X}}_{n+1} - \hat{\mathbf{X}}^G_{n+1} \right\|_2^2$.

Theorem 4.10. *Suppose $(X_n)_{n \in \mathbb{Z}}$ is a functional* ARMA(p, q) *process such that Assumption 3.7 holds. Let $\hat{X}^G_{n+1}$ be the functional best linear predictor of X_{n+1} as defined in* (4.7) *and let $\hat{\mathbf{X}}^G_{n+1}$ be as defined in (4.9). Let furthermore $\hat{\mathbf{X}}_{n+1}$ be the vector best linear predictor of $\mathbf{X}_{n+1}$ based on $\mathbf{X}_1, \dots, \mathbf{X}_n$ as in* (4.2).

(i) In the framework of Proposition 4.4, and if $\sum_{l=1}^\infty \sqrt{\lambda_l} < \infty$, for all $d \in \mathbb{N}$,

$$E \left\| \hat{\mathbf{X}}_{n+1} - \hat{\mathbf{X}}^G_{n+1} \right\|_2^2 \le 4 \Big(\sum_{i=1}^n \| g_{ni} \|_{\mathcal{L}} \Big)^2 \Big(\sum_{l=d+1}^\infty \sqrt{\lambda_l} \Big)^2 < \infty. \tag{4.15}$$

(ii) In the framework of Proposition 4.5, for all $d \in \mathbb{N}$,

$$E\left\|\hat{\mathbf{X}}_{n+1}-\hat{\mathbf{X}}_{n+1}^{G}\right\|_2^2 \le 4\Big(\sum_{i=1}^{n}\Big(\sum_{l=d+1}^{\infty}\|g_{ni}(\nu_l)\|^2\Big)^{\frac{1}{2}}\Big)^2\sum_{l=d+1}^{\infty}\lambda_l<\infty. \tag{4.16}$$

In both cases, $E\left\|\hat{\mathbf{X}}_{n+1}-\hat{\mathbf{X}}_{n+1}^{G}\right\|_2^2$ tends to 0 as $d\to\infty$.

We start with a technical lemma, which we shall need for the proof of the above Theorem.

Lemma 4.11. *Suppose $(X_n)_{n\in\mathbb{Z}}$ is a stationary and causal functional* $\mathrm{ARMA}(p,q)$ *process and $(\nu_l)_{l\in\mathbb{Z}}$ are the eigenfunctions of its covariance operator C. Then for all $j,l\in\mathbb{Z}$*

$$E\left[\left\langle X_{n+1}-\hat{X}_{n+1}^{G},\nu_l\right\rangle\langle Y,\nu_j\rangle\right]=0,\quad Y\in G. \tag{4.17}$$

Proof. For all $j,l\in\mathbb{Z}$ we set $s_{l,j}(\cdot):=\langle\cdot,\nu_l\rangle\nu_j$. We first show that $s_{l,j}\in\mathcal{L}$. Since for all $x\in H$ with $\|x\|\le 1$,

$$\|s_{l,j}(x)\|=\|\langle x,\nu_l\rangle\nu_j\|\le\|x\|\le 1,$$

hence $s_{l,j}\in\mathcal{L}$. Since G is an $\mathcal{L}$-closed subspace, $Y\in G$ implies $s_{l,j}(Y)\in G$ and we get with Remark 4.3(i) for all $j,l\in\mathbb{Z}$,

$$E\left\langle X_{n+1}-\hat{X}_{n+1}^{G},s_{l,j}(Y)\right\rangle=E\left\langle X_{n+1}-\hat{X}_{n+1}^{G},\langle Y,\nu_l\rangle\nu_j\right\rangle=E\left[\left\langle X_{n+1}-\hat{X}_{n+1}^{G},\nu_l\right\rangle\langle Y,\nu_j\rangle\right]=0.$$

$\square$

Proof of Theorem 4.10. First, notice that both under (i) and (ii) there exist $g_{ni}\in\mathcal{L}$ such that $\hat{X}_{n+1}^{G}=\sum_{i=1}^{n}g_{ni}X_{n+1-i}$, and that $\mathcal{S}\subset\mathcal{L}$. Furthermore, recall that $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$ and $\langle\ ,\ \rangle_{\mathbb{R}^d}$ the corresponding scalar product. Now, using the matrix representation of $\hat{\mathbf{X}}_{n+1}^{G}$ in (4.11) and Lemma 4.11, we obtain

$$\begin{aligned}\sum_{j=1}^{d}E\Big[\langle Y,\nu_j\rangle\big\langle X_{n+1}-\hat{X}_{n+1}^{G},\nu_j\big\rangle\Big]&=E\Big\langle\mathbf{Y},\mathbf{X}_{n+1}-\hat{\mathbf{X}}_{n+1}^{G}\Big\rangle_{\mathbb{R}^d}\\&=E\Big\langle\mathbf{Y},\mathbf{X}_{n+1}-\sum_{i=1}^{n}\mathbf{G}_{ni}\mathbf{X}_i-\sum_{i=1}^{n}\mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty}\Big\rangle_{\mathbb{R}^d}=0,\quad Y\in G,\end{aligned} \tag{4.18}$$

where we have set

$$\mathbf{Y}=(\langle Y,\nu_1\rangle,\dots,\langle Y,\nu_d\rangle)^{\top}\in\mathbf{M}. \tag{4.19}$$

Since (4.18) holds for all $\mathbf{Y}\in\mathbf{M}$ and $\mathbf{M}_1\subseteq\mathbf{M}$, it especially holds for all $\mathbf{Y}_1\in\mathbf{M}_1$; i.e.,

$$E\Big\langle\mathbf{Y}_1,\mathbf{X}_{n+1}-\sum_{i=1}^{n}\mathbf{G}_{ni}\mathbf{X}_i-\sum_{i=1}^{n}\mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty}\Big\rangle_{\mathbb{R}^d}=0,\quad\mathbf{Y}_1\in\mathbf{M}_1. \tag{4.20}$$

Combining (4.20) and Remark 4.3(i), we have

$$E\Big\langle\mathbf{Y}_1,\hat{\mathbf{X}}_{n+1}-\sum_{i=1}^{n}\mathbf{G}_{ni}\mathbf{X}_i\Big\rangle_{\mathbb{R}^d}=E\Big\langle\mathbf{Y}_1,\sum_{i=1}^{n}\mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty}\Big\rangle_{\mathbb{R}^d},\quad\mathbf{Y}_1\in\mathbf{M}_1. \tag{4.21}$$

Since both $\hat{\mathbf{X}}_{n+1}$ and $\sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i$ are in $\mathbf{M}_1$, (4.21) especially holds, when

$$\mathbf{Y}_1 = \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \in \mathbf{M}. \tag{4.22}$$

We plug $\mathbf{Y}_1$ as defined in (4.22) in (4.21) and obtain

$$E\Big\langle \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i, \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\rangle_{\mathbb{R}^d} = E\Big\langle \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i, \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\rangle_{\mathbb{R}^d}. \tag{4.23}$$

From the left hand side of (4.23) we read off

$$E\Big\langle \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i, \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\rangle_{\mathbb{R}^d} = E\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\|_2^2, \tag{4.24}$$

and for the right hand side of (4.23) we get by the Cauchy-Schwarz inequality applied twice,

$$\begin{aligned} E\Big\langle \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i, \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\rangle_{\mathbb{R}^d} &\le E\Big[\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\|_2 \Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2 \Big] \\ &\le \Big(E\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\|_2^2 \Big)^{\frac{1}{2}} \Big(E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2 \Big)^{\frac{1}{2}}. \end{aligned} \tag{4.25}$$

Dividing the right hand side of (4.24) by the first square root on the right hand side of (4.25) we obtain

$$E\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\|_2^2 \le E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2. \tag{4.26}$$

Hence, for the mean squared distance we get

$$\begin{aligned} E\Big\| \hat{\mathbf{X}}_{n+1} - \hat{\mathbf{X}}_{n+1}^{G} \Big\|_2^2 &= E\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i - \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2 \\ &\le 2E\Big\| \hat{\mathbf{X}}_{n+1} - \sum_{i=1}^{n} \mathbf{G}_{ni}\mathbf{X}_i \Big\|_2^2 + 2E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2 \\ &\le 4E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2. \end{aligned} \tag{4.27}$$

What remains to do is to bound $\sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty}$, which, by (4.11), is a d-dimensional vector with l-th component

$$\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} \langle X_i, \nu_{l'} \rangle \langle g_{ni}(\nu_l'), \nu_l \rangle = \sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} x_{i,l'} \langle g_{ni}(\nu_l'), \nu_l \rangle.$$

(i) First we consider the framework of Proposition 4.4. We abbreviate $x_{i,l'} := \langle X_i, \nu_l' \rangle$ and calculate

$$E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty}\mathbf{X}_i^{\infty} \Big\|_2^2 = E\Big[\sum_{l=1}^{d} \Big(\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} x_{i,l'} \langle g_{ni}(\nu_l'), \nu_l \rangle \Big)^2 \Big]$$

$$
\begin{aligned}
&= E\Big\| \sum_{l=1}^{d} \Big(\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} x_{i,l'} \langle g_{ni}(\nu_l'), \nu_l \rangle \Big) \nu_l \Big\|^2 \\
&\leq E\Big\| \sum_{l=1}^{\infty} \Big(\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} x_{i,l'} \langle g_{ni}(\nu_l'), \nu_l \rangle \Big) \nu_l \Big\|^2 \\
&= E\Big\| \sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} x_{i,l'} g_{ni}(\nu_l') \Big\|^2
\end{aligned}
\tag{4.28}
$$

by Parseval's identity. Then we proceed using the linearity and orthogonality of ν_l and the Cauchy-Schwarz inequality

$$
\begin{aligned}
&= E\Big[\Big\langle \sum_{i=1}^{n} \sum_{l=d+1}^{\infty} x_{i,l} g_{ni}(\nu_l), \sum_{j=1}^{n} \sum_{l'=d+1}^{\infty} x_{j,l'} g_{n,j}(\nu_l') \Big\rangle \Big] \\
&= \sum_{i,j=1}^{n} \sum_{l,l'=d+1}^{\infty} E(x_{i,l} x_{j,l'}) \langle g_{ni}(\nu_l), g_{n,j}(\nu_{l'}) \rangle \qquad (4.29) \\
&\leq \Big(\sum_{i=1}^{n} \sum_{l=d+1}^{\infty} \sqrt{E(x_{i,l})^2} \| g_{ni}(\nu_l) \| \Big) \Big(\sum_{j=1}^{n} \sum_{l'=d+1}^{\infty} \sqrt{E(x_{j,l'})^2} \| g_{n,j}(\nu_{l'}) \| \Big) \\
&= \Big(\sum_{i=1}^{n} \sum_{l=d+1}^{\infty} \sqrt{\lambda_l} \| g_{ni}(\nu_l) \| \Big) \Big(\sum_{j=1}^{n} \sum_{l'=d+1}^{\infty} \sqrt{\lambda_{l'}} \| g_{n,j}(\nu_{l'}) \| \Big),
\end{aligned}
$$

since $E\langle X_i, \nu_l \rangle^2 = \lambda_l$ by (2.8). Then using the linearity of the operators

$$
\begin{aligned}
&\leq \Big(\sum_{i=1}^{n} \sum_{l=d+1}^{\infty} \sqrt{\lambda_l} \| g_{ni} \|_{\mathcal{L}} \| \nu_l \| \Big) \Big(\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} \sqrt{\lambda_{l'}} \| g_{ni} \|_{\mathcal{L}} \| \nu_{l'} \| \Big) \\
&= \Big(\sum_{i=1}^{n} \sum_{l=d+1}^{\infty} \sqrt{\lambda_l} \| g_{ni} \|_{\mathcal{L}} \Big) \Big(\sum_{i=1}^{n} \sum_{l'=d+1}^{\infty} \sqrt{\lambda_{l'}} \| g_{ni} \|_{\mathcal{L}} \Big) \\
&= \Big(\sum_{i=1}^{n} \| g_{ni} \|_{\mathcal{L}} \Big)^2 \Big(\sum_{l=d+1}^{\infty} \sqrt{\lambda_l} \Big)^2,
\end{aligned}
$$

since $\|\nu_l\| = 1$. Now since $g_{ni} \in \mathcal{L}$, we have $\sum_{i=1}^{n} \|g_{ni}\|_{\mathcal{L}} < \infty$ for all $n \in \mathbb{N}$ and with $\sum_{l=1}^{\infty} \sqrt{\lambda_l} < \infty$, the right hand side tends to 0 as $d \to \infty$.

(ii) In the framework of Proposition 4.5 there exist $g_{ni} \in \mathcal{S}$ such that $\hat{X}_{n+1}^{G} = \sum_{i=1}^{n} g_{ni} X_{n+1-i}$. Then, similarly as before, using the Cauchy-Schwarz inequality, we calculate

$$
\begin{aligned}
E\Big\| \sum_{i=1}^{n} \mathbf{G}_{ni}^{\infty} \mathbf{X}_i^{\infty} \Big\|_2^2 &\leq \Big(\sum_{i=1}^{n} \sum_{l=d+1}^{\infty} \sqrt{\lambda_l} \| g_{ni}(\nu_l) \| \Big) \Big(\sum_{j=1}^{n} \sum_{l'=d+1}^{\infty} \sqrt{\lambda_{l'}} \| g_{n,j}(\nu_{l'}) \| \Big) \\
&\leq \Big(\sum_{i=1}^{n} \Big(\sum_{l=d+1}^{\infty} \lambda_l \Big)^{\frac{1}{2}} \Big(\sum_{l=d+1}^{\infty} \| g_{ni}(\nu_l) \|^2 \Big)^{\frac{1}{2}} \Big) \Big(\sum_{j=1}^{n} \Big(\sum_{l'=d+1}^{\infty} \lambda_{l'} \Big)^{\frac{1}{2}} \Big(\sum_{l'=d+1}^{\infty} \| g_{n,j}(\nu_{l'}) \|^2 \Big)^{\frac{1}{2}} \Big) \\
&= \Big(\sum_{i=1}^{n} \Big(\sum_{l=d+1}^{\infty} \| g_{ni}(\nu_l) \|^2 \Big)^{\frac{1}{2}} \Big)^2 \sum_{l=d+1}^{\infty} \lambda_l
\end{aligned}
\tag{4.30}
$$

Now note that

$$
\sum_{l=d+1}^{\infty} \| g_{ni}(\nu_l) \|^2 \leq \| g_{ni} \|_{\mathcal{S}} < \infty.
$$

Thus, (4.30) is bounded by

$$
\Big(\sum_{i=1}^{n} \Big(\sum_{l=d+1}^{\infty} \| g_{ni}(\nu_l) \|^2 \Big)^{\frac{1}{2}} \Big)^2 \sum_{l=d+1}^{\infty} \lambda_l \leq \Big(\sum_{i=1}^{n} \| g_{ni} \|_{\mathcal{S}}^{\frac{1}{2}} \Big)^2 \sum_{l=d+1}^{\infty} \lambda_l < \infty,
$$

such that (4.30) tends to 0 as $d \to \infty$. □

We are now ready to provide bounds of the mean squared prediction error $E\|X_{n+1} - \hat{X}_{n+1}\|^2$.

Theorem 4.12. *Consider a stationary and causal functional* ARMA(p,q) *process as in* (3.1)*. Then for σ_n^2 as defined in* (4.8)*,*

$$E\left\|X_{n+1} - \hat{X}_{n+1}\right\|^2 \leq \sigma_n^2 + \gamma_{d;n}. \tag{4.31}$$

(i) In the framework of Proposition 4.4, and if $\sum_{l=1}^{\infty}\sqrt{\lambda_l} < \infty$, for all $d \in \mathbb{N}$,

$$\gamma_{d;n} = 4\Big(\sum_{i=1}^{n}\|g_{ni}\|_{\mathcal{L}}\Big)^2\Big(\sum_{l=d+1}^{\infty}\sqrt{\lambda_l}\Big)^2 + \sum_{l=d+1}^{\infty}\lambda_l.$$

(ii) In the framework of Proposition 4.5, for all $d \in \mathbb{N}$,

$$\begin{aligned}\gamma_{d;n} &= \sum_{l=d+1}^{\infty}\lambda_l\left(4\,g_{n;d}+1\right)\\ g_{n;d} &= \sum_{i=1}^{n}\Big(\sum_{l=d+1}^{\infty}\|g_{ni}(\nu_l)\|^2\Big)^{1/2} \leq \sum_{i=1}^{n}\|g_{ni}\|_{\mathcal{S}}^2.\end{aligned}$$

In both cases, $E\left\|X_{n+1} - \hat{X}_{n+1}\right\|_2^2$ tends to σ_n^2 as $d \to \infty$.

Proof. First note that by orthogonality of $\nu_1, \dots, \nu_d$,

$$\begin{aligned}E\left\|X_{n+1} - \hat{X}_{n+1}\right\|^2 &= E\Big\|\sum_{l=1}^{d}\langle X_{n+1} - \hat{X}_{n+1}, \nu_l\rangle\nu_l + \sum_{l=d+1}^{\infty}\langle X_{n+1}, \nu_l\rangle\nu_l\Big\|^2\\ &= \sum_{l=1}^{d}E\left\|\langle X_{n+1} - \hat{X}_{n+1}, \nu_l\rangle\nu_l\right\|^2 + \sum_{l=d+1}^{\infty}E\left\|\langle X_{n+1}, \nu_l\rangle\nu_l\right\|^2\\ &= \sum_{l=1}^{d}E\langle X_{n+1} - \hat{X}_{n+1}, \nu_l\rangle^2 + \sum_{l=d+1}^{\infty}\lambda_l \end{aligned}\tag{4.32}$$

by (2.8) and the fact that $\|\nu_l\| = 1$ for all $l \in \mathbb{N}$. Now recall that similarly as in the first equation of (4.28)

$$\sum_{l=1}^{d}E\langle X_{n+1} - \hat{X}_{n+1}, \nu_l\rangle^2 = E\|\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1}\|_2^2.$$

Furthermore, by Definition 4.2 of $\mathcal{L}$-closed subspaces and Remark 4.3(i) we know that $E\left\langle X_{n+1} - \hat{X}_{n+1}^G, Y\right\rangle = 0$ for all $Y \in G$. Observing that $\hat{X}_{n+1}^G - \hat{X}_{n+1} \in G$, we conclude that

$$E\left\langle X_{n+1} - \hat{X}_{n+1}^G, \hat{X}_{n+1}^G - \hat{X}_{n+1}\right\rangle = 0,$$

and, by Lemma 4.11,

$$E\langle X_{n+1} - \hat{X}_{n+1}^G, \nu_l\rangle\langle \hat{X}_{n+1}^G - \hat{X}_{n+1}, \nu_{l'}\rangle = 0, \quad l, l' \in \mathbb{N}.$$

Hence,

$$E\|\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1}\|_2^2 = E\|\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1}^G\|_2^2 + E\|\hat{\mathbf{X}}_{n+1}^G - \hat{\mathbf{X}}_{n+1}\|_2^2, \tag{4.33}$$

where for the first term of the right hand side,

$$E\|\mathbf{X}_{n+1}-\hat{\mathbf{X}}^G_{n+1}\|_2^2 = E\sum_{l=1}^{d}\langle X_{n+1}-\hat{X}^G_{n+1},\nu_l\rangle^2 \le \sum_{l=1}^{\infty}\langle X_{n+1}-\hat{X}^G_{n+1},\nu_l\rangle^2 = E\|X_{n+1}-\hat{X}^G_{n+1}\|^2 = \sigma_n^2, \tag{4.34}$$

and the last equality holds by Remark 4.3(iii). For the second term of the right hand side of (4.33) we use Theorem 4.10. We finish the proof of both (i) and (ii) by plugging (4.33) and (4.34) into (4.32). □

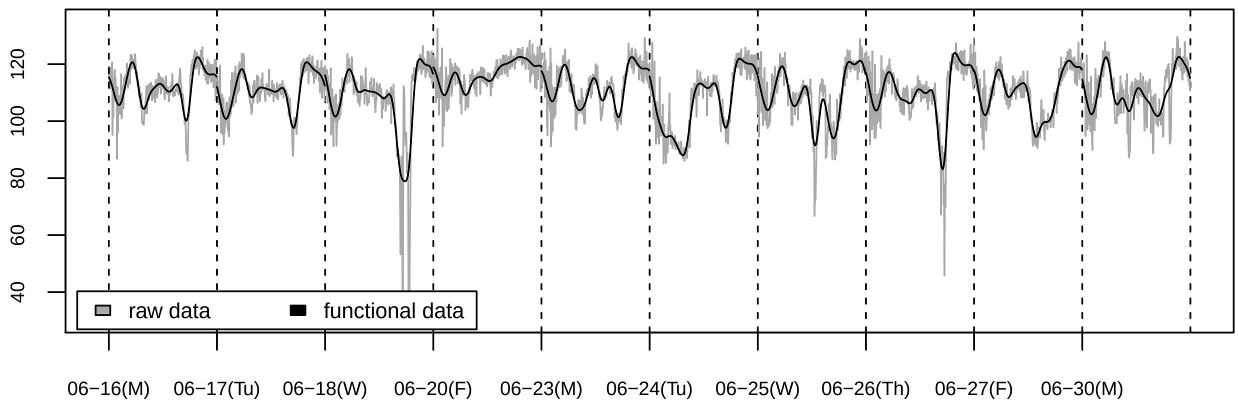

Figure 3: Functional velocity data (black) and raw data (grey) on the last ten working days in June 2014 (June 19th 2014 was a catholic holiday).

5 Real data analysis

In this section we apply the functional time series prediction theory to highway traffic data provided by the Autobahndirektion Südbayern, thus extending previous work by [2]. Our dataset consists of measurements at a fixed point on a highway (A92) in Southern Bavaria, Germany. Recorded is the average velocity per minute from 1/1/2014 00:00 to 30/06/2014 23:59 on three lanes. After taking care of missing values and data outliers, we merge the three lanes (using the weighted average velocity per minute). Finally, we smooth the cleaned daily high-dimensional data, using a Fourier basis to obtain functional data. In Figure 3 we depict the outcome on the working days of two weeks in June 2014. For a precise description we refer to Wei [17], Chapter 6.

As can be seen in Figure 4, different weekdays have different mean velocity functions. To account for the difference between weekdays, we subtract the empirical individual daily mean from all daily data (Monday mean from Monday data, etc.). The effect is clearly visible in Figure 5. However, even after deduction of the daily mean, functional stationarity tests [12] reject stationarity of time series. This is due to traffic flow on weekends: Saturday and Sunday traffic show different patterns than weekdays, even after mean correction. Consequently, we restrict our investigation to working days (Monday-Friday), resulting in a functional time series X_n for $n = 1,\dots,N = 119$.

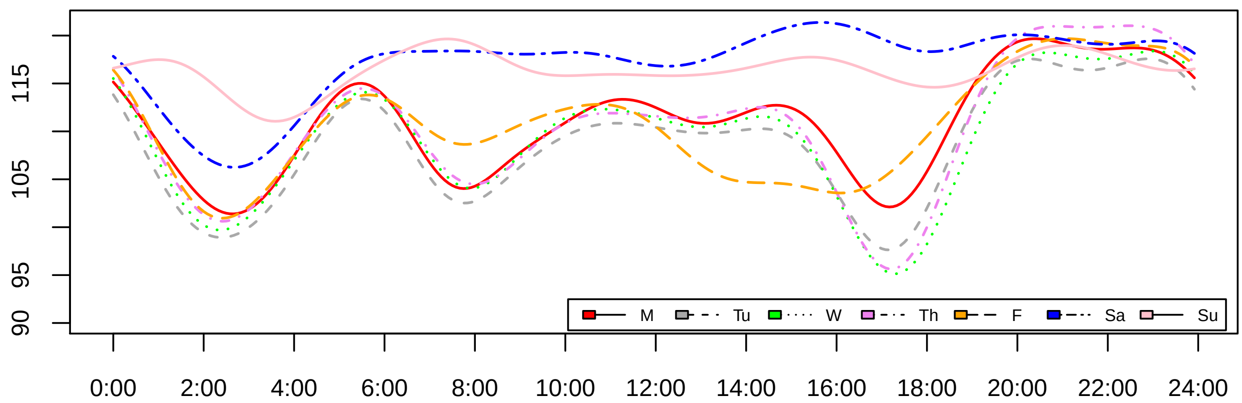

Figure 4: Empirical functional mean velocity on the 7 days of the week

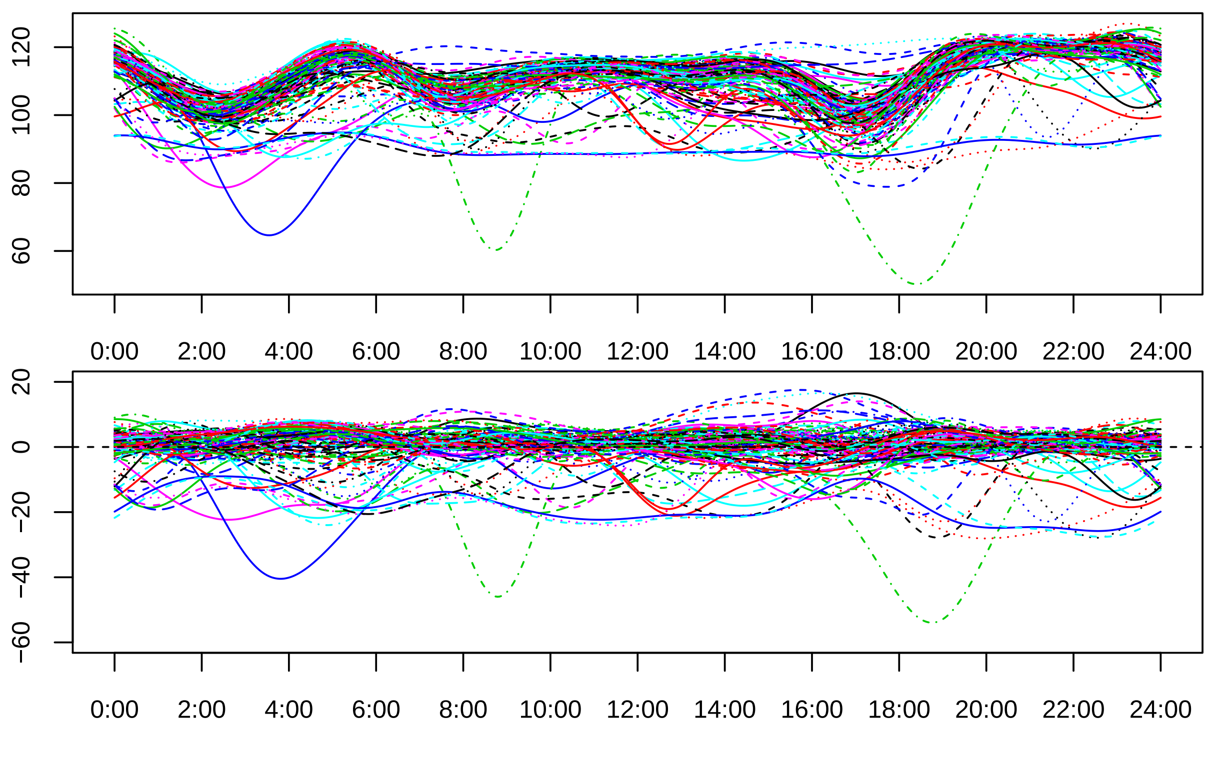

Figure 5: Functional velocity data for 30 working days smoothed by a Fourier basis before and after substracting the weekday mean

A Portmanteau test applied to X_n for $n = 1, \ldots, N$ with $N = 119$ working days (cf. [9]) rejects (with a p-value as small as 10^{-6}) that the daily functional data are uncorrelated. Furthermore, the stationarity tests suggested in [12] do not reject the stationarity assumption.

Figure 6 shows the empirical covariance kernel for the highway functional velocity data on working days, hence the empirical version of $E[(X(t) - \mu(t))(X(s) - \mu(s))]$, for $0 \leq t, s \leq 1$, based on 119 working days.

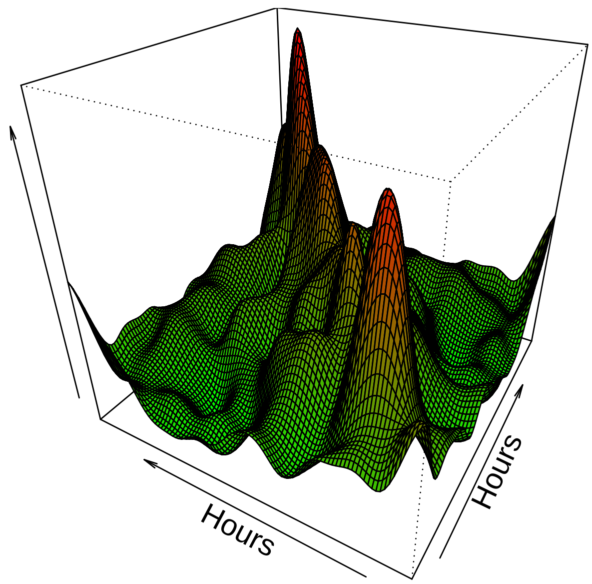

Figure 6: Empirical covariance kernel of functional velocity data on 119 working days.

As indicated by the arrows, the point $(t, s) = (0, 0)$ is at the bottom right corner and estimates the variance at midnight. The empirical variance over the day is represented along the diagonal from the bottom right to the top left corner. The valleys and peaks along the diagonal represent phases of low and high traffic density: for instance, the first peak represents the variance at around 05:00 a.m., where traffic becomes denser, since commuting to work starts. Peaks away from the diagonal represent high dependencies between different time points during the day. For instance, high traffic density in the early morning correlates with high traffic density in the late afternoon, again due to commuting.

Remark 5.1. We want to emphasize that we have developed the prediction theory in its natural framework of a Hilbert space, which follows from the projection theorem. This requires only second order stationarity of all processes involved. Second order stationarity follows from the fact that we used WN as driving process of the functional ARMA(p, q) equations. Consistency of the empirical estimators of e.g. the covariance operator, however, hold under strict stationarity (see e.g. [10]). Strict stationarity of our models (both functional and vector models) follows immediately from using SWN (cf. Definition 3.1 (ii)) as driving process of the functional ARMA(p, q) equations.

All our results remain valid under this more restrictive condition of SWN driving process with the obvious modifications. □

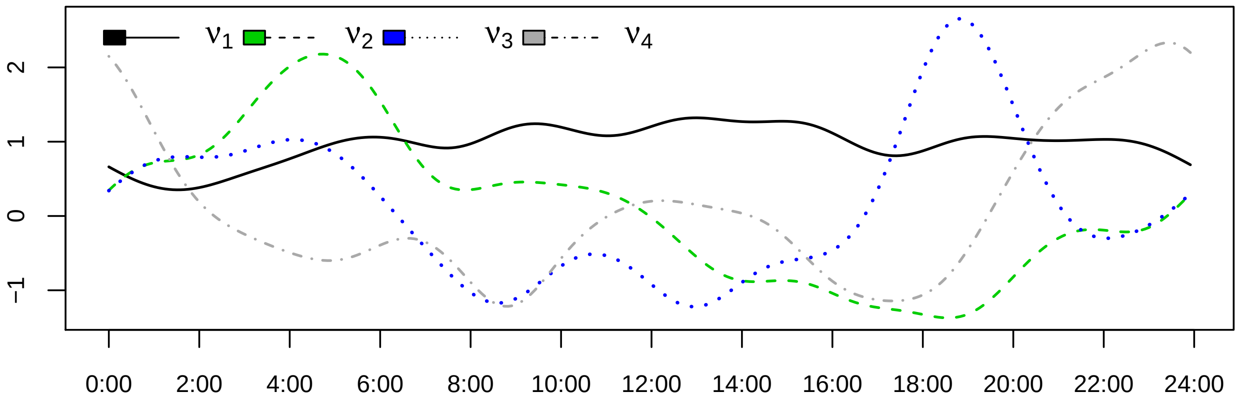

Figure 7: Four empirical eigenfunctions of the $N = 119$ working days functional velocity data. The criterion is 80%; i.e., $\nu_1, \nu_2, \nu_3, \nu_4$ explain together 80% of the total data variability.

Based on the empirical covariance operator with kernel represented in Figure 6 we compute its empirical eigenvalues and eigenfunctions (cf. Figure 7) that we denote by λ_j^e and ν_j^e, for $j = 1, \dots, N$. We are then ready to apply the Algorithm of Section 4.1 to the functional velocity data and implement the following steps.

(1) We apply the CPV method to the highway functional velocity data. From a "CPV(d) vs. d" plot we read off that $d = 4$ functional principal componens explain 80% of the variability of the data. Now for each day $n \in \{1, \dots, N\}$, we use the Karhunen-Loéve Theorem 2.2 and truncate the daily functional velocity curve X_n. This yields

$$X_{n,d} := \sum_{j=1}^{d} \left\langle X_n, \nu_j^e \right\rangle \nu_j^e, \quad n = 1, \dots, N = 119.$$

In Figure 8 we show the (centered) functional velocity data and the corresponding truncation.

We store the $d = 4$ scores in the vector $\mathbf{X}_n$,

$$\mathbf{X}_n = (\langle X_n, \nu_1^e \rangle, \dots, \langle X_n, \nu_4^e \rangle)^\top, \quad n = 1, \dots, N = 119.$$

(2) We now fit different vector ARMA(p, q) models to the multivariate vector data and compare the goodness of fit of the models by their prediction error. We summarize root mean squared errors (RMSE) and mean absolute errors (MAE) for the different models in Table 5.1. To be able to evaluate the performance of the different models, we use standard non-parametric prediction methods from the literature in comparison. All of the linear models significantly outperform methods like exponential smoothing or naively predicting with the mean of the time series. Again details are given in [17]. We find minimal errors for VAR(2) and VMA(1) models, where both prediction errors are equal in case of the MAE, and the RMSE for the VAR(2) model is slightly smaller than that for the VMA(1) model. Since we opt for a parsimonious model, we choose the VMA(1) model, which we fit to the data.

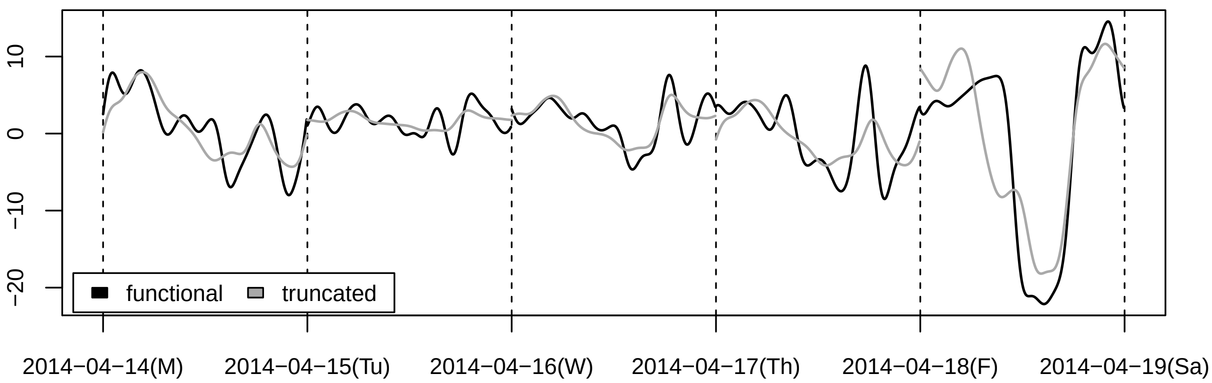

Figure 8: Functional velocity raw data on 5 consecutive working days (black) versus the truncated data by the Karhunen-Loéve representation (grey). The criterion is 80% and the resulting number d of FPC's is 4.

Using the model fit of the VMA(1) model, we compute the best linear predictor $\widehat{\mathbf{X}}_{n+1}$ as in (4.2).

Model fit	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)
RMSE	4.05	**3.87**	3.89	4.78	4.50
MAE	3.19	**3.06**	**3.06**	3.77	3.59

Table 5.1: Average prediction errors of the predictors for the last 10 observations for all working days

(3) We re-transform the vector best linear predictor $\widehat{\mathbf{X}}_{n+1}$ into its functional form $\widehat{X}_{n+1}$, which is depicted in Figure 9.

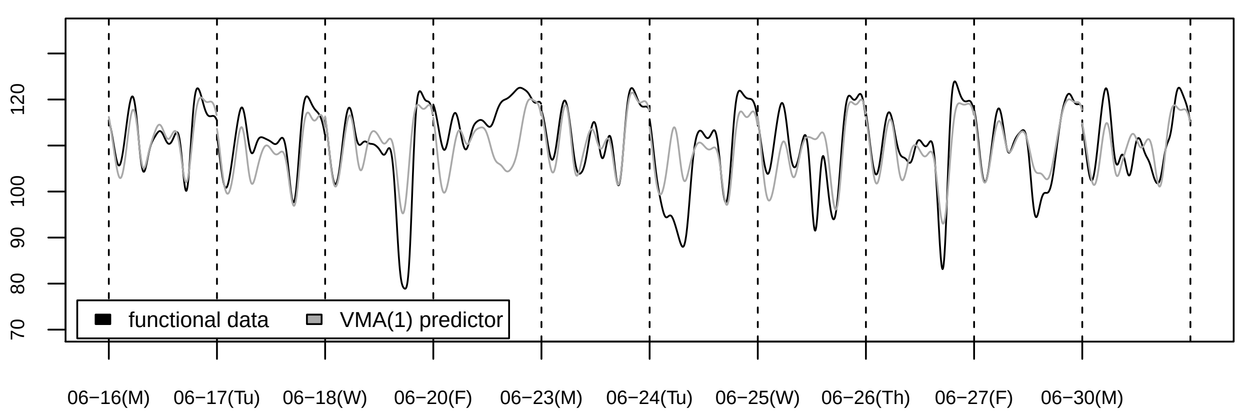

Figure 9: Functional velocity data in black and one-step functional predictor based on VMA(1) in grey for the last working days in June 2014

Acknowledgement: We thank the Autobahndirektion Südbayern and especially J. Grötsch for their support und for providing the traffic data. J. Klepsch furthermore acknowledges financial support from the Munich Center for Technology and Society based

project ASHAD.

References

[1] A. Aue, D. Norinho, and S. Hörmann. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392, 2015.

[2] P. Besse and H. Cardot. Approximation spline de la prévision d'un processus fonctionnel autorégressive d'ordre 1. *Canadian Journal of Statistics*, 24:467–487, 1996.

[3] D. Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Springer New York, 2000.

[4] D. Bosq. General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference*, 137:879–894, 2007.

[5] D. Bosq. Computing the best linear predictor in a Hilbert space. Applications to general ARMA processes. *Journal of Multivariate Analysis*, 124:436–450, 2014.

[6] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods (2nd Ed.)*. Springer, New York, 1991.

[7] L. Debnath and P. Mikusiński. *Introduction to Hilbert Spaces with Applications*. Academic Press, 1999.

[8] N. Dunford and J.T. Schwartz. *Linear Operators Part 1 General Theory*. Wiley, 1988.

[9] R. Gabrys and P. Kokoszka. Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, Vol. 102(No. 480), 2007.

[10] S. Hörmann and P. Kokoszka. Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884, 2010.

[11] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, 2012.

[12] L. Horváth, P. Kokoszka, and G. Rice. Testing stationarity of functional time series. *Journal of Econometrics*, 2013.

[13] T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, 2015.

[14] V. Kargin and A. Onatski. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99:2508–2526, 2008.

[15] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, 2005.

[16] F. Spangenberg. Strictly stationary solutions of ARMA equations in Banach spaces. *Journal of Multivariate Analysis*, 121:127–138, 2013.

[17] T. Wei. Time series in functional data analysis. Master's thesis, Technical University of Munich, 12 2015.