



MOX-Report No. 03/2026

**A neighbour selection approach for identifying differential networks
in conditional functional graphical models**

Mapelli, A.; Carini, L.; Ieva, F.; Sommariva, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

A neighbour selection approach for identifying differential networks in conditional functional graphical models

Alessia Mapelli^{*,1,2}, Laura Carini³, Francesca Ieva^{1,2}, and Sara Sommariva⁴

¹ MOX, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano 20133, Italy

² Health Data Science Research Centre, Human Technopole, Viale Rita Levi-Montalcini, 1, Milano, 20157, Italy

³ Life Science Computational Laboratory, IRCCS Ospedale Policlinico San Martino, Genova, Italy

⁴ Dipartimento di Matematica, Università degli Studi di Genova, via Dodecaneso 35, Genova, 16146, Italy

Estimation of brain functional connectivity from EEG data is of great importance both for medical research and diagnosis. It involves quantifying the conditional dependencies among the activity of different brain areas from the time-varying electric field recorded by sensors placed outside the scalp. These dependencies may vary within and across individuals and be influenced by covariates such as age, mental status, or disease severity. Motivated by this problem, we propose a novel neighbour selection approach based on functional-on-functional regression for the characterization of conditional Gaussian functional graphical models. We provide a fully automated, data-driven procedure for inferring conditional dependence structures among observed functional variables. In particular, pairwise interactions are directly identified and allowed to vary as a function of covariates, enabling covariate-specific modulation of connectivity patterns. Our proposed method accommodates an arbitrary number of continuous and discrete covariates. Moreover, unlike existing methods for direct estimation of differential graphical models, the proposed approach yields directly interpretable coefficients, allowing discrimination between covariate-induced increases and decreases in interaction strength. The methodology is evaluated through extensive simulation studies and an application to experimental EEG data. The results demonstrate clear advantages over existing approaches, including higher estimation accuracy and substantially reduced computational cost, especially in high-dimensional settings.

Key words: Function-on-function regression; Gaussian functional graphical models; Graphical models with covariates; Group-based EEG connectivity; High-dimensional data.

1 Introduction

Estimating brain functional connectivity from electroencephalographic (EEG) data consists in quantifying the conditional dependencies between the activity of different brain areas from the time-varying electric field recorded by sensors placed outside the scalp (Sakkalis 2011). A large number of variables, such as age, mental status, severity of neurodegenerative disorders, may influence these dependencies both within and across individuals (Mueller et al. 2013, Nentwich et al. 2020). Just to provide few examples, EEG functional connectivity have been shown to vary with the clinical phenotype of early-stage patients with Dementia with Lewy Bodies or Alzheimer’s disease (Carini et al. 2025, Hasoon et al. 2024). Similarly, few recent studies have demonstrated that differences in EEG functional connectivity can discriminate alcohol and drug abusers from healthy controls (Khajepour et al. 2019, Mumtaz et al. 2018). Hence quantifying differences in functional connectivity across different conditions may allow to identify timely markers of such disorders. Despite these results, the use of EEG functional connectivity in clinical settings is still limited and the development of rigorous statistical techniques for computing robust estimation of the conditional dependency structure within the brain and its variation under different conditions is still

*Corresponding author: e-mail: alessia.mapelli@polimi.it

an open issue (Cao et al. 2022, Sommariva et al. 2025, Vallarino et al. 2020). Here we introduce a novel approach based on functional graphical models.

Graphical models (Lauritzen 1996) have become a popular strategy for inferring conditional dependencies within the element of a (high-dimensional) random vector. They consist in undirected graphs where the nodes are the elements of the random vector and the edges encode pairwise conditional dependencies. When the random vector follows a multivariate Gaussian distribution such edges correspond to the non-zero elements of its precision matrix. Functional graphical models (Qiao et al. 2019) extend the notion of graphical models to graphs whose nodes represent random functions thus allowing for characterizing conditional dependencies within multivariate random functions. In particular, Gaussian functional graphical models deal with multivariate Gaussian processes.

A typical workflow for estimating (Gaussian) functional graphical models consists in: (i) deriving a finite dimensional representation of the observed functions by computing their projection scores with respect to a defined orthonormal basis (ii) estimating a precision matrix from the computed projection scores to infer the conditional dependence graph. Two main families of approaches currently exist for dealing with this second step (Chen 2024). The first one aims at directly estimating the whole precision matrix by optimizing the corresponding likelihood function with a group-lasso-type penalization that promotes sparsity across the entries of the matrix (Qiao et al. 2019, Zapata et al. 2022). The second family of methods consists in neighbour selection schemes where sparse functional regression is employed for separately estimating the neighbours of each node (Zhao et al. 2024). The main difference with the previous family is that these approaches infer the structure of the covariance operator one column at a time without the need of defining the whole precision matrix which may not be well-defined if the random functions lie in a infinite-dimension space or if the corresponding covariance operator is not partial separable (Zapata et al. 2022). Furthermore, since each node can be treated independently, these approaches efficiently scale to high dimensional settings involving a large number of nodes, and prior information can be exploited for example for limiting the analysis to a subset of nodes of interest as commonly done in seed-based functional connectivity studies.

Most of the existing methods in both families assume that all the observed functions follow the same graphical models without accounting for the effects of external covariates. Only few recent works have focused on characterizing differences in the conditional dependence structures of different populations. When comparing the graphical models associated to two vector-valued samples, the differential graph is commonly characterized by studying the support of the difference between the corresponding precision matrices (Zhao et al. 2014, Xu & Gu 2016, Liu et al. 2014). Zhao et al. (2022) extended this idea to the functional setting and proposed an algorithm called FuDGE (Functional Differential Graph Estimation) for directly estimating the functional differential graph without explicitly computing the functional graphical model of the two populations. However, such a method only accounts for the presence of one binary external variable. Moysidis & Li (2021) proposed a method handling discrete covariate by assuming that all the subgroups in the population share a common graph structure with group-specific level of sparsity. Hierarchical penalization is then exploited to infer the subgroups precision matrices of the coefficients from the Karhunen-Loeve expansion of the functions in input. For handling both continuous and discrete external covariate, Lee et al. (2023) introduce a novel linear operator, the functional precision operator, that extends the notion of precision matrix and its estimation to the functional setting. However, also this approach is based on operator analysis which can be computationally demanding and statistically inefficient in high-dimensional functional settings.

Here we propose a neighbour selection strategy that extends the approach proposed by Zhao et al. (2024) to the estimation of conditional functional graphical models whose structure may vary with an arbitrary number of continuous and discrete external covariates. In detail, for each node we formulate the identification of its neighbours as a function-on-function (fof) regression problem modelling the interaction between the response variables, in our case the EEG time-series, and the covariates. The fof regression model is then approximated by projecting the observed functions in a node-specific basis determined via FPCA. The coefficient matrices of the resulting vector-on-vector (vov) regression model are estimated through a

group-lasso-penalized least squares approach. The result is a fully automated, data-driven approach that, given a set of observed functions, characterizes their conditional dependence structure by directly identifying the pairwise interactions modulated by each covariate. We demonstrate the benefit of our proposed approach in the framework of functional EEG connectivity by comparing its performance with that of FUDGE and three related variants on both simulated data and experimental EEG recordings from a public dataset concerning alcohol use disorder (Ingber 1997, Zhang et al. 1995). The implemented approach is available at <https://github.com/AlessiaMapelli/condFGM>.

The paper is organized as follows. In Section 2 we described the proposed fof-regression based neighbour selection approach. In Section 3 we provide detail on the implementing procedure for estimating the coefficient matrices of the resulting approximated vov regression models. Section 4 and Section 5 collect the results from the analysis of the simulated and the experimental EEG data, respectively. Our conclusions are offered in Section 6.

2 Statistical model

In this section we first introduce the notion of conditional functional graphical model, then we define the fof regression model associated to each node of the graphical model and we derive the corresponding vov approximation.

2.1 Conditional functional graphical model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{H} \subseteq \mathcal{L}^2(\mathcal{T})$ be a Hilbert subspace of the space of square-integrable real-valued functions $\mathcal{L}^2(\mathcal{T})$, where $\mathcal{T} \subseteq \mathbb{R}$ is a closed interval.

Given $p > 0$, we consider a Multivariate Gaussian Process (MGP) $\mathbf{Y} : \Omega \mapsto \mathcal{H}^p$ that we assume to be zero-mean. For all $\omega \in \Omega$ and $t \in \mathcal{T}$ we denote $\mathbf{Y}(\omega)(t) = (Y_1(\omega, t), \dots, Y_p(\omega, t))$, where $Y_j(\omega, \cdot) \in \mathcal{H}$ for all $j = 1, \dots, p$. Throughout the paper, we will omit the dependence on ω and/or t whenever this does not lead to ambiguity. In our application p is the number of EEG sensors and Y_j represent the signal at the j -th sensor.

For each pair (Y_j, Y_l) , $j, l = 1, \dots, p$, the conditional cross-covariance function is defined as (Qiao et al. 2019)

$$C_{jl}(t', t) = \text{Cov}(Y_j(t'), Y_l(t) | \mathbf{Y}_{-\{j,l\}})$$

where $\mathbf{Y}_{-\{j,l\}} = \{Y_k : k \in \{1, \dots, p\} \setminus \{j, l\}\}$ and $(t', t) \in \mathcal{T}^2$. We say that Y_j and Y_l are conditionally independent, and we write $(Y_j(t) \perp\!\!\!\perp Y_l(t) | \mathbf{Y}_{-\{j,l\}})$, if and only if $C_{jl}(t', t) = 0$ for all $(t', t) \in \mathcal{T}^2$. Defined the set of vertices $V = \{1, \dots, p\}$, we aimed at characterizing the undirected graph $G = (V, E)$ where the set of edges

$$E = \{(j, l) \in V^2 : Y_j(t) \not\perp\!\!\!\perp Y_l(t) | \mathbf{Y}_{-\{j,l\}}\} \quad (1)$$

collects all the pairwise conditional interdependence within $\mathbf{Y}$.

In many practical scenarios, the graph G is affected by external variables such as, for example, the age of the patients or their disease stage. Inspired by Lee et al. (2023), to account for this influence we used a conditional functional graphical model. In detail, given a q -dimensional random vector of covariates $\mathbf{X} = (X_1, \dots, X_q)$, for all $\mathbf{x} \in \mathbb{R}^q$ we assume that $\mathbf{Y} | \mathbf{X} = \mathbf{x}$ is a MGP following a graphical model defined by the graph $G(\mathbf{x}) = (V, E(\mathbf{x}))$ being

$$E(\mathbf{x}) = \{(j, l) \in V^2 : Y_j(t) \not\perp\!\!\!\perp Y_l(t) | [\mathbf{Y}_{-\{j,l\}}, \mathbf{X} = \mathbf{x}]\} . \quad (2)$$

2.2 Function-on-function regression

Given the functional graphical model defined by the graph G in Eq. (1), under mild assumptions (Zhao et al. 2024), for each node $j \in V$ there exist $\{\beta_{j,k}(t, t')\}_{j \neq k}$ such that $\|\beta_{j,k}\|_{HS} < \infty$ and

$$Y_j(t) = \sum_{k \neq j} \int_{\mathcal{T}} \beta_{j,k}(t, t') Y_k(t') dt' + e_j(t) \quad (3)$$

being $e_j(\cdot) \perp Y_k(\cdot)$ for all $k \neq j$, and the set of neighbours of node j in graph G is given by

$$\mathcal{N}_j := \{k \in V \setminus \{j\} : (j, k) \in E\} = \{k \in V \setminus \{j\} : \|\beta_{j,k}\|_{HS} > 0\}.$$

To account for the influence of the external covariates, we assume that the regression coefficients in (3) depend linearly on the component of $\mathbf{X}$. Hence for all $c = 0, \dots, q$, there exists $\{\beta_{j,k}^c(t, t')\}_{j \neq k}$ such that

$$\beta_{j,k}(t, t') = \beta_{j,k}^0(t, t') + \sum_{c=1}^q \beta_{j,k}^c(t, t') X_c = \sum_{c=0}^q \beta_{j,k}^c(t, t') X_c \quad (4)$$

where we set $X_0 = 1$. By incorporating (4) into the model in (3) we obtain

$$\begin{aligned} Y_j(t) &= \sum_{k \neq j} \int_{\mathcal{T}} \beta_{j,k}^0(t, t') Y_k(t') dt' + \sum_{c=1}^q \sum_{k \neq j} \int_{\mathcal{T}} \beta_{j,k}^c(t, t') X_c Y_k(t') dt' + e_j(t) \\ &= \sum_{c=0}^q \sum_{k \neq j} \int_{\mathcal{T}} \beta_{j,k}^c(t, t') X_c Y_k(t') dt' + e_j(t) \end{aligned} \quad (5)$$

The complete set of neighbours of node j according to the conditional graph $G(\mathbf{x})$ in Eq. (2) will be given by

$$\mathcal{N}_j(\mathbf{x}) := \{k \in V \setminus \{j\} : (j, k) \in E(\mathbf{x})\} = \left\{ k \in V \setminus \{j\} : \left\| \sum_{c=0}^q \beta_{j,k}^c X_c \right\|_{HS} > 0 \right\}.$$

Our interest is in characterizing how each covariate affects the conditional dependence structure of $\mathbf{Y}$. In model (5), the baseline coefficients $\beta_{j,k}^0(t, t')$ encode the population-level pairwise conditional interdependencies, while $\beta_{j,k}^c(t, t')$ captures how the c -th covariate modulates such relationships. Importantly, throughout we consider a fixed set of vertices $V = \{1, \dots, p\}$, and our goal is to identify, for each covariate X_c , which pair-wise connections $(j, k) \in E(X_c)$ are affected by it. This explicit decomposition into a population network and covariate-specific modulation components and their simultaneous estimation in a unified model is the key novelty of our approach.

To this end, for each $c = 0, \dots, q$ we define the covariate-specific neighbour set

$$\mathcal{N}_j^c = \{k \in V \setminus \{j\} : \|\beta_{j,k}^c\|_{HS} > 0\}, \quad (6)$$

and the associated undirected graph $G^c = (V, E^c)$ with

$$E^c = \{(j, \ell) \in V^2 : \ell \in \mathcal{N}_j^c \text{ AND/OR } j \in \mathcal{N}_\ell^c\}.$$

Here, G^0 represents the baseline (population-level) network, whereas for $c \geq 1$ the graph G^c encodes a differential network collecting the pairs of nodes for which the conditional association is influenced by the c -th covariate X_c .

For any covariate value $x \in \mathbb{R}^q$, the conditional model $G(\mathbf{x})$ in Eq. (2) is obtained by combining the baseline and covariate components through the linear decomposition $\beta_{j,k}(t, t'; x) = \beta_{j,k}^0(t, t') +$

$\sum_{c=1}^q x_c \beta_{jk}^c(t, t')$. Thus, an edge may be present due to the baseline term, due to a covariate term, or due to their combination, while the graphs $\{G^c\}_{c=1}^q$ isolate which edges are differential with respect to each specific covariate.

Suppose we observe n functions $\{\mathbf{y}_i(\cdot) := (y_{i,1}(\cdot), \dots, y_{i,p}(\cdot))\}_{i=1}^n$ and the corresponding covariates $\{\mathbf{x}_i := (x_{i,1}, \dots, x_{i,q})\}_{i=1}^n$ which are realization of n i.i.d. random copies of $\mathbf{Y}$ and $\mathbf{X}$, respectively. In this paper we present a neighbour selection procedure for estimating both the population network $G^0 = (V, E^0)$ and the differential networks $G^c = (V, E^c)$, $c = 1, \dots, q$. Hence, for each node $j = 1, \dots, p$ we seek to estimate the sets of nodes $\mathcal{N}_j^c$, $c = 0, \dots, q$ defined by (6).

Remark 2.1 The fof model in Eq. (5) easily accounts for both continuous and categorical covariates. In particular, a categorical covariates with ℓ categories can be incorporated in the model by fixing a reference category and by defining $\ell - 1$ binary variables such that

$$x_{i,c} = \begin{cases} 1 & \text{if observation } i \text{ belongs to category } c \\ 0 & \text{otherwise} \end{cases} \quad c = 1, \dots, \ell - 1.$$

For each observation in the reference group it then holds

$$y_{i,j}(t) = \sum_{k \neq j} \int_{\mathcal{T}} \beta_{j,k}^0(t, t') y_{i,k}(t') dt' + e_{i,j}(t),$$

while for each observation in the c -th group, $c = 1, \dots, \ell - 1$, we can write

$$y_{i,j}(t) = \sum_{k \neq j} \int_{\mathcal{T}} (\beta_{j,k}^0(t, t') + \beta_{j,k}^c(t, t')) y_{i,k}(t') dt' + e_{i,j}(t).$$

In this case, the population network $G^0 = (V, E^0)$ collect the conditional dependence relationships within the reference group, while the differential network $G^c = (V, E^c)$ represents the pairwise relationship in which the c -th group differs from the reference one.

2.3 Vector-on-vector regression

The fof regression model in (5) can be approximated in a tractable finite dimensional vector-on-vector (vov) linear regression problem as follows. For each node $j \in V$, let $\{\phi_m^{(j)}\}_{m=1}^\infty$ by an orthonormal basis of $\mathcal{H}$, and let $M > 0$ be a positive constant. The actual choice of the basis, which is defined independently for each node, is discussed in Appendix. Each observed function is approximated with the corresponding projection onto the first M elements of the basis, namely

$$y_{i,k} = \sum_{m=1}^{\infty} a_{i,km}^{(j)} \phi_m^{(j)} \tag{7}$$

$$\simeq \sum_{m=1}^M a_{i,km}^{(j)} \phi_m^{(j)} \tag{8}$$

where, for all $k = 1, \dots, p$ and $i = 1, \dots, n$, $a_{i,km}^{(j)} = \langle y_{i,k}, \phi_m^{(j)} \rangle$ are the projection scores.

Since the neighbour selection is typically carried on independently for each target node, for the ease of notation in the following we will focus on the last node, i.e. $j = p$, and we will omit the dependence on the index j of the projection scores and of the elements of the basis.

By exploiting Eq. (7) and the fof model in (5), it follows that for each observation $i = 1, \dots, n$ the following representation holds:

$$\mathbf{a}_{i,p} = \sum_{c=0}^q \sum_{k=1}^{p-1} \mathbf{B}_k^c x_{i,c} \mathbf{a}_{i,k} + \boldsymbol{\rho}_i + \boldsymbol{\varepsilon}_i. \quad (9)$$

In the previous model, $\mathbf{a}_{i,k} = (a_{i,k1}, \dots, a_{i,kM})^\top$ collects the first M projection scores for the k -th node; $\mathbf{B}_k^c = (b_{k,mm'}^c)_{1 \leq m, m' \leq M} \in \mathbb{R}^{M \times M}$ are regression coefficient matrices whose relationship with the parameters of the original fof regression model is given by

$$b_{k,mm'}^c = \int_{\mathcal{T}} \int_{\mathcal{T}} \beta_{p,k}^c(t, t') \phi_m(t) \phi_{m'}(t') dt dt' \quad \forall m, m' \geq 1$$

$\boldsymbol{\rho}_i = (\rho_{i,1}, \dots, \rho_{i,M})$, being $\rho_{i,m} = \sum_{c=0}^q \sum_{k=1}^{p-1} \sum_{m=M+1}^{\infty} b_{k,mm'}^c a_{i,km'} x_{i,c}$, is a bias term arising when the approximation in (8) is employed; and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,M})$, being $\varepsilon_{i,m} = \int_{\mathcal{T}} e_{i,p}(t) \phi_m(t) dt$, is a vector of model errors related to the error term $e_{i,p}(\cdot)$ in (5).

According to the vov model in (9), if M is large enough, the set of neighbours of node p in the differential graph G^c associated to the c -th covariate can be approximated as

$$\mathcal{N}_p^c \simeq \{k \in V \setminus \{p\} : \|\mathbf{B}_k^c\|_F > 0\}.$$

3 Estimation procedure

In this Section, details about the estimation procedure for the vov regression model presented in (9) will be given together with some practical details for the actual implementation of the proposed approach.

3.1 Penalized least square estimation of the regression coefficient and of the differential networks

Inspired by Zhao et al. (2024) we compute a penalized least square estimator of the coefficient matrices within the vov regression model in Eq. (9). Specifically, given n i.i.d samples $\{(\mathbf{y}_i(\cdot), \mathbf{x}_i)\}_{i=1}^n$ we determine

$$\begin{aligned} & \hat{\mathbf{B}}_1^0, \dots, \hat{\mathbf{B}}_{p-1}^0, \hat{\mathbf{B}}_1^1, \dots, \hat{\mathbf{B}}_{p-1}^1, \dots, \hat{\mathbf{B}}_1^q, \dots, \hat{\mathbf{B}}_{p-1}^q \in \\ & \underset{\mathbf{B}_1^0, \dots, \mathbf{B}_{p-1}^0, \dots, \mathbf{B}_1^q, \dots, \mathbf{B}_{p-1}^q \in \mathbb{R}^{M \times M}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^n \left\| \mathbf{a}_{i,p} - \sum_{c=0}^q \sum_{k=1}^{p-1} \mathbf{B}_k^c x_{i,c} \mathbf{a}_{i,k} \right\|_2^2 + \lambda_p \sum_{c=0}^q \sum_{k=1}^{p-1} \|\mathbf{B}_k^c\|_F \right\} \end{aligned} \quad (10)$$

where λ_p is a regularization parameter whose value may depend on the node whose neighbours' set is being computed.

Eq. (10) can be reformulate in a more compact way as

$$\hat{\mathbf{B}} \in \underset{\mathbf{B} \in \mathbb{R}^{M \times M(p-1)(q+1)}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{A}_p - (\mathcal{A} * \mathcal{X}) \mathbf{B}\|_F^2 + \lambda_p \sum_{k=1}^{(p-1)(q+1)} \|\mathbf{B}_k\|_F \right\}. \quad (11)$$

In equation (11), $\mathbf{B} = [\mathbf{B}_1^0, \dots, \mathbf{B}_{p-1}^0, \dots, \mathbf{B}_1^q, \dots, \mathbf{B}_{p-1}^q] \in \mathbb{R}^{M \times M(p-1)(q+1)}$ and $\mathbf{B}_k$ scans the $M \times M$ coefficient matrices within $\mathbf{B}$; $\mathcal{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{p-1}] \in \mathbb{R}^{n \times M(p-1)}$, where for all $j = 1, \dots, p$,

$$\mathbf{A}_j = \begin{bmatrix} \mathbf{a}_{1,j}^\top \\ \vdots \\ \mathbf{a}_{n,j}^\top \end{bmatrix} \in \mathbb{R}^{n \times M};$$

$\mathcal{A} * \mathcal{X} \in \mathbb{R}^{n \times M(p-1)(q+1)}$ denotes the matrix obtained by multiplying element-wise each column of $\mathcal{A}$ by each columns of the matrix

$$\mathcal{X} = \begin{bmatrix} 1 & \mathbf{x}_1 \\ \vdots & \vdots \\ 1 & \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{n \times (q+1)}.$$

The optimization problem in (11) can be further rewritten as

$$\begin{aligned} \underset{\mathcal{P}, \mathcal{Q} \in \mathbb{R}^{M \times M(p-1)(q+1)}}{\operatorname{argmin}} \quad & \left\{ \frac{1}{2n} \|\mathbf{A}_p - (\mathcal{A} * \mathcal{X})\mathcal{Q}\|_F^2 + \lambda_p \sum_{k=1}^{(p-1)(q+1)} \|\mathcal{P}_k\|_F \right\} \\ \text{subject to } & \mathcal{P} = \mathcal{Q} \end{aligned} \quad (12)$$

which can be solved via the Alternating Direction Method of Multipliers (ADMM, Boyd et al. (2011)).

Once the estimates $\hat{\mathcal{B}} = [\hat{\mathbf{B}}_1^0, \dots, \hat{\mathbf{B}}_{p-1}^0, \dots, \hat{\mathbf{B}}_1^q, \dots, \hat{\mathbf{B}}_{p-1}^q]$ of the regression coefficients have been computed, the estimated set of neighbours of node p within the differential graph G^c is given by

$$\hat{\mathcal{N}}_p^c = \left\{ k \in V \setminus \{p\} : \|\hat{\mathbf{B}}_k^c\|_F > \varepsilon_p \right\}, \quad (13)$$

where $\varepsilon_p > 0$ is a threshold parameter to be fixed.

The same procedure is carried out independently for each target node $j \in [p]$, yielding node-wise neighbourhood estimates $\hat{\mathcal{N}}_j^c$, $c = 0, \dots, q$. Following the neighbourhood selection literature (Meinshausen & Bühlmann 2006), we recover an undirected edge set via OR- or AND-symmetrization:

$$\hat{E}^{c, \text{OR}} = \left\{ (j, l) \in V^2 : l \in \hat{\mathcal{N}}_j^c \text{ OR } j \in \hat{\mathcal{N}}_l^c \right\}, \quad (14)$$

or, for a more conservative choice,

$$\hat{E}^{c, \text{AND}} = \left\{ (j, l) \in V^2 : l \in \hat{\mathcal{N}}_j^c \text{ AND } j \in \hat{\mathcal{N}}_l^c \right\}. \quad (15)$$

Characterizing the direction of change. The edge sets in (14)–(15) identify pairs of nodes whose conditional association varies with the c -th covariate. In contrast to approaches that only recover the support of a differential network, our formulation also allows us to quantify the direction of change.

Once $\hat{\mathcal{N}}_p^c$ is obtained from (13), the effect of the c -th covariate on the conditional association between node p and another node k can be further specified by comparing the baseline block $\hat{\mathbf{B}}_k^0$ with its covariate-specific counterpart $\hat{\mathbf{B}}_k^c$. For each $k \in \hat{\mathcal{N}}_p^c$, we define the relative effect

$$Eff_{kp}^c = \frac{\|\hat{\mathbf{B}}_k^0 + \hat{\mathbf{B}}_k^c\|_F}{\|\hat{\mathbf{B}}_k^0\|_F}. \quad (16)$$

Values $Eff_{kp}^c < 1$ indicate that the c -th covariate attenuates the baseline association between nodes p and k , whereas $Eff_{kp}^c > 1$ indicates a strengthening. For visualization and interpretability, one may alternatively report $\log(Eff_{kp}^c)$, which is symmetric around zero (negative values correspond to reductions and positive values to increases).

These directed effects can be summarized for each differential edge (k, j) in G^c into a single weight as follows:

$$w_{kj}^c = \begin{cases} Eff_{kj}^c & \text{if } k \in \mathcal{N}_j^c \text{ and } j \notin \mathcal{N}_k^c \\ Eff_{jk}^c & \text{if } k \notin \mathcal{N}_j^c \text{ and } j \in \mathcal{N}_k^c \\ \sqrt{Eff_{kj}^c \cdot Eff_{jk}^c} & \text{if } k \in \mathcal{N}_j^c \text{ and } j \in \mathcal{N}_k^c \end{cases}. \quad (17)$$

3.2 Pipeline overview

A schematic overview of the implemented pipeline is provided in Algorithm 1. In particular, we observe that in many practical scenarios, including the estimation of EEG functional connectivity, the functions $\{y_i\}_{i=1}^n$ are only observed on a discrete set of time-points $\{t_\ell\}_{\ell=1}^{L_i}$. Therefore, a smoothing step is required before applying the proposed approach. Details on how to perform this step are provided in Appendix A.2 where we also discuss the choice of the basis for the finite-dimensional embedding, providing practical guidance for selecting the truncation level M . Details on the hyperparameter selection procedure can be found in Appendix A.3: the group-lasso neighbourhood regressions are solved via an ADMM routine, and the node-specific tuning parameters $(\lambda_j^*, \epsilon_j^*)$ are selected automatically using selective cross-validation (SCV) with a BIC-type penalty, with an optional randomized search to reduce computational cost.

Algorithm 1 Neighbour selection approach for estimating population and differential functional graphical models.

Input: For each sample $i = 1, \dots, n$ the discrete measurements $\{y_{i,j}^{\text{obs}}(t_{i\ell})\}$ for each node $j \in V$ and the covariates $\mathbf{x}_i \in \mathbb{R}^q$.

- 1: **Smoothing.** For each sample i and node j , smooth the discrete observations to obtain a curve $\tilde{y}_{i,j} \in \mathcal{H}$.
- 2: **Node-wise neighbourhood estimation.** For each $j \in V$ do:

- (a) **Choice of the projection basis.** Define an orthonormal basis $\{\phi_m^{(j)}\}_{m \geq 1}$ (fixed or FPCA on $\{\tilde{y}_{i,j}\}_{i=1}^n$ based). Retain the first M elements.
- (b) **Dimentionality reduction.** For each node $k \in V$ and sample i , compute score vectors $\mathbf{a}_{i,k}^{(j)} = (\langle \tilde{y}_{i,k}, \phi_1^{(j)} \rangle, \dots, \langle \tilde{y}_{i,k}, \phi_M^{(j)} \rangle)^\top$.
- (c) **Penalized vov regression.** Estimate the coefficient matrices $\{\hat{\mathbf{B}}_k^c\}$ for $k \in V \setminus \{j\}$ and $c = 1, \dots, q$ by solving the group-lasso penalized least square problem described by (12).
- (d) **Automatic hyperparameter tuning and thresholding.** Select $(\lambda_j^*, \epsilon_j^*)$ via SCV (see Appendix A.3) and for $c = 0, \dots, q$ form the set of neighbours $\hat{\mathcal{N}}_j^c = \{k \neq j : \|\hat{\mathbf{B}}_k^c\|_F > \epsilon_j^*\}$.
- (e) **Inference of the direction of change** For each $k \in \hat{\mathcal{N}}_j^c$, compute the relative effect Eff_{kj}^c as in (16).

- 3: **Graph construction.** For each $c = 0, \dots, q$, aggregate $\{\hat{\mathcal{N}}_j^c\}_{j=1}^p$ and obtain an undirected edge set $\hat{E}^c$ via OR- or AND-symmetrization (cf. (14)–(15)), yielding $\hat{G}_c = (V, \hat{E}^c)$.
- 4: **Edge weighting.** For each $(k, j) \in \hat{E}^c$, define the weight w_{kj}^c as in (17) to discriminate attenuated vs strengthened dependences.

Output: Estimated graphs $\{\hat{G}_c\}_{c=0}^q$ and edge weights $\{w_{kj}^c\}_{c=1}^q$.

4 Simulation study

We tested the finite sample performance of the proposed approach by using simulated data. In the following, dataset simulation and model's performances are presented.

4.1 Data generation

To be able to quantitatively compare our results with those of existing method directly estimating the differential network between two groups (Zhao et al. 2022), we model the presence of a single binary

external variable representing a grouping factor that splits the sampled data into two groups: $\mathcal{G}^0$, taken as reference group, and $\mathcal{G}^1$. Hence in our simulation $q = 1$ and $x_{i,1} = 0$ for all the observations i that belong to the reference group $\mathcal{G}^0$, while $x_{i,1} = 1$ for all the other observations within $\mathcal{G}^1$.

We assume the functional data are observed at $\tau = 100$ time points equally spaced in $(0, 1]$. For all $i = 1, \dots, n$ and all $j = 1, \dots, p$, the observed values are generated as

$$y_{i,j}^{obs}(t_k) = \alpha_{i,j}^\top \mathbf{f}(t_k) + \eta_{i,j}(t_k) \quad k = 1, \dots, \tau$$

where $\mathbf{f}(t_k) = (f_1(t_k), \dots, f_{M^*}(t_k))^\top$ collects the value of the first M^* Fourier basis functions (Ramsay & Silverman 2005b) at the k -th observed time points; in our simulations we set $M^* = 15$; $\eta_{i,j}(t_k)$ is drawn from a zero-mean Gaussian distribution with variance $\sigma^2 = 0.5$; $\alpha_{i,j}^\top = (\alpha_{i,j1}, \dots, \alpha_{i,jM^*}) \in \mathbb{R}^{M^*}$ are scoring factors such that for each observation, $i = 1, \dots, n$, the vector $(\alpha_{i,1}^\top, \dots, \alpha_{i,p}^\top)^\top \in \mathbb{R}^{pM^*}$ is randomly drawn from a multivariate Gaussian distribution with zero mean and covariance matrix $\Sigma_i = (\Theta^0 + x_{i,1}\Theta^1)^{-1}$. Here, Θ^0 and Θ^1 are symmetric matrices of size $pM^* \times pM^*$ encoding respectively the population network G^0 , that in our simulations describes the conditional interdependencies within data from the reference group, and the differential graph G^1 , collecting pairs of nodes whose conditional dependencies is different in the two observed groups.

The precision matrices Θ^0 and $\Theta^0 + \Theta^1$ are defined by modifying the block-banded matrix $\Theta \in \mathbb{R}^{pM^* \times pM^*}$ whose non-zero blocks are given as

$$\Theta_{kj} = \begin{cases} \mathbf{T} & \text{if } k = j \\ 0.4\mathbf{A} & \text{if } k \in \{j-1, j+1\} \\ 0.2\mathbf{I}_{M^*} & \text{if } k \in \{j-2, j+2\} \end{cases}$$

where $\mathbf{T} \in \mathbb{R}^{M^* \times M^*}$ is a Toeplitz matrix such that $T_{vv} = 1$ and $T_{vw} = 0.5|v - w|$ for all $v \neq w$; $\mathbf{A} \in \mathbb{R}^{M^* \times M^*}$ is a tridiagonal matrix such that $A_{vv} = 1$ and $A_{v(v+1)} = A_{(v+1)v} = 0.5$; $\mathbf{I}_{M^*} \in \mathbb{R}^{M^* \times M^*}$ is the identity matrix. In detail, we initially set both the precision matrices equal to Θ and then we modified them in order to simulate the following scenarios, denoted as S1-S6 and schematically depicted in Fig. 1.

- S1. The samples within $\mathcal{G}^0$ show fewer pairwise conditionally connections than those in $\mathcal{G}^1$. Towards this end we set $\Theta_{kj}^0 = \mathbf{0}$ for all $j, k \in \{1, \dots, \lfloor \frac{p}{3} \rfloor\}$.
- S2. The samples with $\mathcal{G}^0$ show some additional connections with respect to those in $\mathcal{G}^1$. To simulate this condition we set $(\Theta^0 + \Theta^1)_{kj} = \mathbf{0}$ for all $k, j \in \{1, \dots, \lfloor \frac{p}{3} \rfloor\}$.
- S3. $\mathcal{G}^0$ and $\mathcal{G}^1$ involve two different but intersecting set of edges. This condition is simulated by setting $\Theta_{kj}^0 = \mathbf{0}$ for all $j, k \in \{1, \dots, \lfloor \frac{p}{3} \rfloor\}$ and $(\Theta^0 + \Theta^1)_{kj} = \mathbf{0}$ for all $k, j \in \{p - \lfloor \frac{p}{3} \rfloor + 1, \dots, p\}$.
- S4. $\mathcal{G}^0$ and $\mathcal{G}^1$ involve two non-intersecting set of edges. In this case we $\Theta_{kj}^0 = \mathbf{0}$ for all $j, k \in \{1, \dots, \lfloor \frac{p}{2} \rfloor + 1\}$ and $(\Theta^0 + \Theta^1)_{kj} = \mathbf{0}$ for all $k, j \in \{p - \lfloor \frac{p}{2} \rfloor, \dots, p\}$.
- S5. $\mathcal{G}^0$ and $\mathcal{G}^1$ show the same pairwise conditionally pairwise relationship, but some of the connections are stronger in $\mathcal{G}^1$. To simulate this scenario for all $j, k \in \{1, \dots, \lfloor \frac{p}{3} \rfloor\}$, $j \neq k$, we set $\Theta_{kj}^0 = 0.25\Theta_{kj}$.
- S6. Similarly to the previous scenario, $\mathcal{G}^0$ and $\mathcal{G}^1$ show the same pairwise conditionally pairwise relationship, but in $\mathcal{G}^1$ some of the connections are weaker. To simulate this scenario we set $(\Theta^0 + \Theta^1)_{kj} = 0.25\Theta_{kj}$ for all $j, k \in \{1, \dots, \lfloor \frac{p}{3} \rfloor\}$, $j \neq k$.

4.2 Benchmark method and evaluation metrics

In simulation studies we compare the proposed estimator with FuDGE and three joint graphical-lasso baselines (FGL, FFGL, FFGL2) introduced by Zhao et al. (2024). All benchmark methods operate on a finite-dimensional representation of the functional observations (e.g., FPCA scores), and performance is assessed in terms of recovery of the differential graph G^1 .

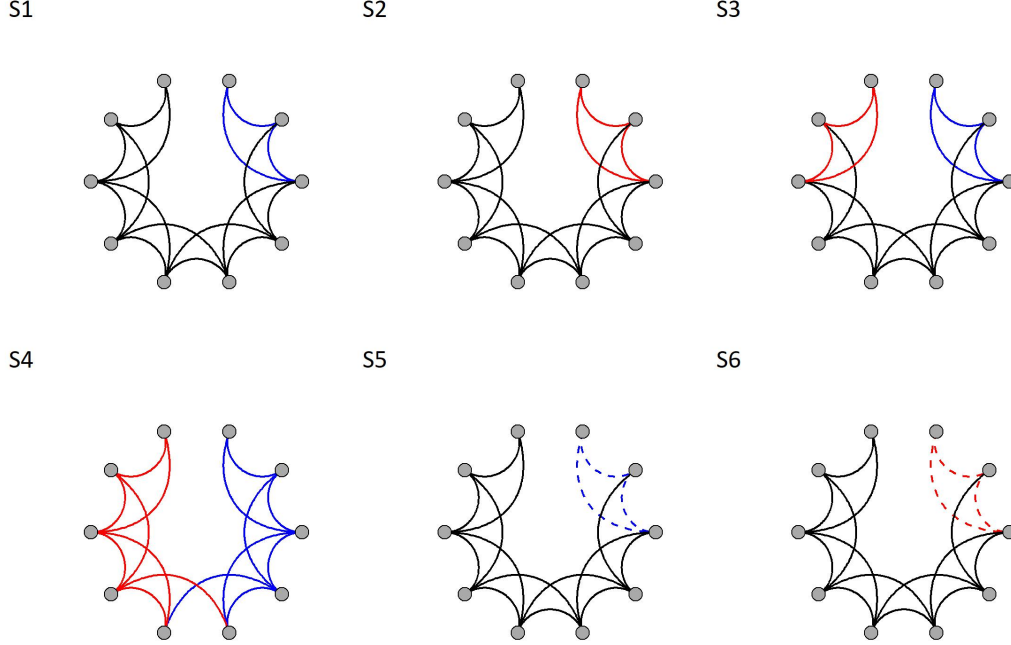


Figure 1 Visual illustration of the graphical models involved in the simulated scenarios S1-S6 described in detail in Section 4.1. In all graphs: black solid lines depict pairwise conditional interdependencies that are common to the two groups; red and blue solid lines depict pairwise conditional connections that are present only in one group, the reference group and the other group, respectively; red and blue dashed lines represent connections that are common to the two groups but that are stronger in the reference or in the other group, respectively.

Benchmark methods. *FuDGE* (Functional Differential Graph Estimation) is a direct estimator of the functional differential graph. Rather than estimating the two group-specific graphs and subtracting them, *FuDGE* targets the difference of the (finite-dimensional) precision operators, Θ^1 , through a convex quadratic loss constructed from the sample covariance matrices of the score vectors, combined with a penalty that promotes sparsity of Θ^1 at the block (edge) level. As a result, *FuDGE* outputs only the differential network, which could be appealing when the two underlying graphs may be dense but their difference is sparse.

FGL (Fused Graphical Lasso) is a joint estimation approach originally developed for multiple related Gaussian graphical models. It estimates the precision matrices of the two groups simultaneously using an ℓ_1 penalty to induce sparsity within each group and an additional fused ℓ_1 penalty on entry-wise differences to encourage similarity of edge values across groups. In the functional setting, *FGL* is applied directly to the (vectorized) score representation and does not explicitly exploit the natural block structure induced by the basis expansion.

FFGL and *FFGL2* are functional extensions of *FGL* that incorporate block structure. Both methods jointly estimate Θ^0 and $\Theta^0 + \Theta^1$ with a block-wise sparsity penalty (Frobenius norms of off-diagonal blocks). They differ in how similarity across groups is enforced: *FFGL* uses a fused Frobenius penalty on differences between corresponding blocks, whereas *FFGL2* retains an element-wise fused ℓ_1 penalty within blocks. Differential edges are then obtained by differencing the two estimated precision matrices.

Relation to the proposed method. Unlike FuDGE, which estimates only Θ^1 and FGL and FFGL that focus their estimation only on the components Θ^1 and $\Theta^0 + \Theta^1$, our approach is formulated to estimate simultaneously the group component Θ^0 and the differential graph Θ^1 within a single optimisation problem. This joint parameterisation allows us to recover both group-specific networks $\mathcal{G}^0$ described by Θ^0 and $\mathcal{G}^1$ given by

$$\hat{\mathcal{N}}_j^{\mathcal{G}^1} = \left\{ k \in V \setminus \{j\} : \|\hat{\mathbf{B}}_k^0 + \hat{\mathbf{B}}_k^1\|_F > \varepsilon_j \right\}$$

while still enabling a fair comparison on the differential-network recovery task used in the simulations.

Evaluation metrics. Across simulations, we evaluate recovery of the target differential adjacency structure using precision, true positive rate (TPR), false positive rate (FPR) and the F_1 score on the undirected graph obtained after symmetrization. Specifically, we report Precision = $\text{TP}/(\text{TP} + \text{FP})$, TPR = $\text{TP}/(\text{TP} + \text{FN})$, FPR = $\text{FP}/(\text{FP} + \text{TN})$, and $F_1 = 2 \text{TP}/(2 \text{TP} + \text{FP} + \text{FN})$. Metrics are computed on the upper-triangular part of the symmetric adjacency matrices. Results are primarily summarized in terms of F_1 , as it provides a single measure balancing precision and recall.

For FuDGE and the related baselines, the public simulation code release does not include an automatic, data-driven hyperparameter selection routine. Therefore, for these methods we computed the metrics over the full hyperparameter sequences provided by the respective implementations and, for each simulation setting, reported the best-performing configuration (highest F_1) in Figure 3. This choice yields an optimistic estimate of benchmark performance relative to our approach, for which tuning parameters are selected in a fully data-driven manner.

4.3 Results

To test the impact of sample size and number of nodes on our proposed method, we focused on scenario S3 and we generated 10 synthetic EEG data for each combination of $p \in \{10, 15, 25, 50\}$ and $n \in 2 \cdot \{50, 75, 100, 150, 200\}$. The two groups, $\mathcal{G}^0$ and $\mathcal{G}^1$, are of equal sample size.

Fig. 2 shows that the OR symmetrization strategy in Eq. (14) should be preferred over the AND strategy in Eq. (15): while the two strategies provide comparable results for the population network $\mathcal{G}^0$, for the differential network $\mathcal{G}^1$ the F1 score of the OR strategy is systematically higher. For this reason, in the remaining of the section, only results for this strategy are shown. Furthermore, Fig. 2 shows that the estimates of the population network are usually more accurate than the estimates of the differential network. However, with enough samples our approach provides accurate estimate of both networks: when the group sample size, $n/2$, is greater than 150 ($n/2 \geq 200$ for $p = 50$) the average F1 score exceeds 0.75 for the differential network and approaches 0.9 for the population network.

As a touchstone, Fig. 3 compares the proposed approach with four competitors. Among the latter, the best performance were reached by FuDGE whose F1 scores associated to the differential network $\mathcal{G}^1$ remained lower than 0.65 also for the highest tested sample sizes. Our approach outperforms FuDGE also from a computational point of view, as shown by Fig. 4. While FuDGE deals with the entire precision matrices of the involved processes, our approach is based on a neighbour selection approach that can be fully parallelized across nodes. As a result, our approach shows an higher scalability: while FuDGE and our method show similar computational times for a small number of nodes ($p = 10$), when $p = 50$ the computation time of FuDGE is more the four time higher than that of our approach.

As a further experiment, we tested the robustness of our results when varying the underlying graphical structure. Towards this end, we fixed $p \in \{10, 50\}$ and the group sample size to $n/2 = 200$ and we generated 10 simulated EEG data for each scenario depicted in Fig. 1. Our approach reached satisfying results in all the considered scenarios. In scenario S1, S3, S4, and S5, the averaged F1-score for the differential network was close to 1 for $p = 10$ and to 0.8 for $p = 50$. Slightly worse results were reached in scenario S2 and S6 where the differential network only includes pairwise interdependencies that are stronger in the reference group $\mathcal{G}^0$. Interesting, in these scenarios the support of the graphical models

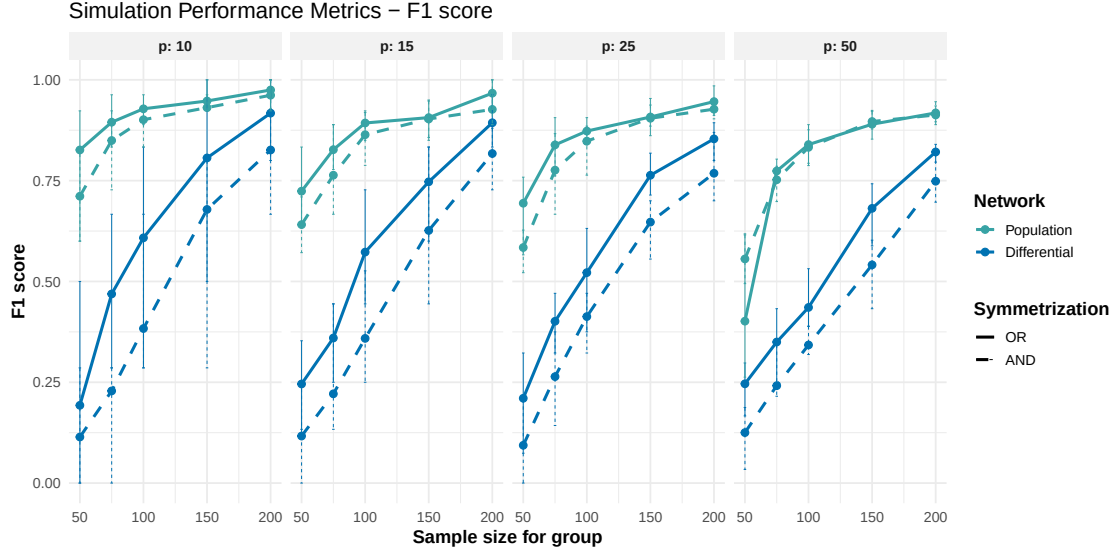


Figure 2 Performance in estimating the population network G^0 and the differential graph G^1 for different numbers of nodes p , sample sizes n , and symmetrization strategies. For each combination of p and n the plots show minimum, mean and maximum F1 score reached over 10 simulated data generated according to scenario S3.

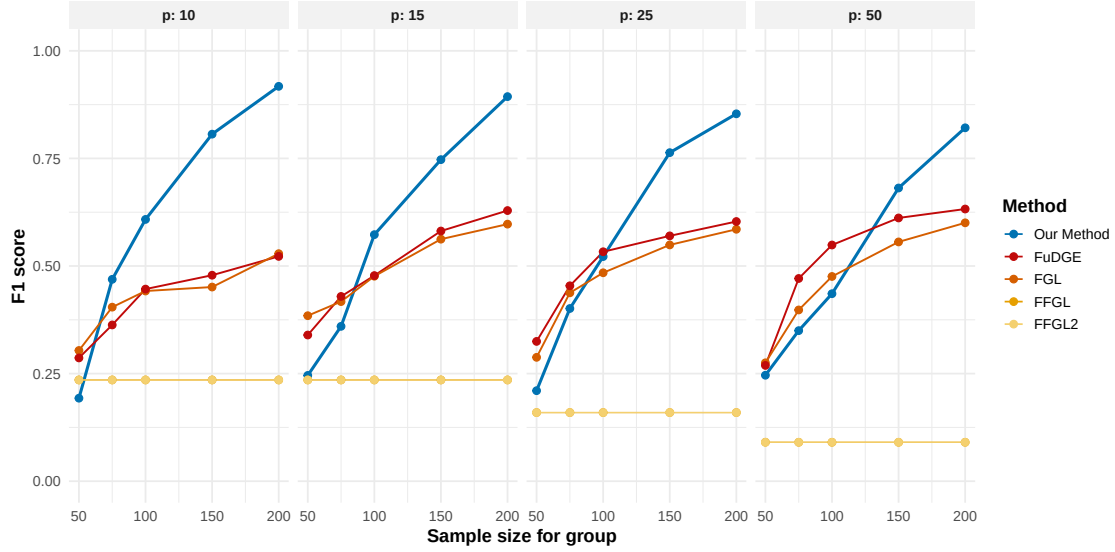


Figure 3 Comparison of the performance in estimating the differential network G^1 of our approach with respect to state-of-the-art methods. Plots show the average F1 score over the 10 simulated data simulated considered in Fig. 2. Note that the F1-scores for FFGL coincide with those of FFGL2.

of the two groups $\mathcal{G}^0$ and $\mathcal{G}^1$ is correctly identified, suggesting that this behaviour may be partially due to penalization term of the loss function in (10) which forces sparsity in the estimates of the differential network.

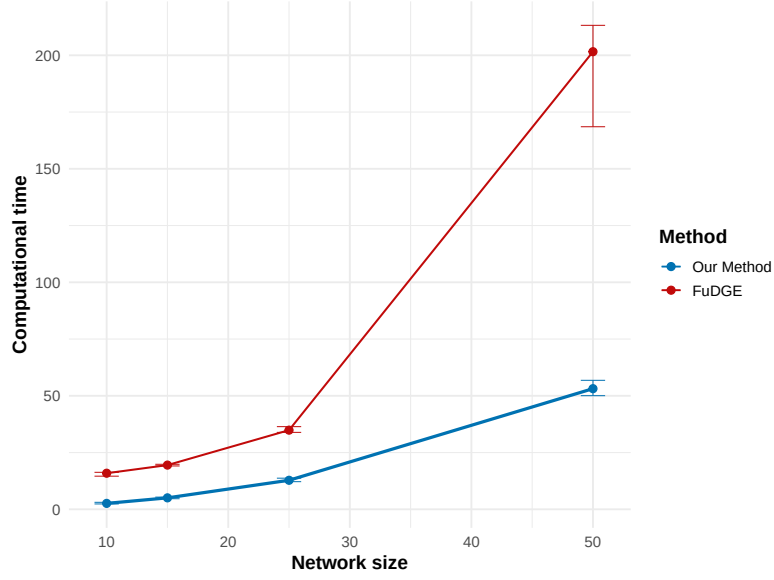


Figure 4 Comparison of the computational time of the proposed approach and of FuDGE. For each combination of number of nodes and methods, considering a sample size of 200 observations per group, the plots show the average computational time in seconds across 10 different data.

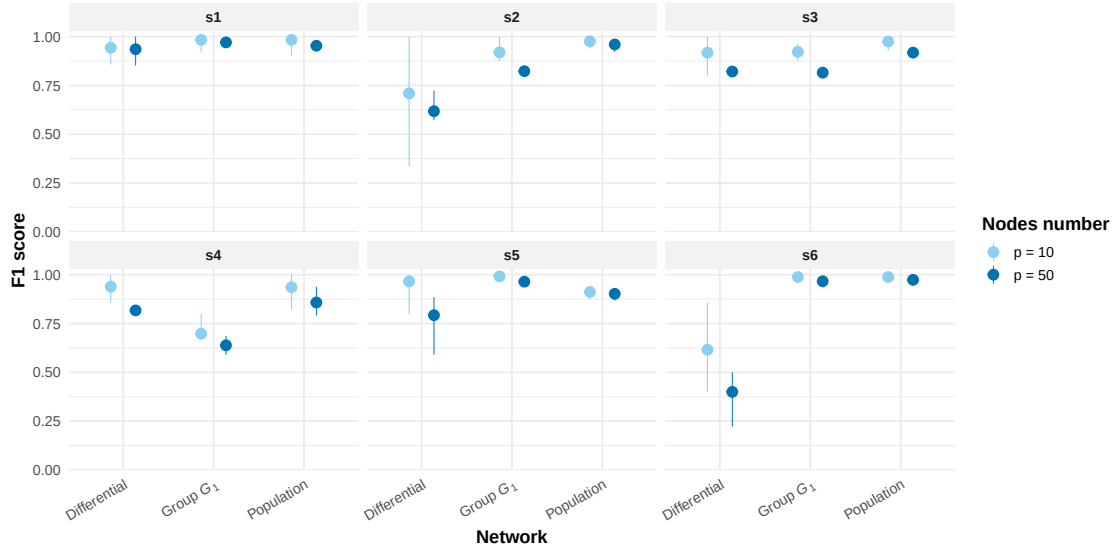


Figure 5 Performance in estimating the differential network G^1 and the graphical models associated to the groups $\mathcal{G}^0$ (population network) and $\mathcal{G}^1$, under the scenarios in Fig. 1 when the number of nodes is $p \in \{10, 50\}$ and the sample size of each group is $n/2 = 200$. For each scenario, plots show minimum, mean and maximum F1 score reached over 10 simulated data.

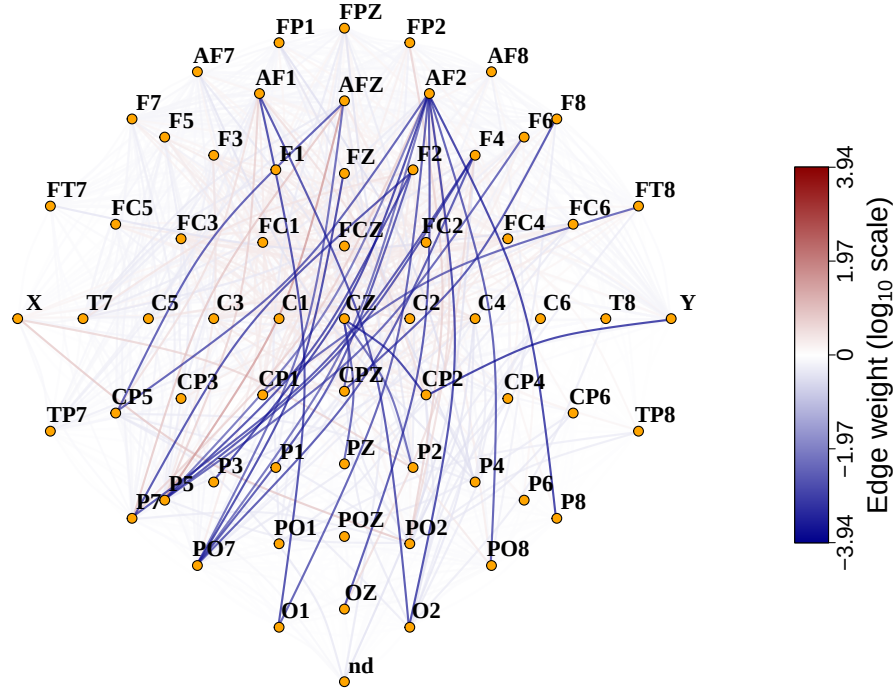


Figure 6 Differential network characterizing changes in EEG functional connectivity in individual with AUD with respect to controls. Edges are coloured based on the weights defined in Eq.(17) so that blue and red edges indicated a reduction and an increase in the connection strength, respectively.

5 Case study: effect of alcohol abuse on EEG functional connectivity

We apply our method to a publicly available EEG dataset (Zhang et al. 1995, Ingber 1997) involving $n = 122$ participants, including 77 individuals diagnosed with alcohol use disorder (AUD) and 45 control subjects. EEG recordings were collected during an object recognition task using a standard 61-sensor EEG cap, one ground sensor placed on the nose, and two bipolar derivations recording the vertical and horizontal electrooculogram (EOG). To obtain results comparable with those presented by Zhao et al. (2022) we thus fixed $p = 64$ and we relied on their preprocessed EEG data, where signals were filtered within the α -frequency band (8–12.5 Hz) (Knyazev 2007, Zhu et al. 2014).

We applied our proposed method equipped with the OR symmetrization strategy, and considering as external variable a binary covariate that distinguishes the controls, chosen as the reference group $\mathcal{G}^0$, from the group, $\mathcal{G}^1$, of individuals with AUD.

Appendix B.1 summarizes the output provided by our approach while Fig. 6 focuses on the estimated differential network describing changes in EEG functional connectivity induced by AUD. In this figure, the direction of the changes are inferred by plotting the decimal logarithm of the weights in Eq. (17).

Our results mainly indicated a decreased functional connectivity between the frontal areas and the posterior-occipital regions. This result is consistent with previous studies reporting impaired connectivity in individuals with alcohol dependence (de Bruin et al. 2005, Herrera-Díaz et al. 2016), although most previous works focused on a different recording condition where subjects are resting. Furthermore, the regions involved in our differential network are similar to those identified by Zhao et al. (2022) using FUDGE, with the main difference being that most regions in their network were connected to node X. It is worth noting that channel X was not included in the analysis of the original study by Zhang et al. (1995),

as it is an unconventional channel for the standard EEG cap used in their work and may have served as a sensor for controlling EOG artifacts rather than for measuring brain activity. More importantly, while we recognize a reduction in the connection strength, FUDGE does not provide any information on the direction of the change. Thus overall, our method provide richer results more adequate to answer to clinical and research questions.

6 Discussion

Motivated by the problem of estimating group-specific EEG functional connectivity, we proposed a neighbourhood selection procedure based on functional-on-functional regression for conditional Gaussian functional graphical models with an arbitrary number of continuous and discrete covariates. Given EEG recordings from n individuals, the resulting pipeline returns (i) a population-level graph G^0 capturing conditional dependencies shared across subjects, and (ii) a covariate-specific modulation graph G^c for each covariate, highlighting the connections whose strength varies with X_c .

The proposed framework contributes to the literature in two main ways. First, it extends neighbourhood selection from unconditional functional graphical models to a conditional setting, providing a unified and practical pipeline to estimate, within a single model, both the baseline population network and the covariate-specific differential components. This joint decomposition enables two complementary types of inference: it allows one to reconstruct group-specific connectivity patterns by combining baseline and covariate terms, while simultaneously isolating differential networks to support clinical insights. Second, beyond identifying which pairwise connections are affected by each covariate, our approach provides a principled way to characterize their direction of change. By leveraging the estimated regression coefficients, we can summarize whether the covariate induces an attenuation or a strengthening of the baseline conditional association, which is not directly available in most methods focusing solely on differential edge support.

Our simulation studies and real-data application support the practical advantages of the method. From a methodological standpoint, we introduced a computationally efficient estimator based on a block-structured group-lasso objective solved by an ADMM routine, combined with a data-driven SCV procedure for node-specific hyperparameter selection, including both the regularization level controlling sparsity in the regression coefficients and the threshold used to identify non-negligible edges. Empirically, across a broad set of simulation scenarios, the method achieved competitive or improved edge-recovery performance relative to existing approaches, with particularly strong performance as the number of subjects increased. In addition, the computational cost scaled favourably in high-dimensional settings, including values of p comparable to commonly used EEG montages, making the approach more suitable for real-world applications where the number of sensors can be large. The method remained robust across the considered data-generating settings, and the real-data analysis produced connectivity patterns that were coherent with prior neurophysiological findings reported in the literature, supporting the reliability of the inferred networks.

A key limitation emerges in the low-sample regime, where reconstruction of differential networks is more challenging. This behaviour is expected given the high dimensionality of the parameter space induced by the functional representation and by the presence of multiple covariate-specific coefficient blocks. Moreover, the current penalization is primarily block-wise, encouraging sparsity at the edge level but not within-edge structure. As a consequence, when sample sizes are limited, the estimator may either retain or discard an entire edge block, even when only a subset of within-block coefficients is reliably supported by the data.

Several directions are promising for improving performance in such settings. A natural extension is to incorporate additional penalties that act within blocks (e.g., sparse-group penalties), enabling the model to reduce the effective number of parameters without forcing complete removal of an edge. A second direction concerns scalability in the number of covariates. While the current framework accommodates an arbitrary set of covariates, inference may become sample-limited as q grows. To address this, future work

will consider introducing an additional layer of regularization at the covariate level (e.g., a penalty on entire covariate-specific coefficient groups), yielding a data-driven mechanism for identifying influential covariates while shrinking negligible covariate effects towards zero. This extension would support applications in which many clinical or behavioural variables are available, while maintaining stable network estimation under realistic sample sizes.

Acknowledgements The present research has been partially supported by MUR, grant Dipartimento di Eccellenza 2023-2027 awarded to the Department of Mathematics of Politecnico di Milano (A.M. and F.I.) and to the Department of Mathematics of The Università di Genova (L.C. and S.S.).

S.S. acknowledges the support of the PRIN PNRR 2022 Project 'Computational mEthods for Medical Imaging (CEMI)' 2022FHCNY3, cup: D53D23005830006.

Data Availability

The EEG data that support the findings of this study are openly available in the GitHub repository associated to the work by Zhao et al. (2022) available at <https://github.com/boxinz17/FuDGE/tree/master/EEG/>.

Code Availability All code required to reproduce the analyses and figures in this manuscript is available at: <https://github.com/AlessiaMapelli/condFGM>.

Conflict of Interest

The authors have declared no conflict of interest.

Appendix

A.1 Algorithm details

The developed method builds on Zhao et al. (2024) and extends their ADMM routine for solving the optimization problem in (12) to accommodate a user-specified number of covariates. The resulting implementation is designed to scale from sequential execution to parallel runs in high-performance computing (HPC) environments. The pipeline is composed of three steps:

1. Sparse and discretely observed trajectories are smoothed into continuous functional curves, for example via spline smoothing, yielding subject-specific longitudinal profiles.
2. Given a chosen basis for $\mathcal{H}$, each curve is projected onto the basis and represented through its corresponding score vector.
3. The main pipeline is executed: for each node $j \in V$, a neighbourhood selection problem is fitted to estimate the conditional associations with the remaining nodes.
4. Neighbourhood estimates are aggregated across j and symmetrized as in (14) – (15), producing the estimated conditional graphs $\hat{G}_c$, for $c = 0, \dots, q$.

A.2 Choice of the basis

Both the smoothing step and the finite-dimensional approximation in (7) require specifying a basis for the Hilbert space $\mathcal{H}$.

In practice, the EEG signals are observed on finite grids and are affected by measurement error. For $i = 1, \dots, n, j \in V$, and observation times $\{t_{i\ell}\}_{\ell=1}^{L_i} \subset T$, we model the recorded values as

$$y_{i,j}^{\text{obs}}(t_{i\ell}) = \tilde{y}_{i,j}(t_{i\ell}) + \eta_{i,j\ell},$$

where $\tilde{y}_{i,j} \in \mathcal{H}$ denotes the underlying smooth trajectory and $\{\eta_{i,j\ell}\}$ are i.i.d. noise terms. A standard approach is to approximate $\tilde{y}_{i,j}$ through a basis expansion, $\tilde{y}_{i,j}(t) \approx \sum_{r=1}^R \theta_{i,jr} \psi_r(t)$, and to estimate the coefficients $\theta_{i,jr}$ via (penalized) least squares, typically including a roughness penalty such as the integrated squared second derivative to encourage smoothness (Ramsay & Silverman (2005a)). This first-stage choice is driven by the qualitative features of the generating process: a Fourier basis is well suited for approximately periodic signals without localized bursts, a condition often encountered in resting-state EEG, whereas B-splines provide greater local adaptivity through their piecewise-polynomial construction and can better accommodate non-stationary patterns. The number of basis functions R to include in this first step is a key tuning parameter, as it controls the flexibility of the reconstructed curves: larger R allows the fit to capture increasingly complex temporal features, but may also lead to overfitting by reproducing measurement noise rather than the underlying signal. Prior knowledge should guide this choice. Within our applications, both in simulations and in the real-data analysis, we used a Fourier basis for the reconstruction step, as it naturally captures the oscillatory nature of EEG signals. We fixed the number of basis functions to $R = 15$, which we found sufficiently large to provide adequate flexibility for representing complex trajectories while maintaining smoothness.

After smoothing, we construct the finite-dimensional representation used in the vector-on-vector model (9) by projecting each curve onto an orthonormal basis of $\mathcal{H}$. In particular, for each neighbourhood selection problem associated with a target node j , we adopt a basis $\{\phi_m^{(j)}\}_{m \geq 1}$ and compute the projection scores $a_{i,km}^{(j)} = \langle y_{i,k}, \phi_m^{(j)} \rangle$, retaining the first M components. Two main strategies are commonly employed for defining $\{\phi_m^{(j)}\}$: (i) a fixed, pre-specified basis (e.g., Fourier, B-spline, wavelet), chosen based on prior knowledge and computational convenience; (ii) a data-driven basis, most often functional principal components, obtained as the eigenfunctions of an estimated covariance operator. The FPCA basis yields the best L_2 approximation among all M -dimensional subspaces and therefore provides an efficient compression when M is moderate. As noted by Zhao et al. (2024), using an estimated basis in place of the population eigenfunctions induces an additional error term in the regression model, reflecting uncertainty due to basis estimation.

Despite this additional source of variability, in the context of neighbourhood selection we favour data-driven bases over fixed choices, as a misspecified pre-selected basis may inadequately capture the dominant modes of variation and thereby bias edge selection. Accordingly, for each target node j , we compute a node-specific basis $\{\phi_m^{(j)}\}$ intended to represent the variability of the signal at that node, and use it consistently across all regressors $k \neq j$ within the corresponding regression fit. This choice aligns the score representations entering (9) and facilitates interpretation of the coefficient blocks. Our implementation supports FPCA-based dimensionality reduction. More generally, user-defined preprocessing can be incorporated whenever alternative bases or alternative FPCA constructions, as the target-driven FPCA basis for each neighborhood regression in the spirit of Zhao et al. (2024), are of interest. The truncation level M is again a key hyperparameter: it governs the fidelity of the finite-dimensional embedding (larger M yields a more accurate representation of the underlying functions) but increases the dimension of the regression problem, and hence the computational burden and the sample size required for stable neighbourhood recovery. In FPCA-based implementations, we recommend selecting M based on the proportion of explained variance, assessed across nodes to ensure a comparable level of functional information is retained throughout the network.

A.3 Optimization hyperparameter selection

In the current implementation, two tuning parameters must be specified for each neighbourhood regression: the thresholding parameter ϵ_j in (13) and the group-lasso penalty λ_j in (12). Both parameters are selected in a node-specific manner via selective cross-validation (SCV) following a similar procedure to Zhao et al. (2024), using an error criterion that combines predictive fit with a BIC-type complexity penalty.

To construct the candidate set for λ , we compute from the data a value $\lambda_{\max}$ large enough that, for every $j \in V$, the solution of (12) is the all-zero estimator. We then consider a grid in log scale of penalties $\lambda \in (0, \lambda_{\max}]$ of user-specified size. Candidate values for ϵ are also user-defined. Given these sets, we select $(\lambda_j^*, \epsilon_j^*)$ by minimizing the average SCV criterion over K folds, as follows.

We fixed node p , and for each candidate pair (λ, ϵ) :

1. Using the full data, estimate the block coefficient matrix $\widehat{\mathbf{B}}(\lambda)$ by solving (12).
2. Define the estimated active block index set $\widehat{\mathcal{N}}_p(\lambda, \epsilon)$ by computing for each $c = 0, \dots, q$

$$\widehat{\mathcal{N}}_p^c(\lambda, \epsilon) = \left\{ k \in V \setminus \{p\} : \|\widehat{\mathbf{B}}_k^c(\lambda)\|_F > \epsilon \right\}$$
3. For each fold $\nu = 1, \dots, K$, let I_ν denote the training indices and $I_{\text{test}, \nu}$ the corresponding test indices. Re-estimate the coefficients by unpenalized least squares on I_ν , constrained to the screened set:

$$\begin{aligned} \tilde{\mathbf{B}}^{(\nu)} \in \underset{\mathbf{B} \in \mathbb{R}^{M \times M(p-1)(q+1)}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{A}_p - (\mathcal{A} * \mathcal{X})\mathbf{B}\|_F^2 \right\} \\ \text{s.t. } \tilde{\mathbf{B}}_k = 0 \text{ for all } k \notin \widehat{\mathcal{N}}_p(\lambda, \epsilon). \end{aligned}$$

4. For $i \in I_{\text{test}, \nu}$, define the refit residual $\widehat{\mathbf{re}}_i^{(\nu)} = \mathbf{A}_i - (\mathcal{A}_{:,i} * \mathcal{X}_{:,i})\tilde{\mathbf{B}}^{(\nu)}$, and compute the fold-specific SCV-RSS score

$$\text{SCV-RSS}_\nu(\lambda, \epsilon) = \sum_{i \in I_{\text{test}, \nu}} \left\| \widehat{\mathbf{re}}_i^{(\nu)} \right\|_2^2 + \log(|I_{\text{test}, \nu}|) \cdot \left| \widehat{\mathcal{N}}_p(\lambda, \epsilon) \right|.$$

We then set $(\lambda_j^*, \epsilon_j^*)$ to the pair minimizing the average criterion $\frac{1}{K} \sum_{\nu=1}^K \text{SCV-RSS}_\nu(\lambda, \epsilon)$.

To reduce computational cost when the candidate grids are large, we additionally implement a randomized hyperparameter search: rather than evaluating all (λ, ϵ) combinations, a user-defined proportion of pairs is sampled without replacement and assessed via the SCV-RSS criterion.

B.1 Complete output for the case-study on EEG functional connectivity.

In the case study described in Section 5, we apply our method to a set of EEG time-series from two groups of individuals (controls vs AUD patients) considering the group membership as covariate. As detailed in Algorithm 1, in this scenario our pipeline returns the population graphical model G^0 , characterizing functional connectivity in the controls, and the differential graphical model G^1 , describing modulation in connectivity induced by AUD. Furthermore, a weight is associated to each edge in G^1 to discriminate impaired or strengthened connections.

Figure 7 displays the adjacency matrices defined by G^0 and G^1 . We observe that the adjacency matrix of G^0 shows a block structures indicating inter-hemisphere connectivity mainly involving the frontal areas and the posterior-occipital ones. G^1 tends to have a sparser support than G^0 , which is the typical scenario where directly estimating the differential network may be beneficial (Zhao et al. 2022). Further comments on the differential network are provided in Section 5.

References

- Boyd, S. P., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. (2011), ‘Distributed optimization and statistical learning via the alternating direction method of multipliers’, *Foundations and Trends in Machine Learning* **3**(1), 1–122.

- Liu, S., Quinn, J. A., Gutmann, M. U., Suzuki, T. & Sugiyama, M. (2014), ‘Direct learning of sparse changes in markov networks by density ratio estimation’, *Neural computation* **26**(6), 1169–1197.
- Meinshausen, N. & Bühlmann, P. (2006), ‘High-dimensional graphs and variable selection with the lasso’, *The Annals of Statistics* **34**(3), 1436–1462.
- Moysidis, I. & Li, B. (2021), ‘Joint functional gaussian graphical models’, *arXiv preprint arXiv:2110.06653*.
- Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J. & Liu, H. (2013), ‘Individual variability in functional connectivity architecture of the human brain’, *Neuron* **77**(3), 586–595.
- Mumtaz, W., Kamel, N., Ali, S. S. A., Malik, A. S. et al. (2018), ‘An EEG-based functional connectivity measure for automatic detection of alcohol use disorder’, *Artificial intelligence in medicine* **84**, 79–89.
- Nentwich, M., Ai, L., Madsen, J., Telesford, Q. K., Haufe, S., Milham, M. P. & Parra, L. C. (2020), ‘Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI’, *NeuroImage* **218**, 117001.
- Qiao, X., Guo, S. & James, G. M. (2019), ‘Functional graphical models’, *Journal of the American Statistical Association* **114**(525), 211–222.
- Ramsay, J. O. & Silverman, B. W. (2005a), *Functional data analysis*, Springer.
- Ramsay, J. & Silverman, B. (2005b), *Functional data analysis*, Springer, New York.
- Sakkalis, V. (2011), ‘Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG’, *Computers in biology and medicine* **41**(12), 1110–1117.
- Sommariva, S., Subramaniam, N. P. & Parkkonen, L. (2025), ‘Cortical parcellation optimized for magnetoencephalography with a clustering technique’, *Scientific Reports* **15**(1), 6404.
- Vallarino, E., Sommariva, S., Piana, M. & Sorrentino, A. (2020), ‘On the two-step estimation of the cross-power spectrum for dynamical linear inverse problems’, *Inverse Problems* **36**(4), 045010.
- Xu, P. & Gu, Q. (2016), ‘Semiparametric differential graph models’, *Advances in neural information processing systems* **29**.
- Zapata, J., Oh, S.-Y. & Petersen, A. (2022), ‘Partial separability and functional graphical models for multivariate gaussian processes’, *Biometrika* **109**(3), 665–681.
- Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W. & Litke, A. (1995), ‘Event related potentials during object recognition tasks’, *Brain research bulletin* **38**(6), 531–538.
- Zhao, B., Wang, Y. S. & Kolar, M. (2022), ‘FuDGE: A method to estimate a functional differential graph in a high-dimensional setting’, *Journal of Machine Learning Research* **23**(82), 1–82.
- 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.
- Zhao, S. D., Cai, T. T. & Li, H. (2014), ‘Direct estimation of differential networks’, *Biometrika* **101**(2), 253–268.
- Zhu, H., Strawn, N. & Dunson, D. B. (2014), ‘Bayesian graphical models for multivariate functional data’, *J. Mach. Learn. Res.* **17**, 204:1–204:27.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 02/2026** Antonietti, P.F.; Beirao da Veiga, L.; Botti, M.; Harnist, A.; Vacca, G.; Verani, M.
Virtual Element methods for non-Newtonian shear-thickening fluid flow problems
- 01/2026** Iapaolo V.; Vergani, A.M.; Cavinato, L.; Ieva, F.
Multi-view learning and omics integration: a unified perspective with applications to healthcare
- Iapaolo, V.; Vergani, A.; Cavinato, L.; Ieva, F.
Multi-view learning and omics integration: a unified perspective with applications to healthcare
- 82/2025** Varetti, E.; Torzoni, M.; Tezzele, M.; Manzoni, A.
Adaptive digital twins for predictive decision-making: Online Bayesian learning of transition dynamics
- 81/2025** Leimer Saglio, C. B.; Pagani, S.; Antonietti, P. F.
A massively parallel non-overlapping Schwarz preconditioner for PolyDG methods in brain electrophysiology
- 79/2025** Zacchei, F.; Conti, P.; Frangi, A.; Manzoni, A.
Multi-Fidelity Delayed Acceptance: hierarchical MCMC sampling for Bayesian inverse problems combining multiple solvers through deep neural networks
- 78/2025** Botti, M.; Mascotto, L.
Trace inequalities for piecewise $W^{1,p}$ functions over general polytopic meshes
- Botti, M.; Mascotto, L.
Trace inequalities for piecewise $W^{1,p}$ functions over general polytopic meshes
- 77/2025** Bonetti, S.; Botti, M.; Vega, P.
A robust fully-mixed finite element method with skew-symmetry penalization for low-frequency poroelasticity
- 75/2025** Crippa, B.; Scotti, A.; Villa, A.
A one-dimensional reduced plasma model for the electrical treeing