Journal of the Royal Statistical Society Series C: Applied Statistics, 2025, **00**, 1–22
https://doi.org/10.1093/jrsssc/qlaf042

Journal of the Statistics Society Series C Applied Statistics

Original Article

Functional Gaussian graphical regression models for air quality data

Rita Fici[1], Gianluca Sottile[1], Luigi Augugliaro[1] and Ernst C. Wit[2]

[1]Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy
[2]Institute of Computing, Università della Svizzera italiana, Lugano, Switzerland

Address for correspondence: Rita Fici, Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze, 8, Palermo 90128, Italy. Email: rita.fici@unipa.it

Abstract

Functional data capture a wide range of processes, including growth curves and spectral absorption patterns. In this study, we analyse air pollution data from the In-service Aircraft for a Global Observing System, focusing on the spatial interactions among atmospheric chemicals and their dependence on meteorological conditions. This analysis necessitates functional regression, where both response and covariates are functional objects evolving throughout the troposphere. Quantifying both the functional dependencies between the response and covariates and the interdependencies among multivariate response functions poses significant challenges. To address these challenges, we introduce a functional Gaussian graphical regression model, which extends conditional Gaussian graphical models to partially separable functional data. We propose a doubly penalized estimator for model inference and develop a novel adaptation of Kullback–Leibler cross-validation, specifically tailored for graphical estimators. This criterion, named joint Kullback–Leibler cross-validation, simultaneously accounts for both precision and regression matrices, particularly in scenarios where the population comprises multiple sub-groups. Model performance is evaluated in terms of Kullback–Leibler divergence and graph recovery power.

Keywords: air pollutants, conditional graphical models, functional regression, graphical lasso, partial separability

1 Introduction

The study of air, ocean, and land pollution frequently involves analysing large-scale datasets with observations distributed irregularly across spatial domains. Air pollution propagates both horizontally and vertically, influencing air quality at different atmospheric elevations. The vertical transport of chemical species in the atmosphere is a key factor in shaping local meteorological conditions and pollution levels. As highlighted by Ji et al. (2021), understanding air quality dynamics requires monitoring the vertical column density of pollutants within the troposphere. In recent years, numerous studies have investigated vertical pollutant distributions and their interactions with meteorological variables, as these factors exhibit altitude-dependent variability (Sherwood et al., 2010; Uno et al., 2014).

Our objective is to develop statistical tools that facilitate the analysis of chemical interactions as functions of altitude. One of the first applications of functional data analysis (FDA) to air quality monitoring was introduced by Martínez Torres et al. (2020). Building upon this foundation, we aim to extend this framework by modelling the dependencies between pollutants and atmospheric conditions using a multivariate function-on-function regression approach. FDA enables us to capture complex, smoothly varying relationships between atmospheric variables across different altitudes. Additionally, we examine the conditional dependence structure of pollutants while simultaneously accounting for external meteorological influences.

Received: November 13, 2024. **Revised:** May 16, 2025. **Accepted:** July 7, 2025

Several early contributions laid the groundwork for canonical analysis of functional operators. Leurgans et al. (1993) and Dauxois and Nkiet (1997) developed fundamental methods for functional canonical correlation analysis. He et al. (2003) explored the conditions under which canonical correlation is well-defined for functional data, representing functional variables using a canonical basis. Functional regression models, where both response and covariates are univariate functional processes, were later investigated by He et al. (2010), who expressed regression parameters in terms of canonical components. Expanding this framework, Chiou et al. (2016) introduced a multivariate functional linear regression model by employing a vectorized basis expansion, ensuring a structured representation of multivariate functions.

A central tool in FDA is the Karhunen–Loëve (K–L) expansion (Karhunen, 1946), which provides a widely used representation of functional data. In multivariate contexts, this expansion can be extended in various ways. Qiao et al. (2019) proposed a truncated K–L expansion approach, modelling functional covariance structures within a finite-dimensional subspace. Their approach simplifies earlier methodologies (Zhou et al., 2010) that explicitly captured complex functional dependencies. However, a limitation of the truncated K–L expansion approach is its failure to preserve the underlying multivariate functional structure (Chiou et al., 2016).

Recently, Zapata et al. (2021) introduced the concept of *partial separability* for the covariance operator of multivariate functional data. This method defines a structured K–L expansion in which variables share a common basis system, allowing for a more interpretable covariance decomposition. Partial separability builds upon prior developments in functional spatial statistics (Genton, 2007; Gneiting et al., 2006) and functional geostatistics (Delicado et al., 2010; Gromenko et al., 2012). The assumption of separability enables efficient factorization of the covariance structure (Aston et al., 2017; Constantinou et al., 2018; Lynch & Chen, 2018), facilitating statistical inference in high-dimensional functional settings.

Graphical models for multivariate functional data were first proposed by Zhu et al. (2016), who formulated a Markov structure for infinite-dimensional random functions. K.-Y. Lee et al. (2023) later introduced conditional functional graphical models, but their approach is computationally intensive due to the lack of separability constraints. A dynamic functional graphical model allowing for time-varying graph structures was proposed by Qiao et al. (2020). However, as noted by Zapata et al. (2021), existing methods fail to recover the true graph structure unless the functional data are assumed to be finite-dimensional, due to the non-invertibility of the covariance operator. To circumvent this issue, Zapata et al. (2021) proposed the partial separability assumption, ensuring that the inverse covariance operator remains well-defined.

A penalized approach for multivariate Gaussian processes (MGP) was introduced by Qiao et al. (2019), utilizing a block-structured graphical lasso (glasso) estimator (Yuan & Lin, 2007). Zapata et al. (2021) extended this framework, demonstrating that partial separability corresponds to a multivariate K–L expansion structure. They introduced a joint graphical lasso (jglasso) estimator, incorporating a double-penalization scheme to infer both graphical and regression structures simultaneously. Alternative nonparametric approaches, such as the additive model proposed by Li and Solea (2018), relax distributional assumptions but fail to establish direct equivalence with classical conditional independence. More recently, Zhao et al. (2024) proposed a functional neighbourhood selection method (Meinshausen & Bühlmann, 2006), avoiding explicit definition of the precision operator.

Estimating both regression and conditional dependence structures in a multivariate setting was first explored by Rothman et al. (2010) and later extended by W. Lee and Liu (2012) through a doubly penalized maximum likelihood approach. More recently, Sottile et al. (2024) developed a conditional graphical lasso to infer interactions among response and predictor functions, incorporating efficient penalty-based inference strategies for high-dimensional settings. Model selection in graphical models can be performed using information criteria such as the Akaike Information Criterion (AIC) or extended Bayesian Information Criterion (eBIC) (Chen & Chen, 2008; Foygel & Drton, 2010). Vujačić et al. (2015) proposed a Kullback–Leibler cross-validation (KLCV) approach optimized for graphical model selection, demonstrating its finite-sample efficiency and convergence properties.

In this article, we extend the conditional graphical model framework to a functional setting. We propose a doubly penalized estimator and an efficient algorithm to jointly recover the conditional independence structure of response functions and their dependencies on predictor functions. Our approach integrates multivariate and functional aspects while promoting sparsity through penalization. To assess model fit, we introduce a modified Kullback–Leibler Cross-Validation criterion tailored for penalized conditional estimators.

The remainder of this article is organized as follows. Section 2 describes the main aspects of the In-service Aircraft for a Global Observing System (IAGOS) global initiative to monitor pollutants in the atmosphere and the vertical sampling strategy used. In Section 3, we present the methodological aspects of our proposed extension of functional graphical regression models. In Section 4, we provide an estimation procedure to recover the edge set using a conditional group lasso estimator, and we propose a novel selection method for the tuning parameter of the estimator. Section 5 provides computational details and evaluates our approach's performance through a simulation study. Finally, in Section 6, we present the results of our analysis on the IAGOS data.

2 IAGOS: high-resolution air quality monitoring

The IAGOS is a European research infrastructure designed to collect high-resolution atmospheric data through instruments installed on commercial aircraft. These instruments continuously measure key air quality parameters, including trace gas concentrations, aerosols, meteorological variables, and geolocation data (latitude, longitude, and altitude). The dataset, publicly accessible at https://iagos.aeris-data.fr, provides an extensive global record of atmospheric composition across a wide range of altitudes.

IAGOS operates in collaboration with multiple international airlines, equipping long-haul passenger aircraft with the IAGOS-CORE rack, a fully automated measurement system. This system, certified by the European Aviation Safety Agency (EASA), is installed within the fuselage and utilizes dedicated inlet probes mounted on a specialized plate, ensuring precise sampling of atmospheric components without interfering with aircraft operations. The sensors record pollutant concentrations and meteorological parameters at a temporal resolution of 4 s, allowing for fine-scale analysis of air quality dynamics across altitudes.

For this study, we consider data from 75 flights conducted in 2020, with each flight serving as an independent statistical unit. Figures 1 and 2 illustrate the geographical trajectories and altitude profiles of these flights. On average, each flight provides approximately 2,450 measurement points spanning altitudes up to 15,000 m. However, to ensure a highly dense observation grid, we restrict our analysis to data collected below 13,000 m.

The study focuses on five key air quality indicators: carbon monoxide (CO), ozone (O_3), nitrogen monoxide (NO), water vapour (H_2O), and air temperature. Each of these variables plays a critical role in atmospheric chemistry:

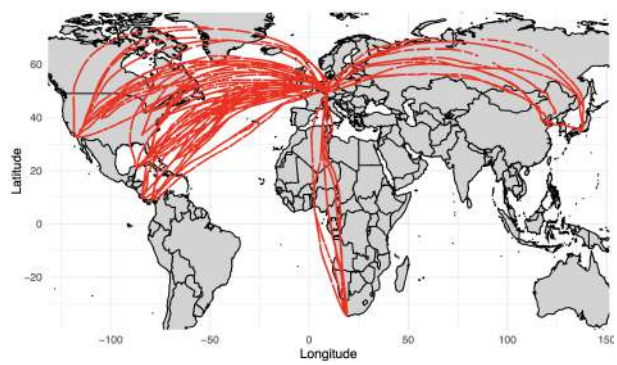

Figure 1. Flight trajectories from the IAGOS dataset.

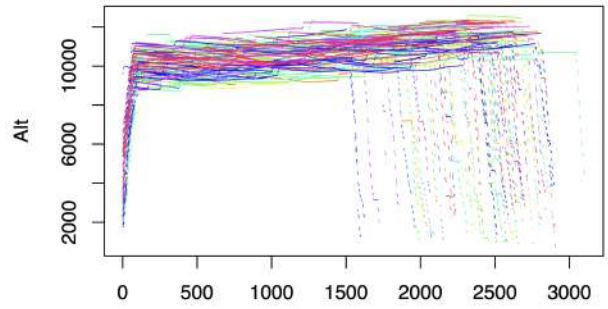

Figure 2. Observed flight altitudes (in metres).

- CO: A byproduct of incomplete fossil fuel combustion, serving as a tracer for anthropogenic pollution and a precursor to ground-level ozone formation.
- O_3: A highly reactive gas with contrasting roles; in the stratosphere, it protects against ultraviolet radiation, while at lower altitudes, it is a harmful pollutant affecting human health.
- NO: A key precursor in ozone formation, strongly linked to vehicular and industrial emissions.
- H_2O: A fundamental component affecting atmospheric transport and pollutant dispersion, with concentration levels varying significantly with altitude.
- Air temperature: A primary meteorological variable influencing chemical reaction rates and atmospheric mixing processes.

Figure 3 depicts the vertical distribution of these atmospheric components. Notably, CO and O_3 exhibit peak concentrations at higher altitudes, while NO and H_2O display an inverse trend. The distribution of measurement points along the altitude axis, particularly the denser clustering at elevated levels, is consistent with the altitude distribution shown in Figure 2. The mean vertical distance between consecutive observations is approximately 8 m, ensuring a fine-grained representation of atmospheric variability.

3 Functional Gaussian graphical regression

In this study, we model the response random functions, denoted by $\boldsymbol{Y}$, which represent the vertical concentration profiles of O_3, CO, NO, and H_2O. The predictor set, $\boldsymbol{X}$, consists of a single covariate: the vertical temperature profile. To ensure methodological generality, we adopt a vector notation for both $\boldsymbol{Y}$ and $\boldsymbol{X}$. Observations are collected over an irregular and discrete grid of altitudes, varying across statistical units. FDA provides a robust framework for capturing complex dependencies among environmental variables evolving smoothly over continuous domains.

We consider two multivariate stochastic processes with functional components, $\boldsymbol{Y} = \{\boldsymbol{Y}(s) \in \mathbb{R}^p : s \in \mathcal{S}\}$ and $\boldsymbol{X} = \{\boldsymbol{X}(s) \in \mathbb{R}^q : s \in \mathcal{S}\}$, observed over a closed and continuous altitude domain $\mathcal{S} \subset \mathbb{R}$. These processes reside in Hilbert spaces, $\mathbb{H}_Y$ and $\mathbb{H}_X$, containing p- and q-dimensional vectors of square-integrable functions, $\mathcal{L}_2^p[\mathcal{S}]$ and $\mathcal{L}_2^q[\mathcal{S}]$, respectively.

Observed data consist of discretized and noisy realizations of the underlying smooth functions:

$$Y_j^o(s) = Y_j(s) + e_j^y(s) \quad \text{and} \quad X_k^o(s) = X_k(s) + e_k^x(s) \text{ for } j = 1, \ldots, p \quad \text{and} \quad k = 1, \ldots, q;$$

where $Y_j^o(s)$ and $X_k^o(s)$ denote the observed versions of $Y_j(s)$ and $X_k(s)$ at discrete altitudes, and $e_j^y(s)$ and $e_k^x(s)$ capture measurement errors. Due to substantial observational noise in atmospheric data, functional variables cannot be directly measured. The subsequent sections describe the smooth functional modelling of $\boldsymbol{Y}$ and $\boldsymbol{X}$, while Section 4.1 details the estimation of error terms.

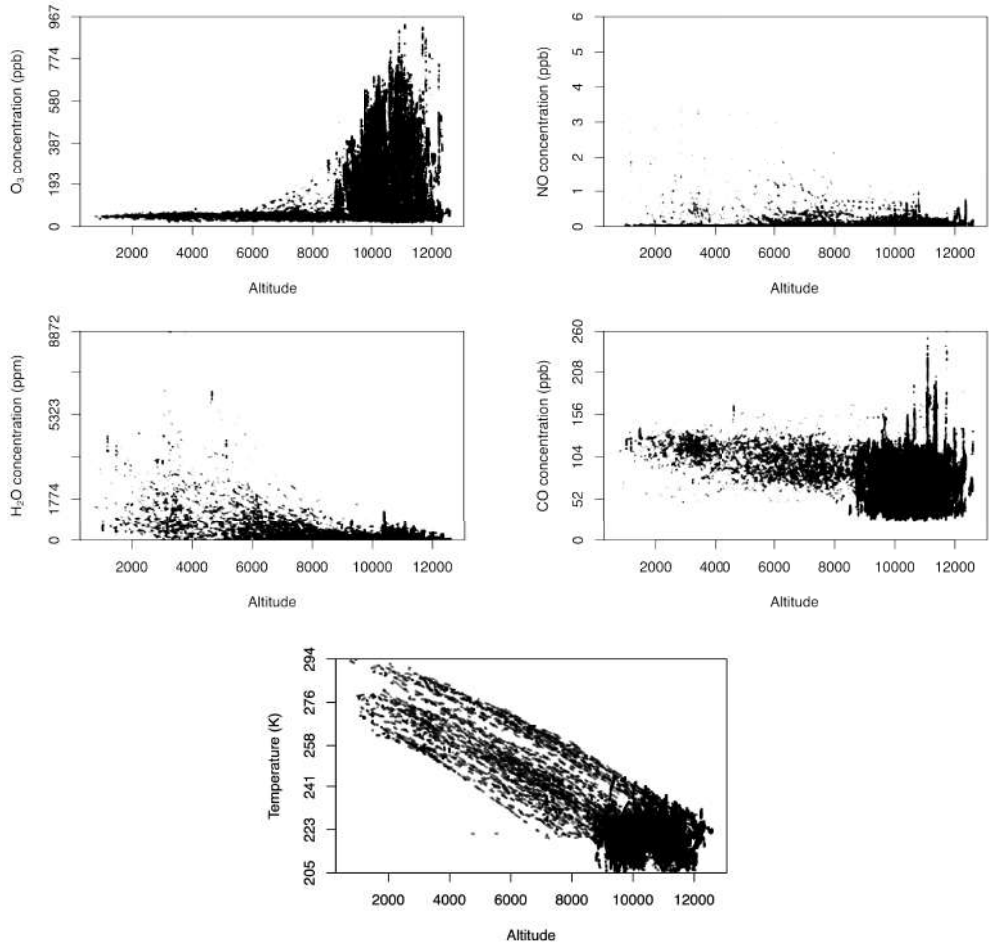

Figure 3. Observed concentrations of atmospheric constituents and temperature across altitude levels. O_3, NO, and CO are measured in parts per billion (ppb), H_2O in parts per million (ppm), and temperature in Kelvin (K).

3.1 Multivariate functional regression model

In this section, we define a functional regression model where the sum of two components gives the response process:

$$\boldsymbol{Y}(t) = \int_{\mathcal{S}} \boldsymbol{\beta}(t, s)\boldsymbol{X}(s)\,\mathrm{d}s + \boldsymbol{K}(t), \quad t \text{ and } s \in \mathcal{S} \tag{1}$$

The first element on the right-hand side of (1) is a structural component that considers the effects of the covariates on the response. It involves the covariate process $\boldsymbol{X}$ and a set of bivariate regression coefficient functions $\boldsymbol{\beta}(t, s) \in \mathcal{L}_2^{p\times q}[\mathcal{S}\times\mathcal{S}]$, which represent how the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ changes over $\mathcal{S}\times\mathcal{S}$. If the L_2 norm of β_{ik} vanishes, $\|\beta_{ik}\| = 0$, then Y_i is independent of X_k, given $\boldsymbol{X}_{-k}$. If $\|\beta_{ik}\| \neq 0$, then X_k is a predictor of Y_i. The term $\boldsymbol{K}(t)$ represents the noise component, independent of $\boldsymbol{X}$, and encodes the mutual dependencies among the responses, net of the effect of $\boldsymbol{X}$.

The two elements on the right-hand side of (1) are associated with a set of directed links and a set of undirected links, respectively. The first set represents the conditional dependencies of the responses on the covariates and is defined as

$$\mathcal{E}^{XY} = \{(i, k) : Y_i \perp X_k \mid \boldsymbol{Y}_{-i}, \boldsymbol{X}_{-k}\}$$

The second set represents the pairwise conditional independence structure of $\boldsymbol{Y}$ and is given by

$$\mathcal{E}^{\boldsymbol{Y}} = \{(i, j) : Y_i \perp Y_j \mid \boldsymbol{Y}_{-ij}, \boldsymbol{X}\} = \{(i, j) : K_i \perp K_j \mid \boldsymbol{K}_{-ij}\}$$

Note that conditional independence holds if the functions are conditionally independent over the entire domain $\mathcal{S} \times \mathcal{S}$. These sets of edges define a chain graphical model with two components denoted by $\mathcal{G} = \{\mathcal{V}, \mathcal{E}^{\boldsymbol{XY}}, \mathcal{E}^{\boldsymbol{Y}}\}$, where $\mathcal{V}$ has cardinality $p + q$. The directed edges in $\mathcal{E}^{\boldsymbol{XY}}$ correspond to the first level factorization of the probability density f_{XY}, which for two chain components is

$$f\ \boldsymbol{X}, \boldsymbol{Y}\ = f_X(\boldsymbol{X}) f_Y(\boldsymbol{Y} \mid \boldsymbol{X}_{pa}), \tag{2}$$

where $\boldsymbol{X}_{pa}$ denotes the set of elements of $\boldsymbol{X}$ that affect $\boldsymbol{Y}$. The undirected links in $\mathcal{E}^{\boldsymbol{Y}}$ correspond to the second-level factorization

$$f(\boldsymbol{Y} \mid \boldsymbol{X}_{pa}) = \prod_{C \in \text{ Cliques}} f_C(\boldsymbol{Y}_C \mid \boldsymbol{X}_{pa}), \tag{3}$$

where C are the cliques in $\mathcal{E}^{\boldsymbol{Y}}$. Although in this analysis, we are not directly interested in the factorization of $f_X(\boldsymbol{X})$, the proposed framework can be extended to include a graphical description of $\boldsymbol{X}$. We assume here that $\boldsymbol{X}$ and $\boldsymbol{K}$ are MGP with mean zero.

3.2 Auto-covariance operator

To define the dependence structure on the Hilbert space and extend the definition of edge set related to a single graph, we consider the correlation among the functions across the entire domain, as suggested by Qiao et al. (2019). For the vector of response variables, the covariance is defined (Happ & Greven, 2018) as a matrix where each entry involves an integral operator with a symmetric and nonnegative definite kernel equal to

$$C_{ij}(t, t') = \text{cov}(Y_i(t), Y_j(t'))\, t \quad \text{and} \quad t' \in \mathcal{S} \tag{4}$$

The auto-covariance operator of the process is defined as

$$(\mathcal{C}\boldsymbol{Y})(t) = \int_{\mathcal{S}} C(t, t') \boldsymbol{Y}(t')\, \mathrm{d}t', \tag{5}$$

where $C(t, t')$ is the $p \times p$ symmetric matrix which collects the kernels (4) for $i, j = 1, \ldots, p$. If $\mathcal{C}$ is **partially separable** (Zapata et al., 2021), there exists a sequence of $p \times p$ dimensional matrices, $\{\Sigma_l\}_{l=1}^{\infty}$, and a set of orthonormal eigenbasis $\{\varphi_l\}_{l=1}^{\infty}$, such that it is possible to rewrite $\mathcal{C}$ as

$$\mathcal{C} = \sum_{l=1}^{\infty} \Sigma_l \varphi_l \otimes \varphi_l, \tag{6}$$

where $\otimes$ is the tensor product and $\{\varphi_l\}_l^{\infty}$ is a set of orthonormal functions in $\mathcal{S}$. The trace of (Σ_l) decreases over l. Thanks to the partial separability assumption, Proposition 2 in Happ and Greven (2018) is satisfied, which makes it possible to represent the multivariate process as,

$$Y(t) = \boldsymbol{\mu}(t) + \sum_{l=1}^{\infty} \boldsymbol{\gamma}_l \varphi_l(t), \tag{7}$$

where $\boldsymbol{\gamma}_l$ is a vector of random scores with dependent entries, $\mathrm{E}[\boldsymbol{\gamma}_l] = 0$ and $\mathrm{Var}[\boldsymbol{\gamma}_l] = \Sigma_l$ with ijth element denoted by σ_{ijl}. The dependence structure of the process is directly related to the dependence among the scores inside each vector $\boldsymbol{\gamma}_l$, i.e. in general $\gamma_{il} \perp \gamma_{jl}$. However, the sequence of vectors of scores are independent of each other, i.e. $\boldsymbol{\gamma}_l \perp\!\!\!\perp \boldsymbol{\gamma}_{l'}$ for $l \neq l'$. It can be noted that the cross-covariance function can be written as $\mathcal{C}_{ij}(t, t') = \sum_l \sigma_{ijl} \varphi_l(t) \varphi_l(t')$. A fundamental advantage of (6) is that, if the scores are normally distributed, the elements in $\{\Sigma_l\}_{l=1}^{\infty}$ identify the pairwise

relations among the variables in the process. Due to the independence of the vectors from each other, these covariance relations can be visualized as a block diagonal infinite-dimensional covariance matrix, denoted by $\boldsymbol{\Sigma}$, easy to invert,

$$\boldsymbol{\Sigma} = \begin{matrix} \Sigma_1 & 0 & \dots \\ 0 & \Sigma_2 & \dots \\ \vdots & \vdots & \ddots \end{matrix} \quad \text{and} \quad \Theta = \boldsymbol{\Sigma}^{-1} = \begin{matrix} \Sigma_1^{-1} & 0 & \dots \\ 0 & \Sigma_2^{-1} & \dots \\ \vdots & \vdots & \ddots \end{matrix}$$

In both matrices, the off-diagonal block entries are equal to zero. In this article, we exploit the advantage of representing some variables in a unique system of orthonormal basis by applying the assumption of partial separability on the autocovariance operator of $\boldsymbol{Y}$, in $\mathbb{H}_Y \times \mathbb{H}_Y$, and on the autocovariance operator of $\boldsymbol{X}$, in $\mathbb{H}_X \times \mathbb{H}_X$. The processes are represented as,

$$\boldsymbol{Y}(t) = \sum_{l=1}^{\infty} \boldsymbol{\gamma}_l \varphi_l(t) \quad \text{and} \quad \boldsymbol{X}(s) = \sum_{h=1}^{\infty} \boldsymbol{\chi}_h \psi_h(s) \tag{8}$$

3.3 Connection of parametrization with graphical structure

Here, we derive the graphical structure of the multivariate functional regression formulation. In particular, it constitutes a chain graph model, whereby we assume a directed graphical structure between the covariates and the response and an undirected graphical structure between the response variables.

3.3.1 Identification of covariate-response conditional independence graph

In factorization (2), the set $\boldsymbol{X}_{pa}$ is identified by the regression functions such that X_k is a parent of Y_i if and only if $\|\beta_{ik}\| \neq 0$. In this section, we derive a representation of the parental set through the normally distributed score vectors $\boldsymbol{\gamma}_l$ and $\boldsymbol{\chi}_h$ in (8).

Theorem 1 Let $\boldsymbol{Y}$ and $\boldsymbol{X}$ be two multivariate Gaussian functional processes, their regression dependence can be represented by the coefficient functions $\boldsymbol{\beta}$ which satisfy three characterizations:

1. the linear model

$$\boldsymbol{Y}(t) = (\mathcal{B}\boldsymbol{\beta}_0)(t) + \boldsymbol{K}(t), \tag{9}$$

where $\mathrm{E}[\|\boldsymbol{Y}(t)\|_{\mathbb{H}_Y} \mid \boldsymbol{X}] < \infty$ and $\mathcal{B}: \mathbb{H}_X \times \mathbb{H}_Y \to \mathbb{H}_Y$ is the regression operator defined as

$$(\mathcal{B}\boldsymbol{\beta})(t) = \int_{\mathcal{S}} \boldsymbol{\beta}(t, s)\boldsymbol{X}(s)\, \mathrm{d}s;$$

2. the population normal equation

$$\mathrm{cov}[\boldsymbol{Y}(t), \boldsymbol{X}(s)] = (\mathcal{D}_{XX}\boldsymbol{\beta}_0)(t, s), \tag{10}$$

where $\mathcal{D}_{XX}: \mathbb{H}_X \times \mathbb{H}_X \to \mathbb{H}_Y \times \mathbb{H}_X$ is the cross-covariance operator defined as,

$$(\mathcal{D}_{XX}\boldsymbol{\beta})(t, s) = \int_{\mathcal{S}} \boldsymbol{\beta}(t, s)\mathrm{cov}[\boldsymbol{X}(s), \boldsymbol{X}(s')]\, \mathrm{d}s';$$

where s' is a point in $\mathcal{S}$;

3. and the optimization on the squared norm of the difference between $\boldsymbol{Y}$ and $(\mathcal{B}\boldsymbol{\beta})$ in the response Hilbert space,

$$\boldsymbol{\beta}_0 = \underset{\beta}{argmin}\, \mathrm{E}\|\boldsymbol{Y} - (\mathcal{B}\boldsymbol{\beta})\|^2_{\mathbb{H}_Y} \tag{11}$$

Theorem 1 follows from the generalization of Proposition 2.2 in He et al. (2010). In He et al. (2010), the authors define the linear regression coefficient between two functions. Theorem 1

defines a linear multivariate regression parameter with multiple covariates. As the operator $\mathcal{D}_{XX}$ does not have a bounded inverse, we use the vectorized K–L expansion to extend the results of Theorem 2.3 in He et al. (2010) to the multivariate case. This allows us to derive a unique solution that satisfies (9), (10), and (11). Considering the expansions (8) with $\sum_{l,h=1}^{\infty} \|\mathrm{E}[\gamma_l \chi_h^\top]\mathrm{E}[\chi_h \chi_h^\top]^{-1}\|^2 < \infty$, we now use Theorem 1 to characterize the regression functions as

$$\boldsymbol{\beta}(t, s) = (\mathcal{D}_{XX}^{-1}\mathrm{cov}[\boldsymbol{Y}, \boldsymbol{X}])(t, s), \tag{12}$$

where $\mathcal{D}_{XX}^{-1}$ exists for a subdomain of $\mathcal{D}_{XX}$ and takes the form

$$(\mathcal{D}_{XX}^{-1} \cdot)(t, s) = \sum_{h,l=1}^{\infty} \langle \cdot, \boldsymbol{\varphi}_l \otimes \boldsymbol{\psi}_h \rangle \mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1} \odot \ \boldsymbol{\varphi}_l(t)\boldsymbol{\psi}_h^\top(s) \ , \tag{13}$$

where $\odot$ is the Hadamard product, $\{\boldsymbol{\varphi}_l \otimes \boldsymbol{\psi}_h\}_{l,h=1}^{\infty}$ is a orthonormal basis system in $\mathbb{H}_Y \times \mathbb{H}_X$ and, for two generic set of functions $\boldsymbol{g}, \boldsymbol{g}' \in \mathcal{L}_2^{p \times q}$, the inner product $\langle \boldsymbol{g}, \boldsymbol{g}' \rangle$ is qual to $\int_{\mathcal{S}}\int_{\mathcal{S}} \boldsymbol{g}(s, t) \odot \boldsymbol{g}'(s, t)\, \mathrm{d}t\, \mathrm{d}s$. As we apply the vectorized K–L expansion, $\boldsymbol{\varphi}_l$ and $\boldsymbol{\psi}_h$ represent sets of p- and q-dimensional functions, respectively, where all functions within each set are identical. Plugging in the covariance functions into $\langle \cdot, \boldsymbol{\varphi}_l \otimes \boldsymbol{\psi}_h \rangle$, we have

$$\int_{\mathcal{S}}\int_{\mathcal{S}} \mathrm{cov}[\boldsymbol{Y}(t), \boldsymbol{X}(s)] \odot \ \boldsymbol{\varphi}_l(t)\boldsymbol{\psi}_h^\top(s) \ \mathrm{d}s\, \mathrm{d}t = \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top] \tag{14}$$

Thanks to (13) and (14), we can write (12) as

$$\boldsymbol{\beta}(t, s) = \sum_{h,l=1}^{\infty} \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top]\mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1} \odot \ \boldsymbol{\varphi}_l(t)\boldsymbol{\psi}_h^\top(s) \tag{15}$$

From (1), (8), and (15), we have

$$\begin{aligned}
\mathrm{E}[\boldsymbol{Y}(t) \mid \boldsymbol{X}] &= \int_{\mathcal{S}} \boldsymbol{\beta}_0(s, t)\boldsymbol{X}(s)\, \mathrm{d}s \\
&= \int_{\mathcal{S}} \sum_{h,l=1}^{\infty} \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top]\mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1}\varphi_l(t)\psi_h(s) \sum_{h'=1}^{\infty} \boldsymbol{\chi}_{h'}\psi_{h'}(s)\, \mathrm{d}s \\
&= \sum_{h,l=1}^{\infty} \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top]\mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1}\varphi_l(t) \sum_{h'=1}^{\infty} \boldsymbol{\chi}_{h'} \int_{\mathcal{S}} \psi_h(s)\psi_{h'}(s)\, \mathrm{d}s \\
&= \sum_{h,l=1}^{\infty} \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top]\mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1}\boldsymbol{\chi}_h\varphi_l(t),
\end{aligned}$$

where $\int_{\mathcal{S}} \psi_h(s)\psi_{h'}(s)\, \mathrm{d}s$ is equal to 1 for $h = h'$ and equal to 0 for $h \neq h'$.

It follows that the conditional expectation of $\boldsymbol{Y}(t)$ can be written as

$$\mathrm{E}[\boldsymbol{Y}(t) \mid \boldsymbol{X}] = \int_{\mathcal{S}} \boldsymbol{\beta}(t, s)\boldsymbol{X}(s)\, \mathrm{d}s = \sum_{h,l=1}^{\infty} \boldsymbol{B}_{hl}\boldsymbol{\chi}_h\varphi_l(t),$$

where $\boldsymbol{B}_{hl} = \mathrm{E}[\boldsymbol{\gamma}_l \boldsymbol{\chi}_h^\top]\mathrm{E}[\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top]^{-1}$. Therefore, we have now shown that the edge set $\mathcal{E}^{XY}$ is characterized by

$$\mathcal{E}^{XY} = \{(i, k) \mid \exists l, h : b_{hl\,ik} \neq 0\},$$

where $b_{lh\,ik}$ is the ikth entry of $\boldsymbol{B}_{hl}$.

3.3.2 *Identification of response conditional independence graph*

Given that the conditional distribution $f_Y(\boldsymbol{Y} \mid \boldsymbol{X}_{pa})$ is a MGP the conditional independence structure can be expressed through the zero pattern on the inverse covariance operator. The covariance of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is equal to the covariance of $\boldsymbol{K}$, and it has kernel equal to

$$\boldsymbol{C}^{Y|X}(t,t') = \operatorname{cov}\ Y_i(t), Y_j\ t' \mid \boldsymbol{X}(\cdot)\ \ {}^{p}_{i,j=1} = \operatorname{cov}\ K_i(t), K_j\ t' \ \ {}^{p}_{i,j=1} = \boldsymbol{C}^{K}(t,t')$$

Thanks to Theorem 3 in Zapata et al. (2021), we have the partial correlation between K_i and K_j related to the precision matrices of the scores $\{\boldsymbol{\Theta}^{\gamma}_l\}_{l=1}^{\infty} = \{(\boldsymbol{\Sigma}^{\gamma}_l)^{-1}\}_{l=1}^{\infty}$

$$\operatorname{cov}(K_i(t), K_j(t') \mid \boldsymbol{K}_{-ij}(\cdot)) = \sum_{l=1}^{\infty} \operatorname{cov}(\gamma_{li}, \gamma_{lj} \mid \boldsymbol{\gamma}_{l-ij})\varphi_l(t)\varphi_l(t') = -\sum_{l=1}^{\infty} \frac{\theta_{lij}\varphi_l(t)\varphi_l(t')}{\theta_{lii}\theta_{ljj} - \theta^2_{lij}},$$

where θ_{lij} is an entry of $\boldsymbol{\Theta}^{\gamma}_l$, which, thanks to the K–L expansion, does not depend on $\mathcal{S}$. We can identify $\mathcal{E}_Y$ as

$$\mathcal{E}^Y = \{(i\,j) \mid \exists l : \theta^{\gamma}_{lij} \neq 0\}$$

4 Functional Gaussian graphical regression estimation

We observe N independent observations of the noisy curves $\{\boldsymbol{y}_n\}_{n=1}^{N}$ and $\{\boldsymbol{x}_n\}_{n=1}^{n}$ at discrete locations $s \in \mathcal{S}$. In Section 4.1, we estimate the underlying true smooth realizations $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the functional variables. Then in Section 4.2, we derive an estimator of relevant functional predictors $\boldsymbol{X}$, as well as the conditional dependence structure on $\boldsymbol{Y}$.

4.1 Evaluation of the scores

For each variable and each unit, we estimate the underlying true functional form via a non-linear regression on the spatial dimension. We use penalized splines to model the relationship between variables and altitude. The level of smoothing is automatically selected by the **gam** function implemented in **mgcv** library in the statistical programming language `R`. Figure 4 shows the resulting smooth functional observations without the noise.

Knowing the full functions $z_n = \{\boldsymbol{y}_n, \boldsymbol{x}_n\}$ over the space $\mathcal{S}$, and assuming the partial separability of $(\mathcal{C}z)$, it is possible to define a unique system of basis for $\boldsymbol{Y}$ and $\boldsymbol{X}$, therefore $\varphi_l = \psi_l$ for all l. From Theorem 2 in Zapata et al. (2021), we know that under partial separability assumption, the eigenvalues of the trace-class covariance operator $\mathcal{H} = (p+q)^{-1} \sum_{i=1}^{p+q} \mathcal{C}^{\boldsymbol{X},Y|\boldsymbol{X}}_{ii}$, in the eigenspace spanned by $\{\varphi_l\}_{l=1}^{\infty}$, is equal to the sum of the variances in $\{\langle z_j, \varphi_l\rangle\}_{j=1}^{p+q}$. The eigenbasis of $\mathcal{H}$ are optimal for preserving the maximum amount of overall variability between the orthonormal basis system and the functions and for choosing the number of expansion terms which capture a percentage of the total variability of the processes. The eigenvalues of $\mathcal{H}$ follow a non-increasing order, and for the Lth element of the expansion, the corresponding eigenvalue approaches zero. We evaluate the operator $\mathcal{H}$ fixing grid of 420 points over the altitude, from 1 to 1,300 m, with steps of 30,

$$(\mathcal{H}\mathbf{z})(t,t') = \frac{1}{5}\sum_{j=1}^{4} \frac{\sum_{n=1}^{75} w_{jn}(t)[z_{jn}(t) - \mu_j(t)][z_{jn}(t') - \mu_j(t')]w_{jn}(t')}{\sum_{n=1}^{75} w_{jn}(t)w_{jn}(t')},$$

where $t = 1, \ldots, 420$, $w_{jn}(t)$ is a weight, and μ_j is the mean function of the jth variable. The weights are calculated as the inverse of the variances of the estimated functions, and $\mu_j(t) = \frac{\sum_{n=1}^{75} w_{jn}(t)z_{jn}(t)}{\sum_{n=1}^{75} w_{jn}(t)}$. The eigen-decomposition of $\mathcal{H}\mathbf{z}$ gives a discrete version of the basis system. The first 5 basis functions explain more than the 99% of the total variability. Indeed, each variable for each unit is represented by a vector of 5 scores equal to

$$\boldsymbol{\gamma}_{jn} = [\boldsymbol{\varphi} W_{jn} \boldsymbol{\varphi}^{\top}]^{-1} \boldsymbol{\varphi} W_{jn} \mathrm{y}_{jn} \quad \text{for} \quad j = 1, \ldots, 4,$$

where $\boldsymbol{\varphi} = [\boldsymbol{\varphi}_1 \mid \cdots \mid \boldsymbol{\varphi}_5]^{\top}$ is a 5×420 matrix that collects the values of the eigen-functions, $\mathrm{y}_{jn} = \mathrm{y}_{jn} - \boldsymbol{\mu}_j$ is the vector of centred observations of length 420, and W_{jn} is a 420×420 matrix where

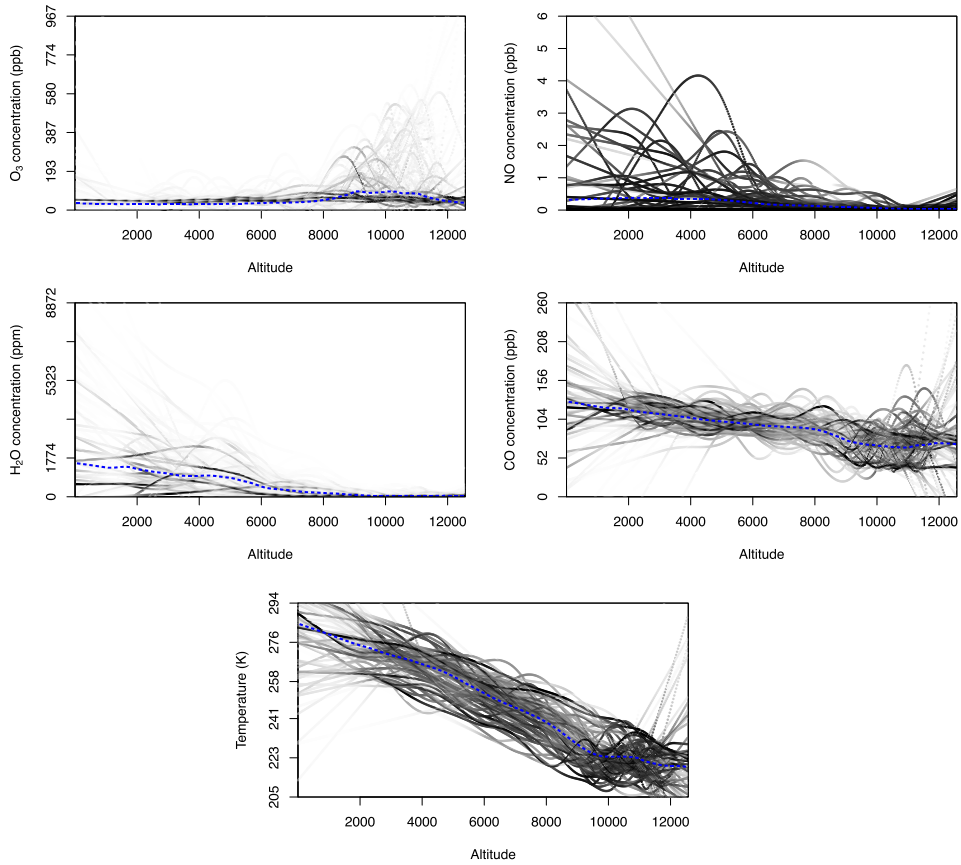

Figure 4. Smooth functional observations $\boldsymbol{y}_n$ and $\boldsymbol{x}_n$ for $n = 1, \ldots, N$. For each altitude value, the points are coloured with a grey scale proportional to the weights $w_{jn}(s)$ for $n = 1, \ldots, N$. In The dashed line indicates the mean function $\boldsymbol{\mu}_j$. O_3, NO, and CO are measured in parts per billion (ppb), while H_2O is measured in parts per million (ppm). Temperature is measured in Kelvin (K).

the diagonal entries are equal to $w_{jn}(t)$ for $t = 1, \ldots, 420$, and the off-diagonal elements are zero. The same is done to estimate the explanatory scores,

$$\boldsymbol{\chi}_n = [\boldsymbol{\varphi} W_{X_n} \boldsymbol{\varphi}^\top]^{-1} \boldsymbol{\varphi} W_{X_n} \mathbf{x}_n,$$

where similarly W_{X_n} is the diagonal matrix of weights, and $\mathbf{x}_n$ is the vector of 420 centred temperature values.

In cases where the functions are not observed densely and the observations are affected by noise, the scores need to be estimated. However, due to the use of an orthonormal basis, these estimates remain unbiased with independent noise terms. This only impacts the marginal variance of the scores without affecting the conditional independence relationships in the underlying graphs. Therefore, as long as each function is observed at more points than the dimension of the underlying functional basis, our estimation procedure remains consistent.

Figure 5 shows the kernel estimation of the densities of the scores for $l = 1, 2$. For each variable and l, the scores are divided by their standard deviation. The assumption of Gaussian distributions appears not completely inappropriate, although the estimated densities present some asymmetries and kurtosis.

4.2 fGGRM estimator

The scores γ_{ln} and $\boldsymbol{\chi}_{ln}$ are assumed to follow multivariate Gaussian distributions. Thanks to the assumption of partial separability of $(\mathcal{C}z)$, the joint distribution is specified as,

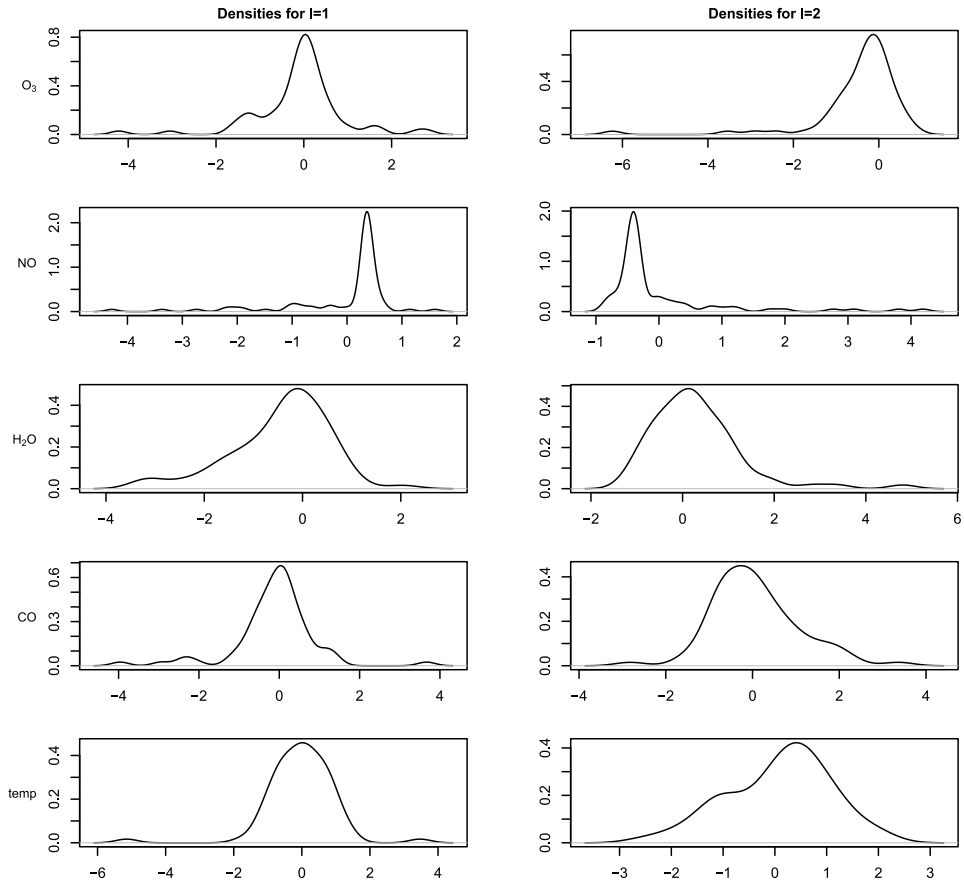

Figure 5. Estimated scores densities for the first and the second term of the expansion.

$$\begin{matrix}\boldsymbol{\chi}_l \\ \boldsymbol{\gamma}_l\end{matrix} \sim \mathcal{N}_{q+p}\ \ 0,\ \ \begin{matrix}\boldsymbol{\Sigma}_l^{\chi} & \boldsymbol{\Sigma}_l^{\chi}\boldsymbol{B}_l \\ \boldsymbol{B}_l^{\top}\boldsymbol{\Sigma}_l^{\chi} & \boldsymbol{B}_l^{\top}\boldsymbol{\Sigma}_l^{\chi}\boldsymbol{B}_l + \boldsymbol{\Sigma}_l^{\epsilon}\end{matrix} \tag{16}$$

The precision matrix of that density is given as,

$$\Theta_l = \begin{matrix}\boldsymbol{\Theta}_l^{\chi} + \boldsymbol{B}_l\boldsymbol{\Theta}_l^{\gamma}\boldsymbol{B}_l^{\top} & -\boldsymbol{B}_l\boldsymbol{\Theta}_l^{\gamma} \\ -\boldsymbol{\Theta}_l^{\gamma}\boldsymbol{B}_l^{\top} & \boldsymbol{\Theta}_l^{\gamma}\end{matrix}, \tag{17}$$

where $\boldsymbol{\Theta}_l^{\chi}$ and $\boldsymbol{\Theta}_l^{\gamma}$ are the inverses of the marginal covariance matrices of $\boldsymbol{\chi}_l$ and $\boldsymbol{\gamma}_l$, respectively. Because of partial separability, the densities (16) are independent over l.

We propose to estimate a functional Gaussian graphical regression model (fGGRM) from $\{\boldsymbol{\chi}_{hn}, \boldsymbol{\gamma}_{ln}\}_{lhn}$, using the double-penalized estimator proposed in Sottile et al. (2024), named in this context as *functional joint conditional graphical lasso* estimator,

$$\{B\}, \{\Theta^{\gamma}\} = \underset{\{B\},\{\Theta^{\gamma}\}}{\operatorname{argmax}} \sum_{l=1}^{L} \log\det(\Theta_l^{\gamma}) - \operatorname{tr}(S(B_l)\Theta_l^{\gamma}) \ - \nu P_1(\{B\}) - \rho P_2(\{\Theta^{\gamma}\}), \tag{18}$$

where

$$S(\boldsymbol{B}_l) = \frac{1}{N}\sum_{n=1}^{N} (\boldsymbol{\gamma}_{ln} - \boldsymbol{B}_l\boldsymbol{\chi}_{ln})^{\top}(\boldsymbol{\gamma}_{ln} - \boldsymbol{B}_l\boldsymbol{\chi}_{ln}) \tag{19}$$

The penalty functions $P_1(\cdot)$ and $P_2(\cdot)$ are convex functions that encourage sparsity in each expansion matrix and specific forms of similarity across the expansion elements. We use the group lasso

penalty functions. The penalty for the regression coefficient matrices and for the precision matrices are

$$P_1(\{B\}) = \sum_{i=1}^{p}\sum_{k=1}^{q}\Big(\sum_{l=1}^{L} b_{lki}^2\Big)^{1/2}, \quad \text{and} \quad P_2(\{\Theta\}) = \sum_{i\neq j}^{p}\Big(\sum_{l=1}^{L} \theta_{lij}^2\Big)^{1/2}$$

Thus, the desired edge sets can be estimated by

$$\mathcal{E}_{\nu\rho}^{XY} = \{(k,i) : \sum_{l=1}^{L} (\hat{b}_{lki}^{\rho\nu})^2 > 0\} \quad \text{and} \quad \mathcal{E}_{\nu\rho}^{Y} = \{(i,j) : \sum_{l=1}^{L} (\hat{\theta}_{lij}^{\rho\nu})^2 > 0\},$$

where $\hat{b}_{lki}^{\rho\nu}$ and $\hat{\theta}_{lij}^{\rho\nu}$ are the matrices entries estimated under the couple of penalization values (ρ, ν).

4.3 Goodness of fit

Kullback–Leibler (KL) divergence serves as a non-symmetric measure to quantify the discrepancy between two probability distributions (Penny, 2001). It gauges the information loss that occurs when one probability distribution is employed to approximate another. The KL measure is defined as:

$$\mathrm{KL}(f_1 \mid f_2) = \mathrm{E}_{f_1}[\log f_1 - \log f_2], \tag{20}$$

where f_1 and f_2 are two probability density functions. In our study, we aim to evaluate the information loss between the true multivariate normal distribution (16) and the estimated distributions derived from (18) under various combinations of tuning parameters values. The couple (ρ, ν) dictate the strength of penalization, leading to distinct estimations of $\{\Theta^{\gamma}\}$, $\{\Theta^{\chi}\}$, and $\{\boldsymbol{B}\}$, which together specify the normal distribution we seek to compare with the true distribution. As shown in Penny (2001), the KL divergence between two zero-mean multivariate normal distributions can be written as

$$\mathrm{KL}_{\mathcal{N}_m}[\boldsymbol{\Theta} \mid \boldsymbol{\Theta}] = \frac{1}{2}\{\mathrm{tr}(\boldsymbol{\Theta}^{-1}\boldsymbol{\Theta}) - \log \mid \boldsymbol{\Theta}^{-1}\boldsymbol{\Theta} \mid - m\} \tag{21}$$

As shown in Section (3), the density function of the joint vector can be factorized as the product between the explanatory and the conditioned response density functions

$$f(\boldsymbol{\chi}_l, \boldsymbol{\gamma}_l \mid \boldsymbol{\chi}_l; \boldsymbol{\Theta}_l) = f(\boldsymbol{\chi}_l; \boldsymbol{\Theta}_l^{\chi}, \boldsymbol{B}_l) f(\boldsymbol{\gamma}_l \mid \boldsymbol{\chi}_l; \boldsymbol{\Theta}_l^{\gamma}, \boldsymbol{B}_l) \tag{22}$$

The KL loss can be calculated as the sum of L divergences

$$\begin{aligned} &\mathrm{KL}_{\mathcal{N}_{q+p}}\{\boldsymbol{\Theta}^{\chi}\}\{\boldsymbol{\Theta}^{\gamma}\}\{\boldsymbol{B}\} \mid \{\boldsymbol{\Theta}^{\chi}\}_{\rho\nu}\{\boldsymbol{\Theta}^{\gamma}\}_{\rho\nu}\{\boldsymbol{B}\}_{\rho\nu} \\ &= \frac{1}{2}\sum_{l=1}^{L}\{\mathrm{tr}(\boldsymbol{\Psi}_l^{-1}\boldsymbol{\Psi}_{l\rho\nu}) - \log \mid \boldsymbol{\Psi}_l^{-1}\boldsymbol{\Psi}_{l\rho\nu} \mid \\ &\quad + \mathrm{tr}(\boldsymbol{\Theta}_l^{\gamma\,-1}\boldsymbol{\Theta}_{l\rho\nu}^{\gamma}) - \log \mid \boldsymbol{\Theta}_l^{\gamma\,-1}\boldsymbol{\Theta}_{l\rho\nu}^{\gamma} \mid - q + p\}, \end{aligned} \tag{23}$$

where $\boldsymbol{\Psi}_l$ is the left-top block in (4.2).

4.4 Model selection methods

The most common approach for comparing and selecting statistical models involves statistics based on the concept of expected KL divergence between the model under examination and the true model. When the true model is elusive, the common strategy is to estimate this discrepancy by considering the sum of two distinct terms. One term is the maximized log-likelihood of the model being tested. This term serves as an indicator of how well the model fits the observed data. The second term, on the other hand, serves a contrasting purpose. It is designed to mitigate

the risk of overfitting, which can occur when a model is too complex and overly tuned to the observed data. This term offers an estimate of the bias that emerges when the true distribution is approximate by the estimated parameters. In essence, this two-term approach strikes a balance between model fit and the risk of overfitting, favouriting a more robust and well-informed model selection process.

Assuming to have N realization of the processes $\boldsymbol{Y}$ and $\boldsymbol{X}$, the log-likelihood of density (22) is

$$\begin{aligned}\boldsymbol{\chi}, \boldsymbol{\gamma} \mid \boldsymbol{\chi}; \quad \boldsymbol{\Psi} \quad , \quad \boldsymbol{\Theta}^{\gamma} \quad &= \frac{N}{2} \sum_{l=1}^{L} {}_{l} \boldsymbol{\chi}; \hat{\boldsymbol{\Psi}}_{l} \quad + \quad {}_{l} \boldsymbol{\gamma}; \boldsymbol{\Theta}_{l}^{\gamma} \\ &= \frac{N}{2} \sum_{l=1}^{L} \log \mid \boldsymbol{\Psi}_{l} \mid -\mathrm{tr} \quad \boldsymbol{S}_{l}^{\chi} \boldsymbol{\Psi}_{l} \quad + \log \mid \boldsymbol{\Theta}_{l}^{\gamma} \mid -\mathrm{tr} \quad \boldsymbol{S}_{\boldsymbol{B}_{l}}^{\gamma} \boldsymbol{\Theta}_{l}^{\gamma} \quad ,\end{aligned}$$

where $\boldsymbol{S}_{l}^{\chi} = N^{-1} \sum_{n=1}^{N} \boldsymbol{\chi}_{ln} \boldsymbol{\chi}_{ln}^{\top}$ and $\boldsymbol{S}_{\boldsymbol{B}_{l}}^{\gamma} = N^{-1} \sum_{n=1}^{N} (\boldsymbol{\gamma}_{ln} - \boldsymbol{\chi}_{ln} \boldsymbol{B}_{l})(\boldsymbol{\gamma}_{ln} - \boldsymbol{\chi}_{ln} \boldsymbol{B}_{l})^{\top}$. The way to estimate the bias term fundamentally characterizes the model selection statistic. Both AIC and BIC penalize the log-likelihood by the degrees of freedom scaled by a parameter. For graphical models, these criteria have the equations

$$\mathrm{AIC}(\mathcal{E}_{\rho\nu}) = -2 \quad \boldsymbol{\chi}, \boldsymbol{\gamma} \mid \boldsymbol{\chi}; \quad \boldsymbol{\Psi} \quad _{\rho\nu}, \quad \boldsymbol{\Theta}^{\gamma} \quad _{\rho\nu} \quad + 2 \mid \mathcal{E}_{\rho\nu} \mid$$

and

$$\mathrm{BIC}(\mathcal{E}_{\rho\nu}) = -2 \quad \boldsymbol{\chi}, \boldsymbol{\gamma} \mid \boldsymbol{\chi}; \quad \boldsymbol{\Psi} \quad _{\rho\nu}, \quad \boldsymbol{\Theta}^{\gamma} \quad _{\rho\nu} \quad + \log (N) \mid \mathcal{E}_{\rho\nu} \mid ,$$

where $\mid \mathcal{E}_{\rho\nu} \mid$ is the number of links in the graph. While BIC is known to be consistent for a fixed number of parameters and increasing sample size, it may not necessarily select a parsimonious model when the model space is large. The extended BIC (eBIC), introduced in Chen and Chen (2008), considers both the number of non-zero estimated parameters and the complexity of the model space. eBIC is particularly useful for variable selection in problems with moderate sample sizes and a large number of variables. It differs from BIC and AIC by an additional penalty term that controls the prior probability of sparse models

$$\mathrm{eBIC}(\mathcal{E}_{\rho\nu}) = -2 \quad \boldsymbol{\chi}, \boldsymbol{\gamma} \mid \boldsymbol{\chi}; \quad \boldsymbol{\Psi} \quad _{\rho\nu}, \quad \boldsymbol{\Theta}^{\gamma} \quad _{\rho\nu} \quad + \log (N) \mid \mathcal{E}_{\rho\nu} \mid + 4g \mid \mathcal{E}_{\rho\nu} \mid \log (p+q),$$

where $p+q$ is the total number of nodes in the graph. The value of g is manually set and larger values of it result in sparser models. When $g = 0$, eBIC is equivalent to the ordinary BIC. The eBIC remains consistent even when the number of parameters grows to infinity with the sample size (Chen & Chen, 2008; Wysocki & Rhemtulla, 2021). This criterion effectively controls the false discovery rate and often shows good performances in terms of graph recovery.

When the goal is to obtain a model with good predicting power, cross-validation is the gold standard. The KLCV leverages the cross-validation of the log-likelihood loss to estimate the KL divergence. This criterion estimates the bias term by approximating leave-one-out-cross-validation and offers a computationally fast alternative to cross-validation. This technique often outperforms other methods, particularly when the sample size is small. The proposed joint KLCV is defined as:

$$\begin{aligned}\mathrm{jKLCV} \quad \boldsymbol{\Psi} \quad _{\rho\nu}, \quad \boldsymbol{\Theta}^{\gamma} \quad _{\rho\nu} \quad &= -\frac{1}{2L} \quad \boldsymbol{\chi}, \boldsymbol{\gamma} \mid \boldsymbol{\chi}; \quad \boldsymbol{\Psi} \quad _{\rho\nu}, \quad \boldsymbol{\Theta}^{\gamma} \quad _{\rho\nu} \\ &\quad + \frac{1}{NL(NL-1)} \sum_{l=1}^{L} \sum_{n=1}^{N} \mathrm{bias}_{l\rho\nu}^{\boldsymbol{\chi}_{n}} + \mathrm{bias}_{l\rho\nu}^{\boldsymbol{\gamma}_{n} \mid \boldsymbol{\chi}}\end{aligned}$$

Adapting the results in Vujačić et al. (2015), for each realization of the process, the bias terms are estimated as,

$$\text{bias}^{\chi_n}_{l\rho v} = \text{vec}\quad \boldsymbol{\Psi}^{-1}_{l\rho v} - S^{\chi_n}_{l\rho v}\ \odot \mathrm{I}^{\chi}_{l\rho v}\ ^{\top} \text{vec}\ \boldsymbol{\Psi}_{l\rho v}\quad S^{\chi}_{l\rho v} - S^{\chi_n}_{l\rho v}\ \odot \mathrm{I}^{\chi}_{l\rho v}\ \boldsymbol{\Psi}_{l\rho v}$$

and

$$\text{bias}^{\gamma_n|\chi}_{l\rho v} = \text{vec}\quad \boldsymbol{\Theta}^{-1}_{l\rho v} - S^{\gamma_n}_{\boldsymbol{B}_{l\rho v}}\ \odot \mathrm{I}^{\gamma}_{l\rho v}\ ^{\top} \text{vec}\ \boldsymbol{\Theta}^{\gamma}_{l\rho v}\quad S^{\gamma}_{\boldsymbol{B}_{l\rho v}} - S^{\gamma_n}_{\boldsymbol{B}_{l\rho v}}\ \odot \mathrm{I}^{\gamma}_{l\rho v}\ \boldsymbol{\Theta}^{\gamma}_{l\rho v}\ ,$$

where $\mathrm{I}^{\chi}_{l\rho v}$ and $\mathrm{I}^{\chi}_{l\rho v}$ are the indicator matrices, whose entry is 1 if the corresponding entry in $\boldsymbol{\Psi}_{l\rho v}$ or $\boldsymbol{\Theta}^{\gamma}_{l\rho v}$ is non-zero, and 0 if the corresponding entry is zero. The observed covariance matrices for the unit n are $S^{\chi_n}_{l} = \boldsymbol{\chi}_{l\,n}\boldsymbol{\chi}^{\top}_{l\,n}$ and $S^{\gamma_n}_{\boldsymbol{B}_{l\rho v}} = (\boldsymbol{\gamma}_{l\,n} - \boldsymbol{\chi}_{l\,n}\boldsymbol{B}_{l\rho v})(\boldsymbol{\gamma}_{l\,n} - \boldsymbol{\chi}_{l\,n}\boldsymbol{B}_{l\rho v})^{\top}$.

In Section 5, the AIC, eBIC, and jKLCV are compared. In general, if the aim is graph identification, then the eBIC is appropriate. On the other hand, if the aim is to find a model with good predictive power, jKLCV and AIC are preferred.

5 Simulation study

In this section, we assess the finite-sample performance of both the proposed fGGRM estimator and the jKLCV goodness-of-fit criterion.

5.1 Impact of sample size on predictive accuracy and graph recovery

We simulate independent realizations of functional responses and covariates under the following setting.

The model consists of $p = 100$ functional responses and $q = 10$ functional covariates, modelled using the multivariate Karhunen–Loève expansion (8) under the assumption of partial separability of the covariance operator. We set $L = 3$ and assume that, for each l, the random score vector $(\boldsymbol{\chi}_l^{\top}, \boldsymbol{\gamma}_l^{\top})^{\top}$ follows a multivariate Gaussian distribution with zero mean and precision matrix as specified in (4.2). In particular, we let Θ^{χ}_l be the identity matrix, while Θ^{γ}_l follows a structured star configuration, where $\theta^{\gamma}_{lr(r+s)}$ is non-zero only for indices $r = 1, 6, 11, \ldots$ and $s = 1, \ldots, 4$ (see Figure 6). The non-zero values are drawn from the uniform distribution $U([-0\,5, -0\,4] \cup [+0\,4, +0\,5])$. Following Zapata et al. (2021), each covariance matrix Σ_l is scaled by a factor of $1\ l^{0\,2}$ to ensure monotonically decreasing traces. The regression coefficient matrices $\boldsymbol{B}_l$ are structured so that each functional response is influenced by exactly two randomly selected functional covariates, with non-zero coefficients drawn from $U([-1\,4, -1\,0] \cup [+1\,0, +1\,4])$. We generate samples of size $N \in \{50, 100\}$.

We compare the proposed fGGRM model against two alternative approaches. The first is the estimator from Zapata et al. (2021), which applies a joint graphical lasso model with a group-lasso penalty to the scores $\boldsymbol{\gamma}_l$ alone. This allows us to assess the impact of omitting functional predictors on precision matrix estimation. The second competitor, termed the 'naive' estimator, fits a conditional graphical lasso separately for each l using $\boldsymbol{\gamma}_l$ and $\boldsymbol{\chi}_l$, and estimates the edge sets $\mathcal{E}^{Y}$ and $\mathcal{E}^{XY}$ using an OR rule, i.e. $(i, j) \in \mathcal{E}^{Y}$ if and only if exists a $\hat{\theta}^{\gamma}_{l\,ij} \neq 0$, with $l = 1, \ldots, L$ and, similarly, $(k, i) \in \mathcal{E}^{XY}$ if and only if exists at least an index l such that $\hat{b}_{l\,ki} \neq 0$.

The considered models were compared both in terms of ability of recovering the edge-sets $\mathcal{E}^{Y}$ and $\mathcal{E}^{XY}$, and the accuracy in estimating the parameters of the true model generating data. In particular, in each simulation run, to evaluate $\{\Theta^{\gamma}\}$ and the corresponding $\mathcal{E}^{Y}$, taking into account the effects coming from $\{\boldsymbol{B}\}$, we used the following strategy. For each method, we computed a decreasing sequence of ten v-values and, for each of these values, a decreasing sequence of ten ρ-values was used to fit the models. The largest v and ρ values were computed as specified in Sottile et al. (2024). The resulting sequences of estimates of the precision matrices were

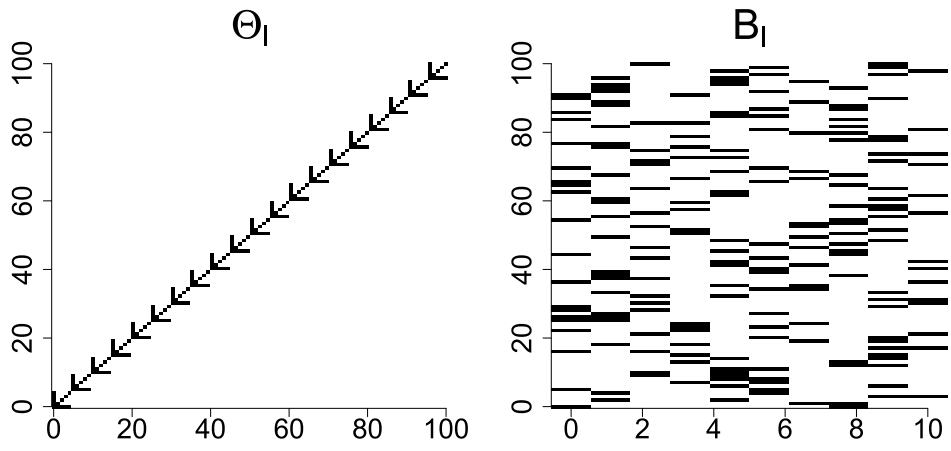

Figure 6. Illustration of the sparsity structure of Θ_l and B_l for $p = 100$, $q = 10$, and $l = 1, \ldots, L$.

summarized using the area under the ROC curve (AUC) and the following estimate of the average mean-square error:

$$\text{aMSE} = \frac{1}{L} \sum_{l=1}^{L} \left\| \Theta_l^{\gamma} - \Theta_l^{\gamma} \right\|_F^2$$

The previous strategy was reversed to evaluate $\{\boldsymbol{B}\}$ taking into account $\{\Theta^{\gamma}\}$, i.e. first a decreasing sequence of ten ρ-values was computed and, for each of these values, the models were fitted using a decreasing sequence of v-values. As before, the behaviour of the resulting estimators was evaluated in terms of AUC and estimate of the average mean-square error.

The results for $N = 50$ are shown in Figure 7. The AUC and aMSE curves consistently demonstrate the superiority of our approach over competitors, reinforcing the benefit of modelling functional responses and predictors jointly with a double penalty. Similar results hold for $N = 100$ (not shown).

5.2 Robustness to violations of partial separability

We further investigate the robustness of the fGGRM estimator under deviations from the partial separability assumption of the covariance operator. We introduce a controlled violation by modifying the covariance structure:

$$\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1} = \begin{pmatrix} \Theta_1 & c \cdot 0\,5(\Theta_1 + \Theta_2) & 0 & 0 \\ c \cdot 0\,5(\Theta_1 + \Theta_2) & \Theta_2 & c \cdot 0\,5(\Theta_2 + \Theta_3) & \vdots \\ \vdots & c \cdot 0\,5(\Theta_2 + \Theta_3) & \ddots & c \cdot 0\,5(\Theta_{L-2} + \Theta_{L-1}) \\ 0 & \cdots & c \cdot 0\,5(\Theta_{L-2} + \Theta_{L-1}) & \Theta_L \end{pmatrix}^{-1}, \tag{24}$$

where c modulates the degree of dependence across levels. When $c \neq 0$, the assumption of conditional independence across l is violated. We consider $c \in \{0\,01, 0\,05, 0\,1\}$, with $c = 0\,1$ representing the highest level of violation that preserves positive definiteness. Across all scenarios, fGGRM remains robust, exhibiting stable predictive accuracy and graph recovery.

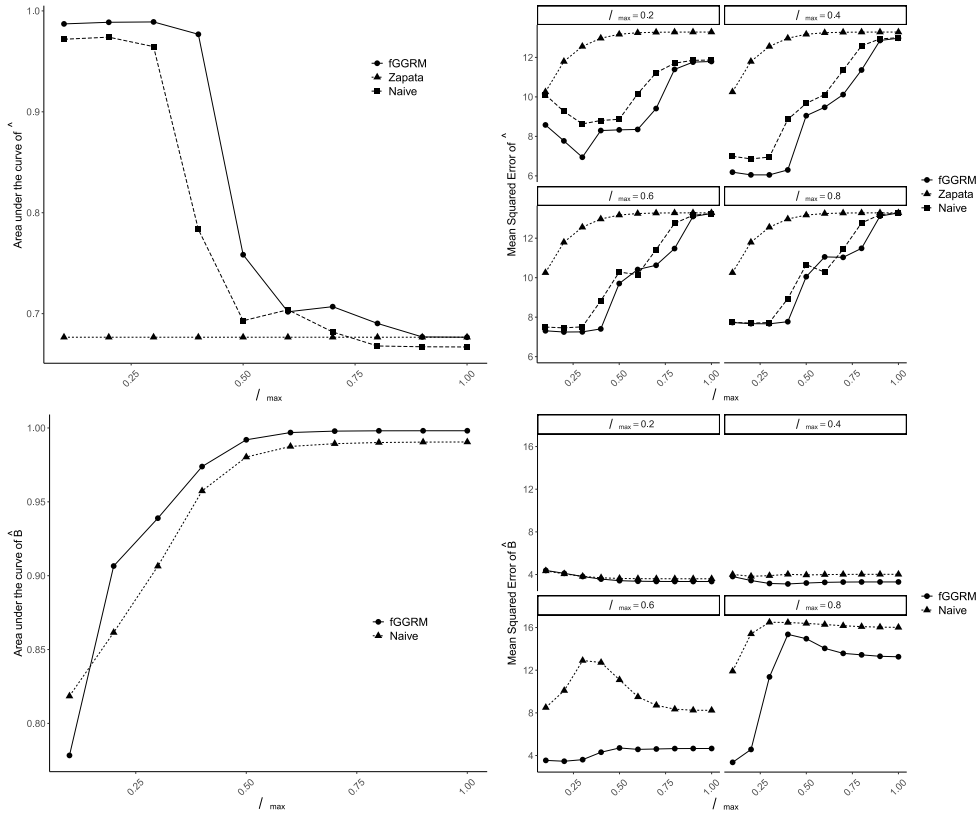

Figure 7. Simulation results for $N = 50$, $p = 100$, and $q = 10$. The upper panels compare fGGRM with the method in Zapata et al. (2021) and a naive estimator. The lower panels omit the method from Zapata et al. (2021) since it does not account for functional predictors. The left panels report AUC curves for the edge sets $\mathcal{E}^{\boldsymbol{Y}}$ and $\mathcal{E}^{\boldsymbol{XY}}$, while the right panels present MSE curves for Θ_l and B_l. AUC values are plotted against the ratios v v_{max} and ρ ρ_{max}, while MSE curves for Θ_l and B_l are computed over ten values of ρ and v, respectively, keeping the regression and precision matrix structures fixed at selected ratios of v v_{max} and ρ ρ_{max}.

Figure 8 presents the results for $N = 50$ and $c = 0$ 1, confirming the previously presented superiority of fGGRM over competing methods. Similar trends are observed for $N = 100$ and lower c values (not shown).

5.3 Comparison of jKLCV with alternative model selection methods

In this section, we assess the effectiveness of different model selection criteria in identifying models that minimize the KL divergence and exhibit strong graphical recovery performance.

5.3.1 jKLCV demonstrates superior predictive accuracy

Table 1 reports the KL loss for models selected using different criteria, across varying sample sizes. The KL loss values in Table 1 (left) are computed using (23) on model estimates selected by the criteria outlined in Section 4.4. When the sample size is small ($N = 10$), jKLCV provides a significantly lower KL loss compared to alternative methods, demonstrating superior predictive accuracy. This advantage stems from its explicit bias correction, which is particularly beneficial in small-sample regimes. As the sample size increases ($N = 100$), the differences among the three criteria diminish, with AIC slightly outperforming the others. This can be attributed to the fact that jKLCV's explicit bias estimation introduces sampling variability, whereas AIC employs a fixed asymptotic correction term, which stabilizes as N grows.

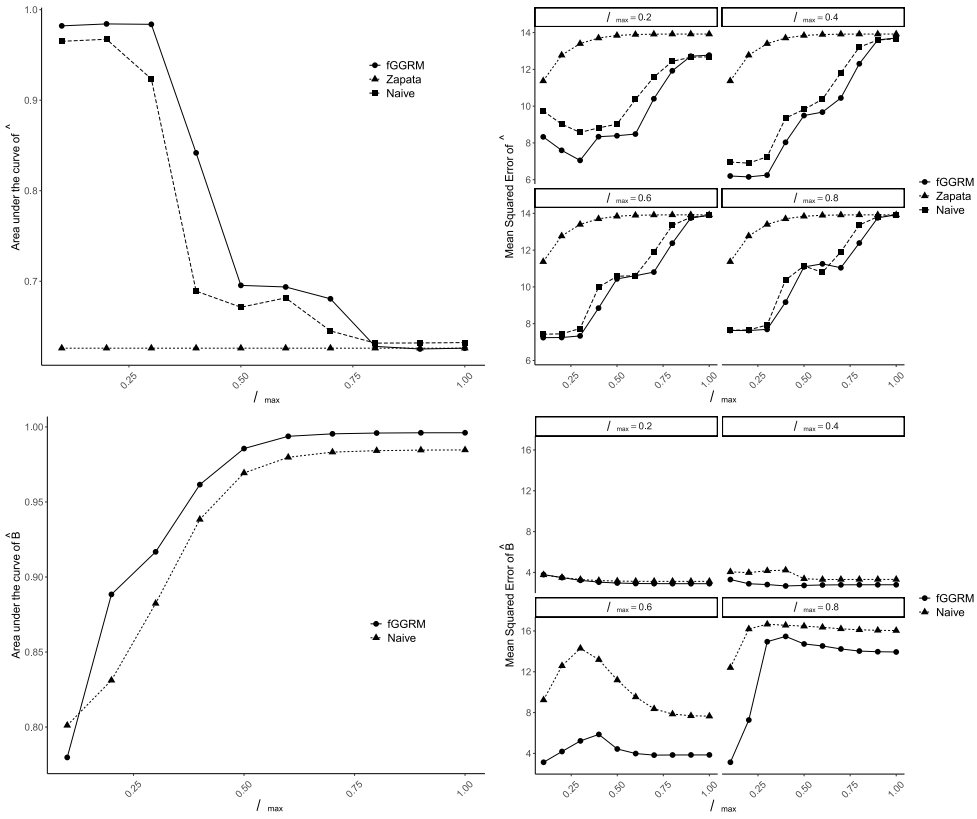

Figure 8. Simulation results for $N = 50$, $p = 100$, $q = 10$, and $c = 0$ 1. The upper panels compare fGGRM with the method in Zapata et al. (2021) and a naive estimator. The lower panels omit the method from Zapata et al. (2021) since it does not account for functional predictors. The left panels report AUC curves for the edge sets $\mathcal{E}^{\boldsymbol{Y}}$ and $\mathcal{E}^{\boldsymbol{XY}}$, while the right panels present MSE curves for Θ_l and B_l. AUC values are plotted against the ratios v $v_{\max}$ and ρ $\rho_{\max}$, while MSE curves for Θ_l and B_l are computed over ten values of ρ and v, respectively, keeping the regression and precision matrix structures fixed at selected ratios of v $v_{\max}$ and ρ $\rho_{\max}$.

Table 1. KL loss and graph recovery accuracy

	KL loss			Graph recovery accuracy		
N	jKLCV	AIC	eBIC	jKLCV	AIC	eBIC
10	**35.55**	54.78	36.20	0.73	0.48	**0.75**
100	15.61	**15.11**	15.76	0.95	**0.96**	0.94

Note. *Left*: KL loss for models selected by different criteria. *Right*: Graph recovery accuracy using different selection methods. Results are the medians over 100 simulations with $L = 3$, across varying sample sizes. For each sample size, the values of the best-performing criteria are indicated in bold.

5.3.2 eBIC excels in graph structure recovery

Table 1 (right) compares the ability of different selection methods to recover the true graphical structure. For small sample sizes ($N = 10$), eBIC exhibits the highest graph recovery accuracy, slightly outperforming jKLCV, and significantly exceeding AIC. This result aligns with eBIC's stronger sparsity enforcement, which is advantageous when estimating graphical structures in high-dimensional settings. For larger samples ($N = 100$), all methods achieve high accuracy in classifying matrix entries correctly, with only marginal differences among them. This suggests that as N increases, the benefits of explicit sparsity control diminish, and all criteria converge toward a similar performance level.

6 Structural interaction of airborne pollutants

This section presents the results of air pollution analysis conducted using the functional Gaussian graphical regression model outlined in this study. The optimization process iteratively determines the edge set by alternately fixing one of the two tuning parameters while varying the other across 101 values, spanning from its maximum to zero. A total of three iterations are performed.

Figure 9 and Table 2 illustrate the selected graphical structures based on the jKLCV, AIC, and eBIC criteria at each iteration.

The AIC criterion yields a fully connected graph for both directed and undirected networks, implying that all pollutants exhibit conditional dependencies when controlling for the remaining variables. Similarly, eBIC selects a complete directed network but a highly sparse undirected network, retaining only the link between CO and O_3. The presence of this connection underscores the crucial role of CO in the formation of tropospheric ozone, as discussed in Choi et al. (2017); Voulgarakis et al. (2011). CO is frequently employed as a proxy for estimating ozone levels generated by anthropogenic sources. However, both fully connected graphs and extremely sparse graphs with a single edge tend to lack interpretability in real-world applications. The jKLCV criterion offers a more balanced network structure. It identifies a directed link from temperature to H_2O and an undirected network where H_2O retains all its connections, along with the crucial link between CO and O_3. This selection highlights the pivotal role of H_2O in pollutant transport (Sherwood et al., 2010). The absence of direct links between NO and both CO and O_3 suggests that their interactions are mediated through other variables. The pronounced effect of temperature on H_2O concentration, as detected by jKLCV, is well-aligned with atmospheric physics. This finding underscores the importance of temperature in regulating humidity levels and emphasizes the significance of water vapour in numerous atmospheric processes and pollutant interactions. Notably, the result that temperature primarily influences H_2O—rather than all pollutants—aligns with the physical reality that water vapour is predominantly controlled by thermal conditions, whereas other pollutants are more directly impacted by emission sources and atmospheric chemical reactions. Previous works, such as Naumann and Kiemle (2020) and Sherwood et al. (2010), have examined the vertical structure of water vapour-temperature interactions within a non-functional framework.

Figure 10 presents the estimated regression function $\beta_{H_2O}(t, t')$ over all pairs $\{(t, t') \mid t, t' = 1, \ldots, 420\}$. The colour scale represents function values, with dark colors indicating the highest positive values and light colors denoting negative values. The prominent dark-colored

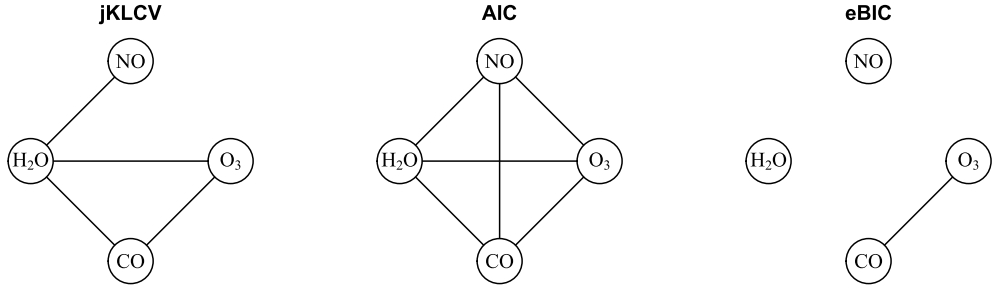

Figure 9. Undirected networks estimated using jKLCV, AIC, and eBIC.

Table 2. Directed links selected using jKLCV, AIC, and eBIC

	O_3	NO	H_2O	CO
AIC	1	1	1	1
eBIC	1	1	1	1
jKLCV	0	0	1	0

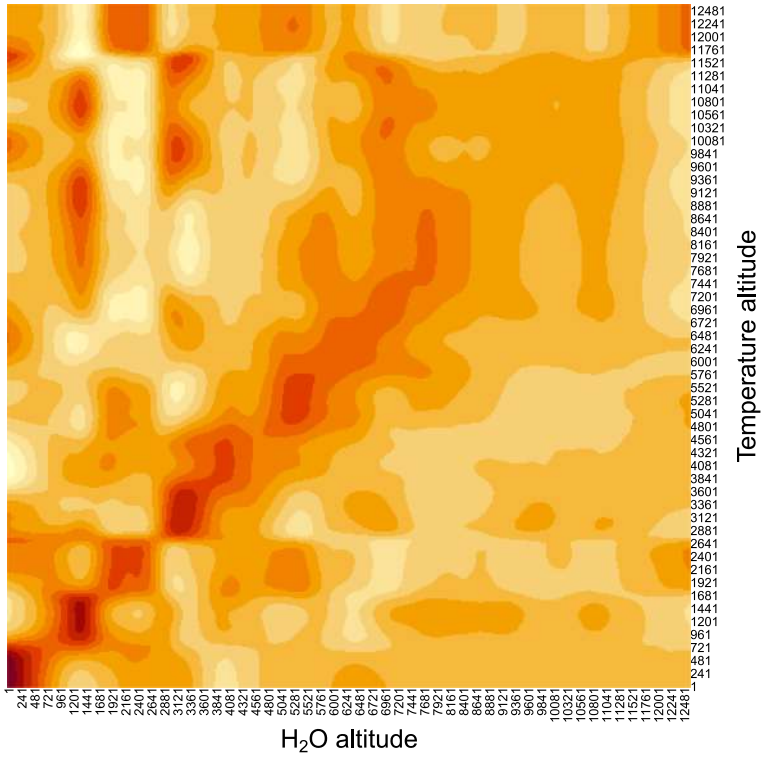

Figure 10. Regression function values $\beta_{H_2O}(t, s)$ for all $t, s = 1, \ldots, 420$. The *x*-axis represents H_2O altitude, while the *y*-axis represents temperature altitude. Darker regions denote the highest regression values (0 00025), while lighter regions indicate the lowest values (–0 00006).

diagonal suggests that temperature exerts the strongest influence on H_2O at the same altitude, extending up to approximately 8,000 m—corresponding to the upper convective boundary, where warm, buoyant air rises and facilitates cloud formation in the presence of water vapour. Convective processes, particularly in the mid-troposphere, are sensitive to humidity levels at various altitudes (Sherwood et al., 2010).

The heatmap reveals distinct vertical patterns in the upper region of Figure 10, suggesting that upper-atmosphere temperature exerts a top-down influence on lower-altitude water vapour levels. The lower triangular region, devoid of dark-colored zones, represents a smoother function, further supporting this top-down effect. This observation aligns with well-established atmospheric mechanisms, wherein solar heating at lower altitudes drives evaporation, leading to water vapour ascent. As the vapour rises, the cooler ambient temperature induces condensation, forming cloud layers—visible as white areas in the heatmap. Above these layers, continued cooling leads to diminished water vapour concentrations, corresponding to positive regression values. At higher altitudes, the colour transitions become more gradual, and the absolute values of $\beta(t, t')$ decrease, indicating a weakening temperature influence on water vapour transport. These findings are consistent with the fundamental thermodynamic and convective processes governing atmospheric moisture dynamics.

7 Conclusion

This article introduces a novel functional regression framework, termed the *functional Gaussian graphical regression model* (fGGRM), which serves as a powerful methodology for uncovering relationships within complex systems of functional variables, such as chemical interactions in the atmosphere. The primary objective of this regression analysis is to reconstruct a graph representation comprising both undirected and directed links.

The undirected links capture the intricate interdependencies among four key atmospheric components: ozone (O_3), nitrogen monoxide (NO), carbon monoxide (CO), and water vapour (H_2O).

These elements exhibit highly complex interactions influenced by various atmospheric conditions. The directed links, on the other hand, represent the dependence of chemical densities on temperature, reflecting its pivotal role in modulating atmospheric composition.

We demonstrate that, under the assumption of partial separability, it is possible to disentangle conditional independence structures within the response functions and establish regression relationships between the covariate and response functions. To estimate the model, we extend the framework of Sottile et al. (2024) to the functional setting, applying it to the scores derived from the expansion of the joint response-explanatory function vector. This refinement leads to the formulation of the *functional joint conditional graphical lasso* estimator.

To refine the model selection process, we introduce the *joint Kullback–Leibler cross-validation* (jKLCV), a tailored criterion designed to determine the optimal configuration for the optimization problem. Comparative analysis of model selection criteria reveals that jKLCV exhibits superior predictive performance, particularly in small-sample settings.

The proposed framework yields promising empirical results. From an atmospheric chemistry perspective, the jKLCV criterion identifies the most plausible model, effectively capturing critical interactions—such as the relationship between O_3 and CO and the distinctive role of H_2O—without introducing spurious connections. In contrast, the AIC-based model fails to provide additional insight beyond the well-known complexity of chemical interactions, while eBIC tends to be overly conservative. Thus, the jKLCV-based model offers a realistic depiction of elemental interactions, emphasizing the conditional nature of certain dependencies and accurately representing the fundamental vertical interplay between temperature and water vapour, a key driver of atmospheric regulation. This approach effectively balances model complexity and interpretability.

A potential extension of the fGGRM framework involves relaxing the assumption of multivariate Gaussianity for the score distributions by incorporating a Gaussian Copula structure, as explored in Solea and Li (2022).

Conflicts of interest: None declared.

Funding

R.F. acknowledges support by PO FSE Sicilia 2014-2020 - POC Sicilia - PJ_AVVISO_6_2022_POC_SICILIA - CUP: G71122000820001. R.F, L.A., and G.S. gratefully acknowledge financial support from the University of Palermo (FFR2024, FFR2025). R.F., L.A., and G.S. were also financially supported by the European Union - Next Generation EU - Mission 4 Component 2 - CUP: B53D23009480006.

Data availability

The data set used in the illustration is from 2020 in L2 format and is available at https://iagos.aeris-data.fr/download/, whereas the latest version of `jcglasso` package can be found at https://github.com/gianluca-sottile/Hematopoiesis-network-inference-from-RT-qPCR-data/tree/main. The R script for the simulation study is available in https://github.com/gianluca-sottile/Functional-Gaussian-Graphical-Regression-Models. The IAGOS data analysis is available in: https://github.com/riifi/Functional-Gaussian-Graphical-Regression-Model.

References

Aston J. A. D., Pigoli D., & Tavakoli S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *Annals of Statistics*, *45*(4), 1431–1461. https://doi.org/10.1214/16-AOS1495

Chen J., & Chen Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Chiou J.-M., Yang Y.-F., & Chen Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, *146*, 301–312. https://doi.org/10.1016/j.jmva.2015.10.003

Choi H.-D., Liu H., Crawford J. H., Considine D. B., Allen D. J., Duncan B. N., Horowitz L. W., Rodriguez J. M., Strahan S. E., Zhang L., Liu X., Damon M. R., & Steenrod S. D. (2017). Global O_3–CO correlations in a chemistry and transport model during July–August: Evaluation with TES satellite observations and sensitivity to input meteorological data and emissions. *Atmospheric Chemistry and Physics*, *17*(13), 8429–8452. https://acp.copernicus.org/articles/17/8429/2017/. https://doi.org/10.5194/acp-17-8429-2017

Constantinou P., Kokoszka P., & Reimherr M. (2018). Testing separability of functional time series. *Journal of Time Series Analysis*, *39*(5), 731–747. https://doi.org/10.1111/jtsa.12302

Dauxois J., & Nkiet G. (1997). Canonical analysis of two Euclidean subspaces and its applications. *Linear Algebra and its Applications*, *264*(11), 355–388. ISSN 0024-3795. https://www.sciencedirect.com/science/article/pii/S0024379596002443. Sixth Special Issue on Linear Algebra and Statistics. https://doi.org/10.1016/S0024-3795(96)00244-3

Delicado P., Giraldo R., Comas C., & Mateu J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, *21*(3-4), 224–239. https://doi.org/10.1002/env.1003

Foygel R., & Drton M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1, NIPS'10* (pp. 604–612). Curran Associates Inc.

Genton M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics: The Official Journal of the International Environmetrics Society*, *18*(7), 681–695. https://doi.org/10.1002/env.854

Gneiting T., Genton M., & Guttorp P. (2006). *Statistical Methods for Spatio-Temporal Systems (Monographs on Statistics and Applied Probability 107)*. https://doi.org/10.1201/9781420011050

Gromenko O., Kokoszka P., Zhu L., & Sojka J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, *6*(2), 669–696. https://doi.org/10.1214/11-AOAS524

Happ C., & Greven S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, *113*(522), 649–659. https://doi.org/10.1080/01621459.2016.1273115

He G., Müller H.-G., & Wang J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, *85*(1), 54–77. https://doi.org/10.1016/S0047-259X(02)00056-8

He G., Müller H.-G., Wang J.-L., & Yang W. (2010). Functional linear regression via canonical analysis. *Bernoulli*, *16*(3), 705–729. https://doi.org/10.3150/09-BEJ228

Ji X., Hu Q., Hu B., Wang S., Liu H., Xing C., Lin H., & Lin J. (2021). Vertical structure of air pollutant transport flux as determined by ground-based remote sensing observations in Fen-Wei plain, China. *Remote Sensing*, *13*(18), 3664. ISSN 2072-4292. https://www.mdpi.com/2072-4292/13/18/3664. https://doi.org/10.3390/rs13183664

Karhunen K. (1946). *Zur spektraltheorie stochastischer prozesse*. Annales Academiae scientiarum Fennicae. Series A. 1 Mathematica-physica. https://books.google.ch/books?id=X3F3SwAACAAJ.

Lee K.-Y., Ji D., Li L., Constable T., & Zhao H. (2023). Conditional functional graphical models. *Journal of the American Statistical Association*, *118*(541), 257–271. https://doi.org/10.1080/01621459.2021.1924178

Lee W., & Liu Y. (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *Journal of Multivariate Analysis*, *111*, 241–255. https://doi.org/10.1016/j.jmva.2012.03.013

Leurgans S. E., Moyeed R. A., & Silverman B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, *55*(3), 725–740. http://www.jstor.org/stable/2345883 https://doi.org/10.1111/j.2517-6161.1993.tb01936.x

Li B., & Solea E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Journal of the American Statistical Association*, *113*(524), 1637–1655. https://doi.org/10.1080/01621459.2017.1356726

Lynch B., & Chen K. (2018). A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika*, *105*(4), 815–831. https://doi.org/10.1093/biomet/asy048

Martínez Torres J., Pastor Perez J., Sancho Val J., McNabola A., Martínez Comesaña M., & Gallagher J. (2020). A functional data analysis approach for the detection of air pollution episodes and outliers: A case study in Dublin, Ireland. *Mathematics*, *8*(2), 225. https://doi.org/10.3390/math8020225

Meinshausen N., & Bühlmann P. (2006). High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, *34*(3), 1436–1462. https://doi.org/10.1214/009053606000000281

Naumann A. K., & Kiemle C. (2020). The vertical structure and spatial variability of lower-tropospheric water vapor and clouds in the trades. *Atmospheric Chemistry and Physics*, *20*(10), 6129–6145. https://acp.copernicus.org/articles/20/6129/2020/https://doi.org/10.5194/acp-20-6129-2020

Penny W. (2001). Kullback-Liebler divergences of normal, gamma, Dirichlet and Wishart densities. http://www.fil.ion.ucl.ac.uk/wpenny/publications/densities.ps.

Qiao X., Guo S., & James G. M. (2019). Functional graphical models. *Journal of the American Statistical Association*, *114*(525), 211–222. https://doi.org/10.1080/01621459.2017.1390466

Qiao X., Qian C., James G. M., & Guo S. (2020). Doubly functional graphical models in high dimensions. *Biometrika*, *107*(2), 415–431. https://doi.org/10.1093/biomet/asz072

Rothman A. J., Levina E., & Zhu J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, *19*(4), 947–962. https://doi.org/10.1198/jcgs.2010.09188

Sherwood S., Roca R., Weckwerth T., & Andronova N. (2010). Tropospheric water vapor, convection, and climate. *Reviews of Geophysics*, *48*(2), RG2001. https://doi.org/10.1029/2009RG000301

Solea E., & Li B. (2022). Copula Gaussian graphical models for functional data. *Journal of the American Statistical Association*, *117*(538), 781–793. https://doi.org/10.1080/01621459.2020.1817750

Sottile G., Augugliaro L., Vinciotti V., Arancio W., & Coronnello C. (2024). Sparse inference of the human haematopoietic system from heterogeneous and partially observed genomic data. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *74*(1), 204–228. https://doi.org/10.1093/jrsssc/qlae052

Uno I., Sugimoto N., Shimizu A., Yumimoto K., Hara Y., & Wang Z. (2014). Record heavy PM2. 5 air pollution over China in January 2013: Vertical and horizontal dimensions. *Sola*, *10*(0), 136–140. https://doi.org/10.2151/sola.2014-028

Voulgarakis A., Telford P. J., Aghedo A. M., Braesicke P., Faluvegi G., Abraham N. L., Bowman K. W., Pyle J. A., & Shindell D. T. (2011). Global multi-year O_3-CO correlation patterns from models and TES satellite observations. *Atmospheric Chemistry and Physics*, *11*(12), 5819–5838. https://acp.copernicus.org/articles/11/5819/2011/. https://doi.org/10.5194/acp-11-5819-2011

Vujačić I., Abbruzzo A., & Wit E. (2015). A computationally fast alternative to cross-validation in penalized Gaussian graphical models. *Journal of Statistical Computation and Simulation*, *85*(18), 3628–3640. https://doi.org/10.1080/00949655.2014.992020

Wysocki A. C., & Rhemtulla M. (2021). On penalty parameter selection for estimating network models. *Multivariate Behavioral Research*, *56*(2), 288–302. https://doi.org/10.1080/00273171.2019.1672516

Yuan M., & Lin Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35. https://doi.org/10.1093/biomet/asm018

Zapata J. A., Oh S.-Y., & Petersen A. (2021). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika*, *109*(3), 665–681. https://doi.org/10.1093/biomet/asab046

Zhao B., Zhai P. S., Wang Y. S., & Kolar M. (2024). High-dimensional functional graphical model structure learning via neighborhood selection approach. *Electronic Journal of Statistics*, *18*(1), 1042–1129. https://doi.org/10.1214/24-EJS2219

Zhou S., Lafferty J., & Wasserman L. (2010). Time varying undirected graphs. *Machine Learning Journal*, *80*(2-3), 295–319. https://doi.org/10.1007/s10994-010-5180-0

Zhu H., Strawn N., & Dunson D. B. (2016). Bayesian graphical models for multivariate functional data. *Journal of Machine Learning Research*, *17*(204), 1–27. http://jmlr.org/papers/v17/16-164.html