



# Optimal design of experiments for functional linear models with dynamic factors

Caterina May<sup>1,2</sup> · Theodoros Ladas<sup>1</sup> · Davide Pigoli<sup>1</sup> · Kalliopi Mylona<sup>1</sup>

Received: 3 July 2025 / Accepted: 6 February 2026  
© The Author(s) 2026

## Abstract

In this work, we build optimal experimental designs for precise estimation of the functional coefficient of linear regression model where both the response and the factors are continuous functions. After obtaining the variance–covariance matrix of the estimator of the functional coefficient which minimizes the integrated sum of square of errors, we extend the classical definition of optimal design to this estimator, and we provide the expression of the A-optimal and of the D-optimal designs. Examples of optimal designs for dynamic experimental factors are then computed through a suitable algorithm, and we discuss different scenarios in terms of the set of basis functions used for their representation. Finally, we present an illustrative example inspired by a real application in a pharmaceutical manufacturing process, to illustrate the feasibility and the advantages of our methodology.

**Keywords** Dynamic experimental conditions · Estimation · Function-on-function linear model · Functional data analysis · Optimal design of experiments

**Mathematics Subject Classification** 62K05 · 62R10

---

✉ Caterina May  
caterina.may@uniupo.it; caterina.may@kcl.ac.uk

Theodoros Ladas  
theodoros.ladas@kcl.ac.uk

Davide Pigoli  
davide.pigoli@kcl.ac.uk

Kalliopi Mylona  
kalliopi.mylona@kcl.ac.uk

<sup>1</sup> Department of Mathematics, King's College London, Strand Campus, London WC2R 2LS, UK

<sup>2</sup> Dipartimento DiSEI, Università del Piemonte Orientale, V. Ettore Perrone 18, 28100 Novara, Italy

## 1 Introduction

In many applications, data are collected as curves over time, and it is therefore natural to consider statistical models with curves (or functions) as response and as covariates. This kind of problems has been considered for example in areas such as biomedical sciences (Shi et al. 2012), genetics (Müller et al. 2008), pharmaceutical sciences (Rahman et al. 2019) and transport engineering (Zhang et al. 2018), to mention just a few. However, most of these works have been concerned with observational data and little attention has been devoted to experimental design for settings where both the covariates and the response are functional in nature. An example of a laboratory experiment that is within this framework and motivated this research is the one described in Pigoli et al. (2023), in the context of forensic entomology. Entomologists are interested in studying the relationship between larval growth (and growth rate) and temperature, but both these quantities vary in time. In a laboratory setting, the current practice is to keep the incubator temperature constant for each experimental run, and then having various runs at different temperatures. However, this is a choice based only on practicality, since it does not require the operator to change the incubator temperature during the run. A natural question is if there are better experimental designs to explore the relationship between the growth rate curve and the temperature curve. This is an example of the kind of scenarios we wish to address in this work. Another example is the ergonomic experiment discussed in Aletti et al. (2016a, b) to forecast car driver's arm motion. Here, the control variables are the spatial coordinates of a target object inside the car cabin, which are scalar. The functional response is the angle formed at the right elbow between the upper and the lower arm varying during time to reach this object. As an extension to functional control variables, the target locations can change to specific trajectories that the hand must follow to move an object from one position to another. Other examples arise in industrial settings, where the response is functional (e.g., near infra-red spectroscopy measurements in chemical or pharmaceutical manufacturing, Blanco et al. 1998) and the process parameters (such as temperature and speed) can be varied over time. More recently, Chen et al. (2021) illustrated an application in bioengineering, where function-on-function Gaussian process regression is used to emulate the tissue-mimicking performance of synthetic materials. A motivating example, inspired by the pharmaceutical industry, will be described in the next section.

One of the main tasks of optimal experimental design is the choice of the best experimental conditions to estimate the unknown parameters of a statistical model. D-optimality and A-optimality, studied in this paper, are popular criteria that aim to minimize the generalized variance and the average variance of the parameter estimators, respectively. Classical references on optimal experimental design are, for instance, Fedorov (1972), Silvey (1980), Pukelsheim (2006) and Atkinson et al. (2007). The use of optimal design is crucial, since it can drastically lower the cost of experimentation by reducing the total number of observations needed to guarantee the desired precision in parameter estimation.

We consider a framework where both the experimental factors and the responses have a functional nature. In practice, the data analysis always starts from raw measurements, and a preprocessing smoothing step is required to represent observations

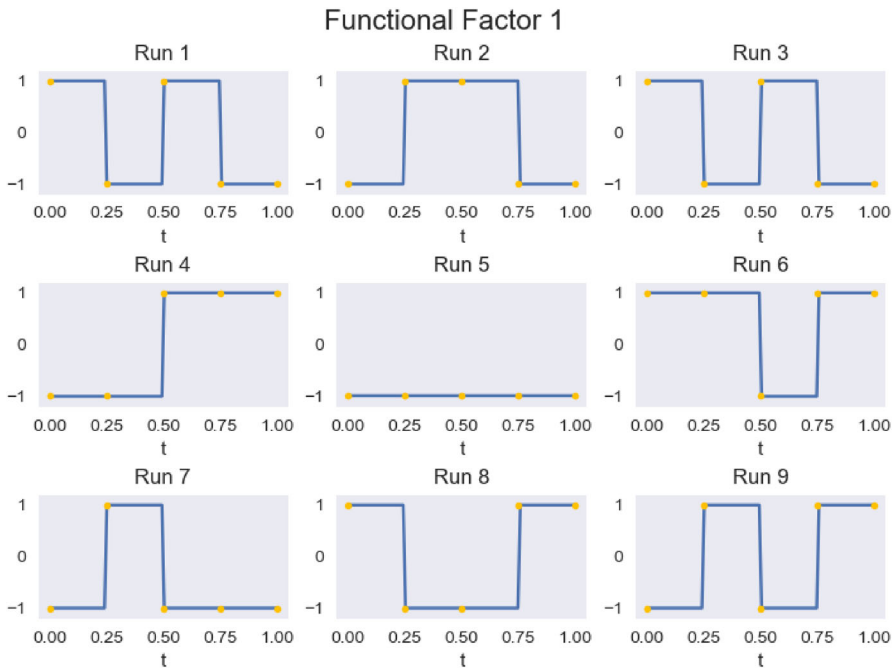
as functions. We refer to Ramsay and Silverman (2005) for an extensive discussion of the smoothing procedures that can be used to reconstruct the functional data; the starting point of our work will be responses and factors that are already represented as functions. Also, we restrict ourselves to the case of *densely observed* functional data, (Gertheiss et al. 2024), where smoothing can indeed be treated as a preprocessing step, while we will not discuss the case of *sparsely observed* functional data, where the location of the measurements can pose in itself an experimental design issue (see, e.g., Rha et al. 2020; Kao and Huang 2024).

In terms of statistical models, in this work we will consider function-on-function linear models (see, e.g., Ramsay and Silverman (2005) or Horváth and Kokoszka (2012)). To the best of our knowledge, statistical inference on these models has not received much attention. The original contribution of this paper is therefore twofold. First, we derive the expression of the covariance of the estimator obtained after basis expansion of the functional coefficient (Ramsay and Silverman 2005). Second, we use this result to construct optimal experimental designs with dynamic factors, which guarantees the most precise estimation. This will expand the current state of the art on experimental design for functional data, where optimal designs for experiments for scalar-on-function linear models (i.e., models with functional covariates but scalar response) have been considered by Michaelides et al. (2023) and for function-on-scalar linear models have been discussed in Aletti et al. (2014) and Aletti et al. (2016b). On the other hand, previous work on optimal design of experiment for problems with time-varying factors and response (see, e.g., Fedorov and Hackl 1997) has focused on models in which the functional form of the relationship between factors and responses is known (e.g., quadratic) or on kriging estimators (Morris 2014). In our approach, we instead consider the functional data framework where the model coefficients are themselves functional objects, represented in a flexible way through a basis expansion.

The paper is organized as follows. Section 2 describes in detail an illustrative example. The model is presented in Sect. 3. In Sect. 4, the estimator of the functional coefficient of multiple factors is derived, and its properties are proved. Section 5 is devoted to the construction of exact optimal experimental design on the base of the inference results previously obtained. In addition, Sect. 6 illustrates and discusses some concrete examples of optimal designs, obtained with an optimization algorithm. In Sect. 7, inspired by the motivating example in Sect. 2, we present a proof-of-concept example with simulated data, as an illustration of our methodology for practitioners. In Sect. 8, conclusions and future developments are discussed.

## 2 Motivating example

We illustrate the applicability of our framework with an industrial scenario based on a pharmaceutical manufacturing process (Van Snick et al. 2019). The process involves feeding material into a continuous line through a gravimetric feeder, while the outlet concentration of the material is measured in-line downstream. Subsequently, the mixture of material is abruptly changed, and it is measured how the resulting outlet concentration changes over the time of the experimental run.



**Fig. 1** Example of a nine-run A-optimal design for the industrial case study. Each panel shows the dosing trajectory of a single experimental run as a stepwise input over normalized time  $t \in [0, 1]$

Hence, each experimental run can be viewed as a dynamic input–output system. The input is the dosing program applied at the feeder, which specifies the exact mixture of material that is fed over time. In practice, these dosing profiles are typically stepwise functions, as the feeder may be turned on or off, or operated at a few different levels in successive time intervals. The output is the concentration measured at the outlet, also observed over time. Because of transport delays and mixing, changes in the input are reflected in the output only after some lag, and the response tends to be smoother than the input. This leads to a setting where both the experimental factors and the responses are functions of time. Each run corresponds to one pair of functional trajectories: a dosing profile and the resulting concentration curve.

In the original study of Van Snick et al. (2019), these functional quantities were simplified into scalar or low-dimensional summaries before designing the experiment. For instance, researchers worked with cumulative or normalized versions of the concentration curve (such as residence-time distributions). The experiment was then designed for these scalar parameters rather than for the underlying functional relationship.

Our framework treats the dosing program and the measured concentration directly as functions. Instead of designing an experiment around scalar summaries, we design the specific trajectories of mixture of material to be tested. The goal is to select a small number of profiles that provide the most precise information about how input activity at one time influences the response at another time.

A possible A-optimal design with 9 runs obtained under our framework is illustrated in Fig. 1. Each of the nine runs corresponds to a distinct dosing program applied to the feeder, where the feed level switches between  $-1$  (lowest setting) and  $1$  (highest setting) at different time intervals. These stepwise profiles were obtained using a step-function basis for the input factor, reflecting the discrete operating modes of the feeder. The corresponding process response was represented using a smoother, second-order spline basis to capture the gradual transport and mixing effects observed in the outlet concentration. This combination of stepwise input and smooth output is consistent with the physical behavior the experimental setup of Van Snick et al. (2019). Collectively, they provide rich coverage of the input domain, ensuring that changes in the input mixture occur at various stages of the run. This diversity allows the resulting data to capture how early, mid-run, and late changes in dosing propagate through the system.

### 3 Model and notation

Throughout this paper, we consider the following function-on-function linear model, which generalizes to multiple factors the model introduced by Ramsay and Dalzell (1991) and further discussed, for example, by Ramsay and Silverman (2005) and Kokoszka and Reimherr (2017):

$$y_n(t) = \beta_0(t) + \sum_{i=1}^p \int_0^{\mathcal{T}} \beta_i(s, t) x_{ni}(s) ds + \varepsilon_n(t), \quad n = 1, \dots, N, \quad p \geq 1 \quad (1)$$

where  $y_n \in L^2([a, b])$ ,  $x_{ni} \in L^2([0, \mathcal{T}])$ ,  $\beta_0 \in L^2([a, b])$  and  $\beta_i \in L^2([0, \mathcal{T}] \times [a, b])$  are square-integrable functions, and  $\varepsilon_1(t), \dots, \varepsilon_N(t)$  are i.i.d. zero-mean stochastic processes (not necessarily Gaussian). Without loss of generality, we assume that all factors are defined over the same interval  $[0, \mathcal{T}]$ . If this is not the case—i.e., if the factors are defined on different intervals or involve different independent variables such as time, frequency, or spatial coordinates—they can nevertheless be rescaled to a common standardized interval  $[0, \mathcal{T}]$ . Similarly, the independent variable  $t$  in the response is not necessarily the same as in the factors and even when it is the same, it can be defined on a different interval.

Furthermore, the model formulation in (1) assumes that the response at any point may depend on the entire domain of the factor. In the special case where the independent variable is time for both the response and one or more of the factors, this assumption may be unrealistic, as it would imply that the response depends on future values of the factor. In such cases, a more appropriate approach is the *historical functional linear model* (Malfait and Ramsay 2003), in which the integration is restricted to the interval  $[0, t]$ . While the general methodology proposed in this paper can be extended to the historical model, it requires additional numerical computations to approximate the integrals. For this reason, we do not consider the historical model further in this work.

In matrix notation, model (1) can be rewritten as

$$\mathbf{y}(t) = \int_0^{\mathcal{T}} \mathcal{X}(s) \boldsymbol{\beta}(s, t) ds + \boldsymbol{\varepsilon}(t), \quad (2)$$

where  $\mathbf{y}(t) = (y_1(t), \dots, y_N(t))^T$  is the vector of responses,  $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \dots, \varepsilon_N(t))^T$  is the vector of errors,  $\boldsymbol{\beta}(s, t) = (\beta_0(t), \beta_1(s, t), \dots, \beta_p(s, t))^T$  is the functional coefficient to be estimated, and  $\mathcal{X}(s)$  is the  $N \times (p + 1)$  model matrix:

$$\mathcal{X}(s) = [\mathbf{1}_N \mid \mathbf{X}(s)], \quad (3)$$

with  $\mathbf{X}(s)$  being the  $N \times p$  design matrix of the dynamic predictors:

$$\mathbf{X}(s) = [\mathbf{x}_1(s) \ \mathbf{x}_2(s) \ \dots \ \mathbf{x}_p(s)] = \begin{bmatrix} x_{11}(s) & \dots & x_{1p}(s) \\ x_{21}(s) & \dots & x_{2p}(s) \\ \vdots & & \vdots \\ x_{N1}(s) & \dots & x_{Np}(s) \end{bmatrix} \quad (4)$$

We consider a basis  $\{\theta_l(t), l \geq 1\}$  to expand the responses  $y_n(t)$ , and the errors  $\varepsilon_n(t)$ , for  $n = 1, \dots, N$ . We can then expand the intercept  $\beta_0(t)$  and the kernels  $\beta_i(s, t)$ , for any  $i = 1, \dots, p$ , according to the basis  $\{\theta_l(t), l \geq 1\}$  and to  $p$  bases  $\{\eta_k^1(s), k \geq 1\}, \dots, \{\eta_k^p(s), k \geq 1\}$ :

$$\beta_0(t) = \sum_l b_l^0 \theta_l(t),$$

$$\beta_i(s, t) = \sum_{k,l} b_{k,l}^i \eta_k^i(s) \theta_l(t).$$

Since, by the Gram–Schmidt orthonormalization procedure, any basis can be transformed into an orthonormal basis, from now on we assume that the basis  $\{\theta_l(t), l \geq 1\}$  is orthonormal, that is,

$$\langle \theta_h(t), \theta_k(t) \rangle_{L^2} = \int \theta_h(t) \theta_k(t) dt = \delta_{hk},$$

where is the Kronecker delta symbol:  $\delta_{hk} = 1$  if  $h = k$  and  $\delta_{hk} = 0$  if  $h \neq k$ .

In order to estimate the parameters of the functional model (1), we follow the ideas contained in Ramsay and Silverman (2005) and in Horváth and Kokoszka (2012); their fully functional model is here extended to multiple factors and, in Sect. 4, we derive explicitly the estimator for the parameters of our model. Let us project the unknown coefficients in the subspaces generated by the finite bases  $\{\theta_l(t), 1 \leq l \leq L\}$  and  $\{\eta_k^1(s), 1 \leq k \leq K^1\}, \dots, \{\eta_k^p(s), 1 \leq k \leq K^p\}$ . Note that each coefficient  $\beta_i(s, t)$  for  $i = 1, \dots, p$ , can be expanded, with respect to their first variable  $s$ , in a different basis and with a proper dimension  $K^i$ , according to the features of the corresponding factor.

We have:

$$\beta_0^*(t) = \mathbf{b}^{0T} \boldsymbol{\theta}(t),$$

$$\beta_i^*(s, t) = \boldsymbol{\eta}^i(s)^T \mathbf{B}^i \boldsymbol{\theta}(t),$$

for any  $i = 1, \dots, p$ , where

$$\mathbf{b}^0 = (b_1^0, \dots, b_L^0)^T,$$

$$\boldsymbol{\theta}(t) = (\theta_1(t), \dots, \theta_L(t))^T,$$

$$\boldsymbol{\eta}^i(s) = (\eta_1^i(s), \dots, \eta_{K^i}^i(s))^T,$$

and

$$\mathbf{B}^i = [b_{k,l}^i]$$

is a  $K^i \times L$  matrix. Then, the functional coefficient  $\boldsymbol{\beta}(s, t)$  becomes

$$\boldsymbol{\beta}^*(s, t) = (\beta_0^*(t), \beta_1^*(s, t), \dots, \beta_p^*(s, t))^T = \mathbf{H}(s)^T \mathbf{B} \boldsymbol{\theta}(t) \tag{5}$$

where  $\mathbf{H}(s)^T$  is a  $(p + 1) \times (1 + K^1 + \dots + K^p)$  diagonal block matrix given by

$$\mathbf{H}(s)^T = \text{diag}(1, \boldsymbol{\eta}^1(s)^T, \dots, \boldsymbol{\eta}^p(s)^T),$$

and  $\mathbf{B}$  is an  $(1 + K^1 + \dots + K^p) \times L$  block matrix of coefficients

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}^{0T} \\ \mathbf{B}^1 \\ \vdots \\ \mathbf{B}^p \end{bmatrix}$$

to be estimated. Model (2) becomes:

$$\mathbf{y}(t) = \int_0^T \mathcal{X}(s) \boldsymbol{\beta}^*(s, t) ds + \boldsymbol{\varepsilon}(t),$$

that is, putting the (5) into it,

$$\mathbf{y}(t) = \mathbf{Z} \mathbf{B} \boldsymbol{\theta}(t) + \boldsymbol{\varepsilon}(t), \tag{6}$$

where  $\mathbf{Z}$  is the  $N \times (1 + K^1 + \dots + K^p)$  matrix given by

$$\mathbf{Z} = \int \mathcal{X}(s) \mathbf{H}(s)^T ds. \tag{7}$$

## 4 Estimation of the model

In this section, we derive an estimator of the unknown functional coefficients of model (3), and we obtain the expression of its variance. This is equivalent to estimating the unknown block matrix  $B$  of coefficients in (6); to this aim, we minimize the Integrated Sum of Square of Errors (ISSE), and then, we define the matrix estimator

$$\hat{B} = \arg \min_B \text{ISSE}, \quad (8)$$

where

$$\begin{aligned} \text{ISSE} &= \int \|\mathbf{y}(t) - Z B \boldsymbol{\theta}(t)\|^2 dt \\ &= \int \text{tr}(\mathbf{y}(t) - Z B \boldsymbol{\theta}(t))^T (\mathbf{y}(t) - Z B \boldsymbol{\theta}(t)) dt \\ &= \text{tr} \int (\mathbf{y}(t)^T \mathbf{y}(t) + \boldsymbol{\theta}(t)^T B^T Z^T Z B \boldsymbol{\theta}(t) - 2\boldsymbol{\theta}(t)^T B^T Z^T \mathbf{y}(t)) dt \\ &= \text{tr} \int \mathbf{y}(t)^T \mathbf{y}(t) dt + \text{tr} B^T Z^T Z B J_{\theta\theta} - 2 \text{tr} B^T Z^T \int \mathbf{y}(t) \boldsymbol{\theta}(t)^T dt; \quad (9) \end{aligned}$$

the last equality follows from the cyclic property of the trace and by defining

$$J_{\theta\theta} = \int \boldsymbol{\theta}(t) \boldsymbol{\theta}(t)^T dt.$$

To obtain the estimator  $\hat{B}$ , we calculate the matrix derivatives of (9) and set them equal to zero; according to the properties of matrix calculus we have:

$$\begin{aligned} \frac{\partial}{\partial B} \text{tr} B^T Z^T Z B J_{\theta\theta} - 2 \frac{\partial}{\partial B} \text{tr} B^T Z^T \int \mathbf{y}(t) \boldsymbol{\theta}(t)^T dt \\ = 2 Z^T Z B J_{\theta\theta} - 2 Z^T \int \mathbf{y}(t) \boldsymbol{\theta}(t)^T dt = 0, \end{aligned}$$

and hence, we obtain the normal equations:

$$Z^T Z B J_{\theta\theta} = Z^T \int \mathbf{y}(t) \boldsymbol{\theta}(t)^T dt.$$

Since the basis  $\{\theta_l(t), l \geq 1\}$  is orthonormal, then  $J_{\theta\theta} = I_L$ ; if  $Z$  in (7) has maximum rank  $(1 + K^1 + \dots + K^p)$ , that is,  $Z^T Z$  is invertible, we have the following expression of the estimator (8):

$$\hat{B} = (Z^T Z)^{-1} Z^T \int \mathbf{y}(t) \boldsymbol{\theta}(t)^T dt. \quad (10)$$

The following result provides the variance of the vectorization of the matrix estimator (10). Given a  $p \times q$  matrix  $M$ ,  $\text{vec}(M)$  is the  $pq$ -vector obtained by stacking the columns of  $M$ ; the symbol  $\otimes$  denotes the Kronecker product on two matrices.

**Proposition 1** *The estimator (10) is an unbiased estimator of the matrix  $B$  of coefficients of  $\beta^*(s, t)$ , and its variance is*

$$Var(\text{vec } \hat{B}) = \Sigma \otimes (Z^T Z)^{-1}.$$

where  $\Sigma$  is a  $L \times L$  matrix whose elements are

$$\Sigma_{il} = Cov(e_{ni}, e_{nl}),$$

not depending on  $n$ , with  $e_{ni}$  and  $e_{nl}$  being the  $i$ -th and the  $l$ -th coefficients, respectively, of the expansion of  $\varepsilon_n(t)$  in the basis  $\{\theta_l(t), l = 1, \dots, L\}$ .

**Proof** The unbiasedness follows from the interchangeability between expectation and integration, and because  $J_{\theta\theta} = I_L$ , from the orthonormality of  $\{\theta_l(t), l = 1, \dots, L\}$ .

By considering the representation  $y(t) = Y\theta(t)$  of the response vector, where  $Y$  is a  $N \times L$  matrix of coefficients, we have that

$$\int y(t)\theta(t)^T dt = Y J_{\theta\theta} = Y,$$

and thus, the estimator (10) can be expressed as  $\hat{B} = (Z^T Z)^{-1} Z^T Y$ . Note that the random coefficients in each  $n$ -th row of  $Y$ , for  $n = 1, \dots, N$ , are dependent between them, since they expand the same functional response  $y_n(t)$ , while each row is independent on the others since each row expands a different response.

In order to obtain an expression of the variance of estimator (10), it is then convenient to concatenate the rows of  $Y$  and, to get it, to vectorize its transpose  $\hat{B}^T$ :

$$\hat{B}^T = Y^T Z (Z^T Z)^{-1}.$$

If  $\Phi$  and  $\Psi$  are a  $p \times q$  and an  $r \times s$  matrices, respectively, it is possible to show that

$$\text{vec}(\Phi\Psi) = (\Psi^T \otimes I_p) \text{vec}(\Phi).$$

Hence,

$$\text{vec } \hat{B}^T = [((Z^T Z)^{-1} Z^T) \otimes I_L] \text{vec}(Y^T)$$

and

$$Var(\text{vec } \hat{B}^T) = F \Delta F^T,$$

where

$$F = [((Z^T Z)^{-1} Z^T) \otimes I_L] \text{ and } \Delta = Var(\text{vec } Y^T),$$

which are, respectively, a  $(1 + \sum_{i=1}^p K^i) L \times NL$  and a  $NL \times NL$  matrices. To obtain the matrix  $\Sigma$ , observe that since  $y_1(t), \dots, y_N(t)$  are independent and identically distributed stochastic processes, then the columns of  $Y^T$ , that is,  $\int y_1(t)\theta(t)^T dt, \dots, \int y_N(t)\theta(t)^T dt$ , are i.i.d. random vectors. Then,

$$\Delta = I_N \otimes \Sigma,$$

where

$$\Sigma_{il} = \text{Cov} \left( \int y_n(t)\theta_i(t) dt, \int y_n(t)\theta_l(t) dt \right) \quad (11)$$

which does not depend on  $n$ . As a consequence of model (1), the second term in (11) is equal to

$$\text{Cov} \left( \int \varepsilon_n(t)\theta_i(t) dt, \int \varepsilon_n(t)\theta_l(t) dt \right);$$

moreover, for any  $n$  and  $i$ ,

$$\int \varepsilon_n(t)\theta_i(t) dt = \int \sum_{l=1}^L e_{nl}\theta_l(t)\theta_i(t) dt = \sum_{l=1}^L e_{nl} \int \theta_l(t)\theta_i(t) dt = e_{ni},$$

where the last equation follows from the orthonormality of  $\{\theta_l, l = 1, \dots, L\}$ . We finally obtain

$$\begin{aligned} \text{Var}(\text{vec } \hat{\mathbf{B}}^T) &= [((\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T) \otimes \mathbf{I}_L] (\mathbf{I}_N \otimes \Sigma) [( \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}) \otimes \mathbf{I}_L] \\ &= (((\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T) \otimes \Sigma) [(\mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}) \otimes \mathbf{I}_L] \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \Sigma. \end{aligned} \quad (12)$$

The two last equalities are due the following property of the Kronecker product: If  $\mathbf{A}$ ,  $\Phi$ ,  $\Psi$  and  $\Omega$  are matrices of sizes that can form the matrices products  $\mathbf{A} \Psi$  and  $\Phi \Omega$ , then

$$(\mathbf{A} \otimes \Phi)(\Psi \otimes \Omega) = (\mathbf{A} \Psi) \otimes (\Phi \Omega).$$

Finally, to obtain  $\text{Var}(\text{vec } \hat{\mathbf{B}})$  from the (12), let us denote by  $K^{(p,q)}$  the commutation matrix of dimension  $pq \times pq$  such that  $\text{vec}(M^T) = K^{(p,q)} \text{vec}(M)$  for any  $p \times q$  matrix  $M$  (see (Magnus and Neudecker (2019), Chapt. 18)). Then,

$$\text{vec } \hat{\mathbf{B}} = K^{(L, 1 + \sum_{i=1}^p K^i)} \text{vec}(\hat{\mathbf{B}}^T),$$

and hence,

$$\begin{aligned} \text{Var}(\text{vec } \hat{\mathbf{B}}) &= K^{(L, 1 + \sum_{i=1}^p K^i)} ((\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \Sigma) K^{(1 + \sum_{i=1}^p K^i, L)} \\ &= \Sigma \otimes (\mathbf{Z}^T \mathbf{Z})^{-1}, \end{aligned}$$

where the first equality is justified by the property  $K^{(p,q)T} = K^{(q,p)}$ , while the final equality follows from the property of the commutation matrices with Kronecker product (see again Magnus and Neudecker (2019, Chapt. 18)).  $\square$

## 5 Optimal experimental designs

An experimental design for the model described in Sect. 3 is defined as a set of experimental conditions  $x_{n1}(s), \dots, x_{np}(s)$  in the interval  $[0, T]$  chosen by the experimenter

for each experimental run  $n = 1, \dots, N$ , according to a specific experimental objective; an experimental design is then identified by the design matrix  $X(s)$  in (4).

Extending the classical definition of the optimal design literature, see, for instance, (Atkinson et al. 2007), to the functional framework considered in this paper, an optimal experimental design is an experimental design  $X^*(s)$  which minimizes a proper function  $\Phi$  of the variance–covariance matrix of the estimator. From Proposition 1, it follows that an optimal design is given by

$$X^*(s) = \arg \min_{X(s) \in \mathcal{C}} \Phi(\Sigma \otimes (Z^T Z)^{-1}), \tag{13}$$

with the  $Z$  given by (7) and  $\mathcal{C}$  is the functional space where the factors are defined. Up to this point, we have maintained a general functional representation for the factors. However, to proceed with the design, it is necessary to express these factors using a basis expansion. The choice of basis functions defines the functional space, denoted by  $\mathcal{C}$ , in which the factors reside. In practical applications, the selection of basis functions may be influenced by constraints arising from the underlying physics of the problem or from economic considerations. For instance, if the factors are required to be smooth, higher-order spline bases may be appropriate; if monotonicity is required, I-splines may be used (Ramsay 1988). In these cases, the constraints guide the selection of the basis functions directly. In other cases, the constraints must be incorporated into the optimization problem explicitly. Since such requirements are typically application-specific, we assume in the following that  $\mathcal{C}$  is the space spanned by the selected basis functions, and that the optimization problem is unconstrained with respect to the basis coefficients. Nonetheless, it is straightforward to modify the formulation to include additional constraints by specifying certain coefficient values as inadmissible.

Let us expand each factor  $\mathbf{x}_i(s)$ , for  $i = 1, \dots, p$ , according to a proper basis  $\{c_k^i(s), k \leq 1\}$ , and write the experimental conditions as finite linear combinations of the first  $K_x^i$  elements of the respective basis:

$$x_{ni}(s) = \boldsymbol{\gamma}_n^{i T} \mathbf{c}^i(s),$$

for any  $n = 1, \dots, N$  and  $i = 1, \dots, p$ , where  $\boldsymbol{\gamma}_n^i$  is a  $K_x^i$ -dimensional vector of coefficients.

Recalling that the model matrix (3) is obtained by adding a column of one’s to the matrix of the experimental conditions, we have that it can be rewritten as

$$\mathcal{X}(s) = \Gamma C(s),$$

where

$$C(s) = \text{diag}(1, \mathbf{c}^1(s), \dots, \mathbf{c}^p(s)),$$

and  $\Gamma$  is a  $N \times (1 + K_x^1 + \dots + K_x^p)$  block matrix given by

$$\Gamma = (\mathbf{1}_N \mid \Gamma^1 \mid \dots \mid \Gamma^p), \tag{14}$$

with

$$\Gamma^i = \begin{bmatrix} \boldsymbol{\gamma}_n^{i T} \\ \vdots \\ \boldsymbol{\gamma}_N^{i T} \end{bmatrix}$$

Hence, the matrix  $Z$  in (7) can be rewritten as

$$Z = \Gamma \int C(s) H(s)^T ds = \Gamma J_{CH} \quad (15)$$

where

$$J_{CH} = \text{diag}(1, J_{c\eta^1}, \dots, J_{c\eta^p}) \quad (16)$$

and, for any  $i = 1, \dots, p$ ,

$$J_{c\eta^i} = \int \mathbf{c}^i(s) \boldsymbol{\eta}^i(s)^T ds.$$

Hence, an optimal design  $X^*(s)$  defined by Eq. (13) is identified by a matrix  $\Gamma^*$  such that

$$\Gamma^* = \arg \min_{\Gamma} \Phi(\Sigma \otimes (J_{CH}^T \Gamma^T \Gamma J_{CH})^{-1}). \quad (17)$$

## 5.1 A- and D-optimality

In this section, we extend A- and D-optimality criteria to the dynamic experimental conditions: A functional A-optimal design  $X_A^*(s)$  and a functional D-optimal design  $X_D^*(s)$  satisfy the (13), with  $\Phi = \text{tr}$ , and  $\Phi = \det$ , respectively. Hence, the D- and the A-optimal designs ensure the precise estimation of the functional coefficients by minimizing the generalized variance and the average variance of the parameter estimators, respectively. In the following propositions, we derive the expression of these criteria.

**Proposition 2** *A functional A-optimal design  $X_A^*(s)$  is obtained by minimizing*

$$\text{tr}(Z^T Z)^{-1},$$

where  $Z$  is the matrix defined in (7).

**Proof** From Proposition 1 and from the property of Kronecker product that

$$\text{tr}(A \otimes B) = (\text{tr } A)(\text{tr } B),$$

we have that

$$\text{tr } \text{Var}(\text{vec } \hat{B}) = \text{tr } \Sigma \text{tr}(Z^T Z)^{-1}; \quad (18)$$

note that  $\Sigma$  is the covariance matrix of the error process and it is not affected by the design; then, if a functional design minimizes  $\text{tr}(Z^T Z)^{-1}$ , it minimizes also the (18), and hence, it is A-optimal.  $\square$

**Proposition 3** A functional D-optimal design  $X_D^*(s)$  is obtained by maximizing

$$\det(Z^T Z),$$

where  $Z$  is defined in (7).

**Proof** Again, from Proposition 1 and from the property of Kronecker product that if  $A$  and  $B$  are an  $n \times n$  and a  $m \times m$  matrices, respectively, then

$$\det(A \otimes B) = (\det A)^m (\det B)^n,$$

we have that

$$\det \text{Var}(\text{vec } \hat{B}) = \frac{(\det \Sigma)^{1+\sum_{i=1}^p K^i}}{(\det Z^T Z)^L}; \tag{19}$$

then, if a functional design maximizes  $\det Z^T Z$ , it minimizes the (19), and hence, it is D-optimal.  $\square$

Propositions 2 and 3 show that that A-optimal and D-optimal designs do not depend on the unknown covariance matrix  $\Sigma$  of the error process. This is a crucial result because it allows us to generate optimal designs without the need of information from the data, which would be available only after the experiment.

Note that the expression therein contained is still based on the general representation of the design matrix, which is integrated to obtain the matrix  $Z$ . According to our method, the experimental conditions are expressed as linear combinations of proper basis of functions, and hence the optimization need to be done with respect to the basis coefficients, as it can be seen in equation (17). The optimal designs will then depend on the original choices of the bases for the factors and of the bases for the functional coefficients in the direction of  $s$ . The range of admissible functional shapes can be freely chosen by the experimenter, taking into account the experimental constraints. Several examples will be provided and discussed in Sect. 6.

## 6 Results: optimal designs

This section is devoted to offer some examples of functional optimal experimental designs for precise estimation of the functional coefficient of model (3), based on the theory developed in the previous sections. The examples are organized as follows: First, we consider the same basis of functions to expand the functional coefficient and the runs of one functional predictor (Sect. 6.1); then, we extend to the case of choosing two different bases (Sect. 6.2); finally, we present an example with more than one predictor (Sect. 6.3).

In all of the examples, various B-splines (Smith 1980) are used as a choice for expanding the basis for both the functional factors and parameters. A B-spline is

defined by its degree  $D$  and its knot vector. If we assume that the knot vector is constructed by equidistant points in the interval  $[0, 1]$ , then we can define it in terms of the number of breakpoints  $\mathcal{K}$  (in our notation including the two bound points 0 and 1). For example, the knots vector with  $\mathcal{K} = 5$  should be  $\{0, 0.25, 0.5, 0.75, 1\}$ . The length for a basis of degree  $D$ , using  $\mathcal{K}$  knots in the knot vector, is  $D + \mathcal{K} - 1$ .

In the examples with one functional predictor, we assume that  $N = 12$  runs are available and we investigate how increasing the number of breakpoints for expanding the functional factor  $\mathbf{x}(s)$  affects the optimality. Additionally, we increase the number of break points in the expansion of  $\beta(s, t)$  with respect to  $s$ , to investigate how the design output changes for a constant number of breakpoints in  $X(s)$ . For a design to be estimable, we need the total number of parameters (plus one to include a constant term according to (3)) to be less than the number of experimental runs. So if we expand  $\beta(s, t)$  with a B-spline of degree  $D_\beta$  and  $\mathcal{K}_\beta$  breakpoints, we need  $N > D_\beta + \mathcal{K}_\beta$ . Additionally, if we use a B-spline of degree  $D_X$  and  $\mathcal{K}_X$  knots for expanding the functional factor  $X(s)$ , we need to enforce  $D_X + \mathcal{K}_X \geq D_\beta + \mathcal{K}_\beta$ , to ensure identifiability of (3).

Additionally, we explore scenarios involving multiple functional predictors to demonstrate how the interplay between different factors and their respective basis expansions influences the optimal design. Specifically, Sect. 6.3 presents an example with more than one predictor, illustrating the complexities and considerations that arise when extending the design framework to accommodate multiple functional inputs.

The algorithm used is a version of the coordinate exchange (Meyer and Nachtsheim 1995) using 1000 random starts implemented in Python 3.10 (Van Rossum and Drake 2009). The code is available by the authors upon request. The examples are referred to A-optimality, and similarly, D-optimum designs could be obtained.

## 6.1 Same basis of functions to expand the dynamic factor and the parameter

In this scenario, we consider one functional factor and we expand both  $X(s)$  and the functional parameter  $\beta(s, t)$  using zero-degree B-spline basis functions ( $D_X = 0$  and  $D_\beta = 0$ ). Zero-degree B-splines are defined over a set of equidistant breakpoints (or knots) within the domain of the function. Between each pair of consecutive breakpoints, the basis function is constant.

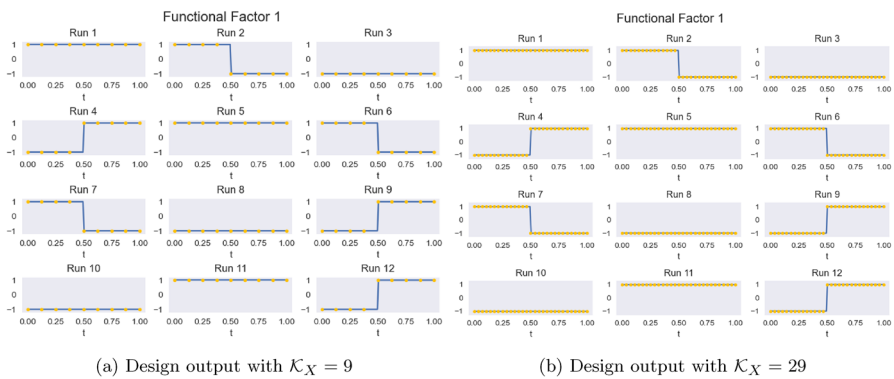
In our analysis, we investigate how varying the number of breakpoints in the B-spline basis for  $X(s)$  affects the A-optimality criterion value while keeping the basis and number of breakpoints of  $\beta(s, t)$ , with respect to the  $s$  variable, fixed. More specifically, we assume  $\mathcal{K}_X \in \{3, 5, 9, 15, 19, 29\}$  and  $\mathcal{K}_\beta \in \{3, 5, 7\}$  and we vary the number of breakpoints  $\mathcal{K}_X$  for each fixed  $\mathcal{K}_\beta$ .

Table 1 summarizes the A-optimality values and the relative efficiencies for different combinations of  $\mathcal{K}_X$  and  $\mathcal{K}_\beta$ . The computed efficiency is relative to the best A-optimality value achieved in each case.

From Table 1, we observe that when  $\mathcal{K}_\beta = 3$ , the efficiency remains constant at 100% no matter the number of breakpoints for  $X(s)$ . This indicates that increasing the number of breakpoints of the predictor's basis functions does not improve the

**Table 1** A-optimality values and relative efficiencies for zero-degree B-splines expansion of  $X(s)$  and  $\beta(s, t)$  with varying numbers of breakpoints in  $X(s)$

Breaks	3		5		7	
	A-opt	Efficiency	A-opt	Efficiency	A-opt	Efficiency
3	0.75	1.00	—	—	—	—
5	0.75	1.00	5.42	1.00	—	—
9	0.75	1.00	5.42	1.00	27.25	0.69
15	0.75	1.00	6.33	0.86	23.46	0.80
19	0.75	1.00	6.10	0.89	18.70	1.00
29	0.75	1.00	5.42	1.00	21.04	0.89



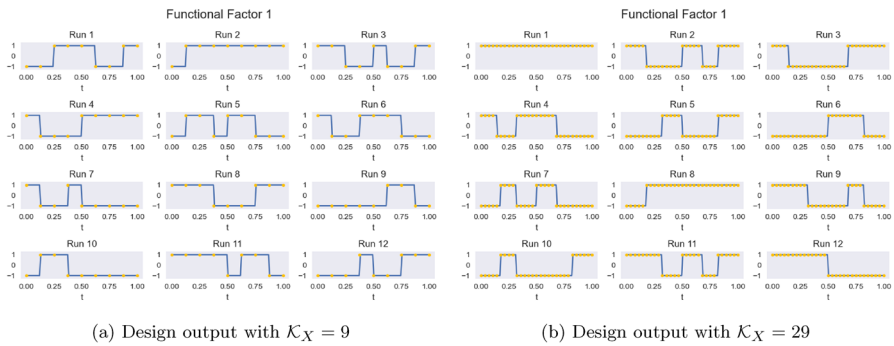
**Fig. 2** Comparison of design output for different breakpoints when  $\mathcal{K}_\beta = 3$

estimation precision of  $\beta(s, t)$  when the coefficient function is represented with only two ( $D_\beta = 0, \mathcal{K}_\beta = 3$ ) basis functions.

This observation is further illustrated in Fig. 2, where we compare the optimal design outputs for  $\mathcal{K}_X = 15$  and  $\mathcal{K}_X = 29$  with  $\mathcal{K}_\beta = 3$ . Despite the increase in the number of breakpoints for  $X(s)$ , the overall shape of the optimal design remains unchanged. Essentially, the simplicity of the coefficient function limits the complexity that can be captured in the design, and thus, even with more breakpoints, the optimal design does not change.

In contrast, when  $\mathcal{K}_\beta = 7$ , allowing for a more flexible representation of the factor by increasing the number of breakpoints improves the efficiency of the design, albeit with diminishing returns. A more flexible basis for  $X(s)$  allows the design to capture the increased complexity of the more intricate coefficient function  $\beta(s, t)$ , enhancing the estimation precision.

This can be seen in the design output in Fig. 3. Increasing the number of breakpoints for  $X(s)$  from 9 to 29 changes the shapes. Upon closer inspection of Fig. 3, we see that there is no run in the left panel that matches in shape any from the right panel. This reinforces our belief that a more complicated expansion for the functional factor is better at capturing a more complicated expansion of the functional parameter.



**Fig. 3** Comparison of design output for different breakpoints when  $K_\beta = 7$

An interesting observation arises from Table 1 concerning the case with 19 breakpoints for  $X(s)$  and 7 breakpoints for  $\beta(s, t)$ . Notably, this specific configuration yields a better A-optimality criterion value than the case with 29 breakpoints for  $X(s)$ . To understand this why this happens, it is essential to examine the knot vectors of  $X(s)$  and  $\beta(s, t)$  in this scenario.

We deduce that the knot vector of  $\beta(s, t)$  is included within the knot vector of  $X(s)$  when 19 breakpoints are used. This inclusion means that every knot of  $\beta(s, t)$  aligns with a knot of  $X(s)$ , facilitating a more precise estimation of the coefficient function. In contrast, this alignment does not occur for  $X(s)$  with 9 or 15 breakpoints, where the step sizes do not result in such inclusion. This observation suggests that carefully aligning the knot vectors of  $X(s)$  and  $\beta(s, t)$  can enhance the estimation precision, and the knot vector of the predictor should be thoughtfully selected based on that of the coefficient function.

A similar rationale applies to the case with 5 breakpoints for  $\beta(s, t)$  and 15 or 19 breakpoints for  $X(s)$ . In these scenarios, the knot vector of the coefficient function is not included within that of the predictor. As a result, the design cannot achieve 100% efficiency compared to cases where such inclusion exists.

These findings suggest the importance of matching not only the number of knots of the predictor's basis functions with the complexity of the coefficient function but also ensuring the alignment of their knot vectors. When the coefficient function is more intricate, increasing the number of knots of the predictor's basis functions is beneficial. However, optimal efficiency is further achieved when the knot vectors are carefully constructed so that the knots of the coefficient function are included within those of the predictor. This alignment allows the predictor to capture the essential features of the coefficient function more effectively, leading to improved estimation precision.

In the next section, we will explore scenarios where different families of basis functions are employed for the functional predictor and the parameter. This investigation will further elucidate how the choice of basis functions and the alignment of knot vectors impact the efficiency and effectiveness of the experimental design.

## 6.2 Different bases of functions

In this scenario, we explore the optimal experimental design when the functional predictor  $X(s)$  and the functional coefficient  $\beta(s, t)$  are expanded using different bases functions. Specifically, we consider cases where the predictor and the coefficient function are represented using B-splines of different degrees. This investigation aims to understand how the choice of basis functions for the predictor and the coefficient affects the efficiency of the design.

We focus on two sub-scenarios:

1.  $D_X = 1$  and  $D_\beta = 2$ : This setup allows us to assess the impact of having a more complex coefficient function relative to the input factor.
2.  $D_X = 1$  and  $D_\beta = 0$ : This contrast helps us understand the effect when the factor's flexibility exceeds that of the coefficient function.

In both sub-scenarios, we vary the number of breakpoints (knots) in the B-spline basis functions for  $X(s)$  to observe how increasing the number of knots of the predictor influences the A-optimality criterion and the efficiency of the design.

### 6.2.1 First-degree B-splines for $X(s)$ and second-degree B-splines for $\beta(s, t)$

In this sub-scenario, we investigate the optimal design when the functional predictor  $X(s)$  is expanded using first-degree B-spline basis functions (piecewise linear functions), and the coefficient function  $\beta(s, t)$  is expanded using second-degree B-spline basis functions (piecewise quadratic functions). This setup allows us to examine the effect of using a basis for the coefficient function that is allowed to change in more points relative to the predictor.

We consider varying the number of breakpoints in the B-spline basis for  $X(s)$  ( $\mathcal{K}_X \in \{5, 9, 15, 19, 29\}$ ), while keeping the number of breakpoints for  $\beta(s, t)$  fixed at three different values as per the previous example ( $\mathcal{K}_\beta \in \{3, 5, 7\}$ ). By analyzing both cases, we can assess how increasing the complexity of the coefficient impacts the efficiency of the design and precision of estimation when combined with predictor flexibility.

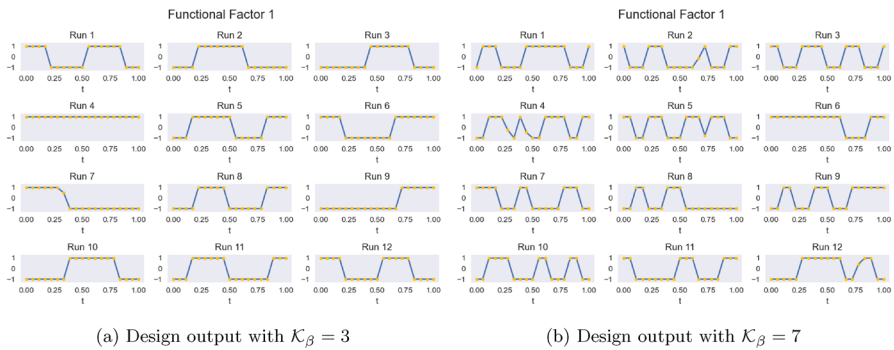
Table 2 summarizes the A-optimality values and relative efficiencies for these settings. The efficiency is calculated relative to the best A-optimality value achieved in each case, which corresponds to the maximum number of breakpoints for  $X(s)$ .

In Fig. 4, we present two designs with a fixed number of breakpoints for  $X(s)$  ( $\mathcal{K}_X = 19$ ), for  $\mathcal{K}_\beta \in \{3, 7\}$ . We can observe that increasing the complexity of the functional parameter significantly impacts the design. The optimal design adapts by utilizing the available knots in  $X(s)$  to generate more varied predictor functions for the more complicated  $\beta(s, t)$ . This is clear when comparing the changes in slope between Figs. 4a, and 4b. Fig. 4a shows much fewer changes in slope for each experimental run because the coefficient function is simpler.

This sub-scenario demonstrates that when the predictor's representation is fixed, increasing the complexity of the coefficient function requires the design to make full use of the available knots to achieve optimal estimation. The designs become more

**Table 2** A-optimality values and relative efficiencies for first-degree B-splines expansion of  $X(s)$  and second-degree B-splines of  $\beta(s, t)$  with varying numbers of breakpoints in  $X(s)$

Breaks	3		5		7	
	A-opt	Efficiency	A-opt	Efficiency	A-opt	Efficiency
5	36.82	0.60	—	—	—	—
9	23.71	0.92	112.84	0.71	352.12	0.60
15	22.49	0.98	84.44	0.95	256.99	0.79
19	22.19	0.99	83.93	0.96	216.76	0.94
29	21.96	1.00	80.30	1.00	207.83	1.00



**Fig. 4** Comparison of design output for different breakpoints when  $K_X = 19$

intricate, with more changes in slope in the predictor functions, reflecting the need to capture the additional complexity in  $\beta(s, t)$ .

**6.2.2 First-degree B-splines for  $X(s)$  and zero-degree B-splines for  $\beta(s, t)$**

In this last sub-scenario, we explore the optimal design when the functional predictor  $X(s)$  is expanded using first-degree B-spline basis functions, while the coefficient function  $\beta(s, t)$  is expanded using zero-degree B-spline basis functions. This setup allows us to examine the effect of using a more flexible basis with respect to the number of breakpoints for the predictor relative to a simpler coefficient function.

We will again vary the number of breakpoints  $K_X$  in the B-spline basis for  $X(s)$  while keeping the number of breakpoints for  $\beta(s, t)$  fixed. We assess how the predictor’s number of knots impacts the efficiency of the design and the estimation precision when the coefficient function is relatively simple.

In Table 3, we again observe the same thing as with the previous sections. The efficiency of the design increases for all cases of  $K_\beta$  when we allow  $X(s)$  to vary between more breakpoints, with diminishing returns.

In scenario 6.1, we assumed the same zero-degree B-spline basis for the expansion of  $\beta(s, t)$  as in this scenario. Hence, the A-optimality values presented in Table 3 are directly comparable to Table 1 from scenario 6.1. The only thing that changed between those two scenarios is the type of basis function chosen for expanding the functional





relatively simple coefficient function  $\beta_1(s, t)$ . In contrast, the second-degree B-spline expansion for  $\mathbf{x}_2(s)$  provides the enough complexity to capture the intricacy of  $\beta_2(s, t)$  which is expanded using first-degree B-splines.

Introducing multiple factors into a model increases the number of parameters that need to be estimated. In our experiments, we found that the coordinate exchange algorithm struggles to find an optimal design when the number of parameters is high. To address this issue, we can either expand the experiment by increasing the number of experimental runs or employ an alternative optimization algorithm.

In conclusion, the case of multiple functional predictors demonstrates that optimal experimental design can account for different complexities of the basis functions of each predictor and their associated coefficient functions. This example highlights that it is possible to accommodate the individual properties of each functional predictor and their associated coefficient functions in the design of experiments on the fully functional model.

## 7 Proof-of-concept example

We develop a proof-of-concept example of our methodology from the continuous manufacturing line setting (Van Snick et al. 2019) described in Sect. 2. In each experimental run, a mixture is dosed into a gravimetric feeder (input), and the outlet concentration is monitored in-line downstream (output). Because of transport delays and mixing, sharp changes in the dosing program produce smoothed responses after a temporal lag. This naturally yields a function-on-function regression problem, where both factors and responses are trajectories on the normalized interval  $[0, 1]$ .

### Simulation setup and data generation

To generate data, we specify a bivariate coefficient surface  $\beta(s, t)$  on  $[0, 1]^2$  and simulate responses

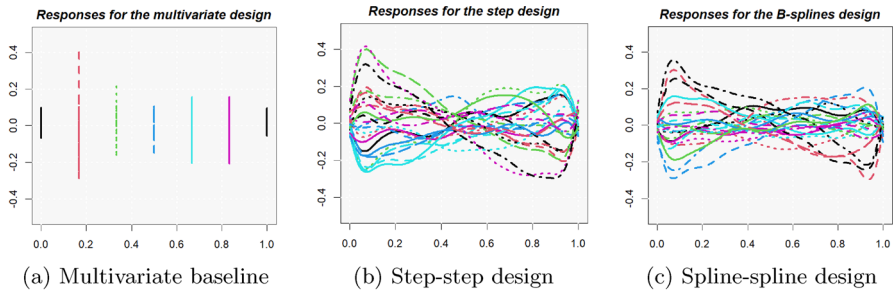
$$Y_j(t) = \int_0^1 X_j(s) \beta(s, t) ds + \varepsilon_j(t), \quad j = 1, \dots, n,$$

where  $X_j(\cdot)$  is the dosing profile of run  $j$  and  $\varepsilon_j(\cdot)$  is a mean-zero functional noise process. The *true* coefficient is represented on a B-spline basis with  $4 \times 4$  cubic splines and random coefficients. The error is a Gaussian process with radial-basis-function kernel and small variance ( $\gamma = 10^{-3}$ ,  $\sigma^2 = 10^{-2}$ ), represented on a Fourier basis with 81 terms. The choice of a rich, smooth basis for the noise ensures realistic smoothness and Dispersion of the response curves while keeping the signal-to-noise ratio moderate.

### Designs under comparison.

We compare three different designs, each one with  $n = 24$  runs (the exact designs are provided in the Appendix and visualized there):

1. **Multivariate baseline (naïve approach):** we discretize the dosing profiles at 10 equally spaced segments and 7 different responses and apply a classical multivariate A-optimal design on the resulting 10-dimensional factor space. This design



**Fig. 7** Simulated response curves for the three 24-run designs

was generated using the software JMP (SAS 2024), treating the discretized values of  $X(\cdot)$  as independent factors. This corresponds to approximating the functional input by a finite grid and designing the experiment as a standard multivariate linear model.

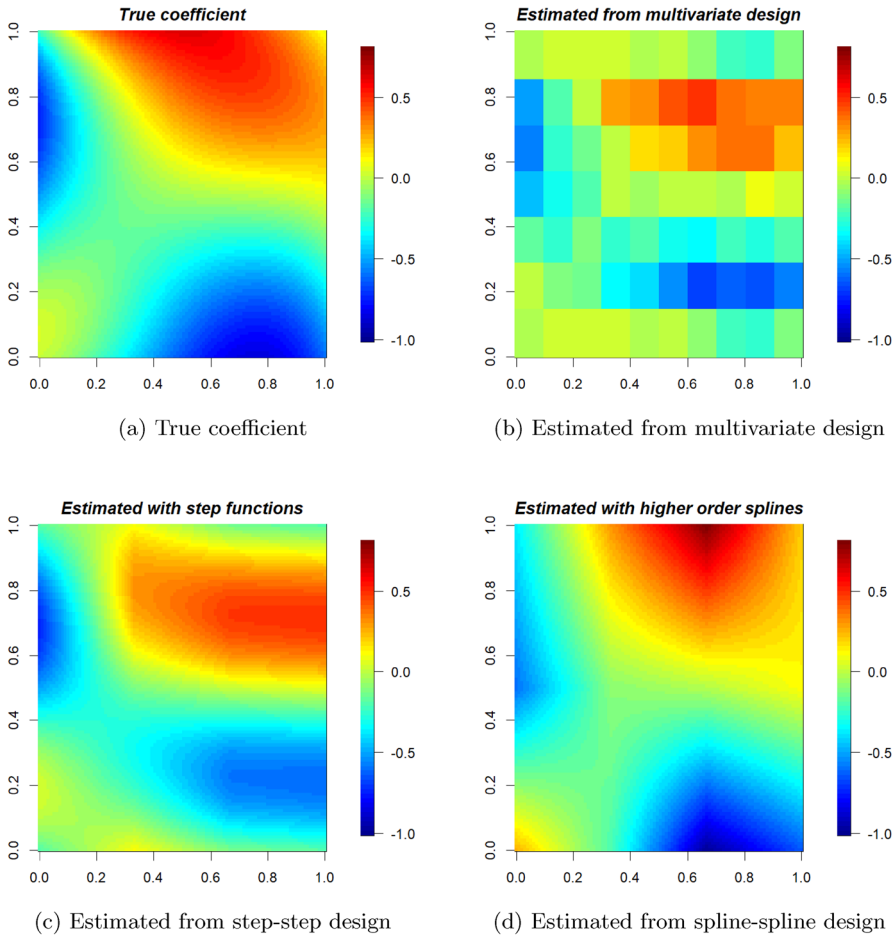
2. **Step–step (basis-aligned to physical actuation)**: the dosing trajectories are represented on a degree-0 B-spline basis with 11 knots, yielding 10 stepwise segments over  $s \in [0, 1]$ . The coefficient surface used in the design criterion is expressed on the same degree-0 B-spline basis (11 knots), and the response direction is represented on a 7-term Fourier basis. This setup produces A-optimal designs whose runs correspond to stepwise input profiles consistent with realistic feeder actuation.
3. **Spline–spline (smooth dosing)**: the dosing trajectories are represented using degree-2 B-splines with 9 knots, yielding 8 interior segments and smoother functional inputs. In the design criterion, the coefficient surface is expressed using degree-3 B-splines with 5 knots in the factor direction, while the response direction is again represented on a 7-term Fourier basis. This basis choice yields A-optimal designs whose runs correspond to smooth dosing programs.

The first design serves as a baseline obtained by discretization and classical multivariate optimal design. The other two are purely functional A-optimal designs constructed directly in the appropriate functional basis, reflecting either stepwise or smooth actuation of the dosing mechanism.

Figure 7 displays typical responses produced by the three designs in a single simulated experiment. The multivariate baseline produces responses that are less diverse and exhibit—by design—blocking at the discretization points, whereas both basis-driven designs (step and B-spline) generate trajectories that explore the response space more effectively and more smoothly, better matching the expected physics of the process.

The corresponding coefficient estimates are shown in Fig. 8. As expected, the multivariate fit yields a blocky coefficient estimate that misses smooth transitions, while the functional designs recover the main features of the true surface. Among them, the estimates produced by the spline–spline design are visually the closest to the true coefficient.

**Simulation study (RMSE).** To quantify the performances of the different designs, we repeated the simulation 100 times using the same true  $\beta(s, t)$  and error process and



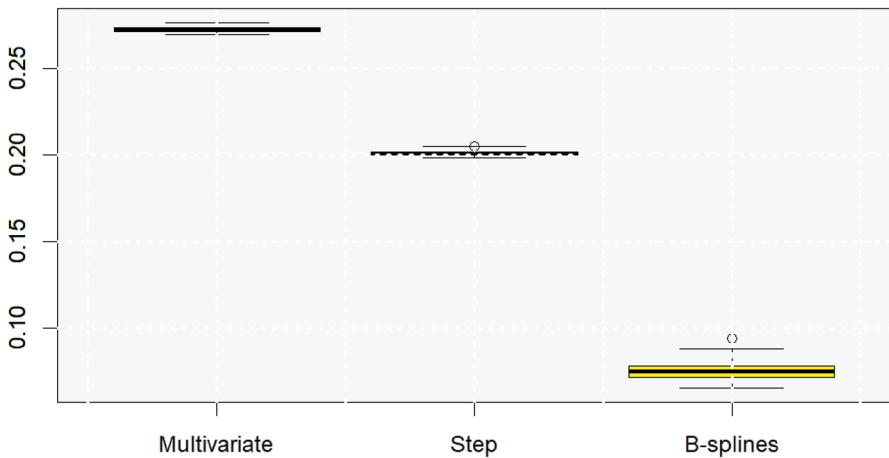
**Fig. 8** True coefficient and estimated coefficient surfaces under the three designs

computed the RMSE on a  $100 \times 100$  grid:

$$\text{RMSE} = \left\{ \frac{1}{10^4} \sum_{g,h} (\hat{\beta}(s_g, t_h) - \beta(s_g, t_h))^2 \right\}^{1/2}.$$

Figure 9 summarizes the distribution of RMSE across replicates. The spline–spline design attains the smallest errors, the step–step design is second, and the multivariate baseline is the worst. This ranking is consistent across repetitions and confirms that designing directly in function space yields more informative experiments than the naïve “discretize-then-design” approach.

These results demonstrate that functional A-optimal designs tailored to the structure of the factor and response spaces provide substantial gains over naïve, discretization-based, strategies. When the design space is expressed in an appropriate functional basis,



**Fig. 9** RMSE over 100 simulated experiments. Left to right: multivariate baseline, step–step, spline–spline

capturing either the stepwise actuation of the process or its expected smoothness, the resulting experiments yield more informative response trajectories and hence, more accurate estimates of the coefficient surface.

## 8 Conclusion

This work presents new results on optimal experimental design for estimating the functional coefficient of a linear model with multiple dynamic factors and where also the response is functional. The first part of the paper develops inference results and extends optimality criteria for selecting designs of experiments, proving that both A-optimal and D-optimality do not depend on the covariance of the error process.

The second part of the paper is dedicated to illustrating the proposed methodology in concrete examples. In order to construct an experimental design in this context, an experimenter should set in advance the set of basis functions (e.g., type, order, dimension, etc.) to be used for the predictors, and the set of basis functions to expand the functional coefficients in the direction of the predictors. As we have shown, the functional form of the optimal design is affected by this choice.

We describe then a real life example in an industrial scenario, and we apply our method with a simulation study. Simulations show that functional A-optimal designs provide substantial gains over naïve, discretization-based, strategies by yielding more informative response trajectories and hence, more accurate estimates of the coefficient surface.

The function-on-function model considered in this paper could be straightforwardly extended by including also non-functional factors and concurrent functional interactions. The extension to time-delayed interactions is more challenging, since the model includes functional coefficients in higher dimensions. This will be scope for future work. Moreover, in some cases it can be of interest to represent the functional coef-

ficients using a higher number of basis functions and this can make the model not identifiable with the available number of runs. In this case, we need to include a penalization in the estimator, usually imposing a penalty on the derivatives of the functional coefficient. A further development will be to derive the expression of the optimal designs for this different estimator. Finally, other optimality design criteria, particularly those focusing on precise prediction, will also represent an interesting direction for additional development of the present work.

## Supplementary information

- An exhaustive database of the experimental designs generated in this paper is available online.<sup>1</sup> The dataset consists of 576 records, each representing one experimental design with 12 columns detailing the settings and results. The columns include the design ID (id); constants such as the number of random starts of the coordinate exchange algorithm (epochs) and the number of experimental runs (Runs); choices made for the functional factor—basis type (X family), degree (X degree), and number of breakpoints (X breaks); and similar choices for the functional coefficient (B family, B degree, B breaks). The results comprise the design coefficients of the basis expansion of the functional factor, the A-optimality criterion value, and the figure of the design. If a design cannot be estimated for specific settings, the design column will have the value “NONE,” the criterion column will be null, and the figure column will contain no image.
- The three optimal designs constructed for Sect. 7 are showed in the Supplementary Material. This contains a table with the optimal multivariate design, a figure with the step function optimal design and a figure with the B-splines optimal design.

## Appendix A Supplementary Material

We report the exact experimental designs used in the realistic proof-of-concept example discussed in Sect. 7 of the main paper. We provide (i) the 24-run multivariate A-optimal design used for the naïve discretize-then-design baseline and (ii) graphical representations of the corresponding 24-run functional A-optimal designs in the step-step and spline-spline settings. This material is intended to facilitate replication and to give additional insight into the structure of the proposed functional designs.

### A.1 Multivariate A-optimal design

Table 5 reports the 24-run multivariate A-optimal design (SAS 2024) employed for the baseline approach in Sect. 7 of the main paper. Each factor represents the value of the discretized input trajectory  $X(\cdot)$  at one of ten equally spaced points on  $[0, 1]$ , coded at levels  $-1$  and  $1$ . The resulting design is obtained by applying standard A-optimal design methodology to this 10-dimensional multivariate regression setting.

<sup>1</sup> <https://airtable.com/appX72fWQTQ081Ykg/shrfeUixNUS8kR2iV/tbILl0ckw7EVoaaj7j>.

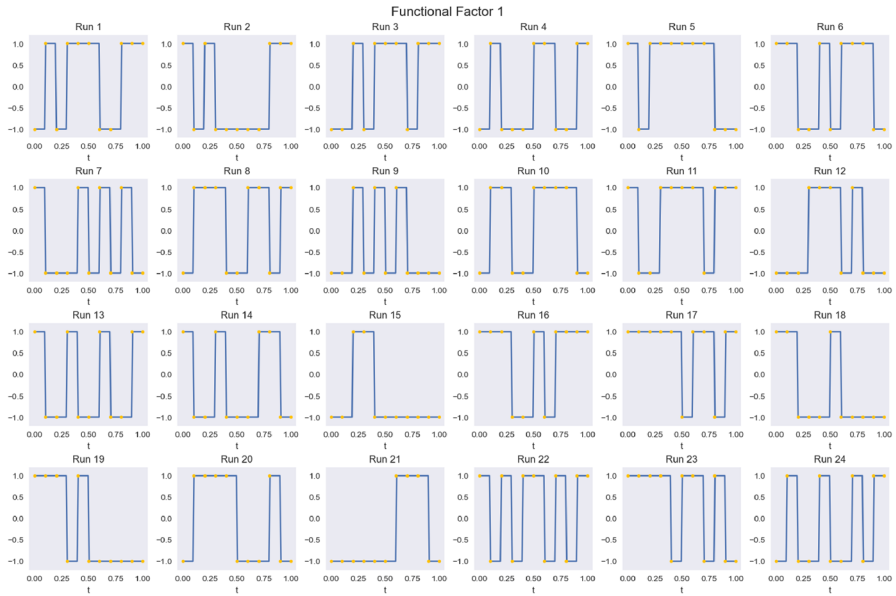
**Table 5** Twenty-four-run A-optimal multivariate design used in the proof-of-concept study

Run	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1	1	-1	1	-1	-1
3	1	1	-1	-1	-1	-1	1	-1	1	-1
4	1	1	1	-1	-1	-1	-1	1	-1	1
5	1	1	1	1	-1	-1	-1	-1	1	-1
6	-1	1	1	1	1	-1	-1	-1	-1	1
7	-1	-1	1	1	1	1	-1	-1	-1	-1
8	-1	-1	-1	1	1	1	1	-1	-1	-1
9	-1	-1	-1	-1	1	1	1	1	-1	-1
10	-1	-1	-1	-1	-1	1	1	1	1	-1
11	-1	-1	-1	-1	-1	-1	1	1	1	1
12	1	-1	-1	-1	-1	-1	-1	1	1	1
13	1	1	-1	-1	-1	-1	-1	-1	1	1
14	1	1	1	-1	-1	-1	-1	-1	-1	1
15	1	1	1	1	-1	-1	-1	-1	-1	-1
16	-1	1	1	1	1	-1	-1	-1	-1	-1
17	-1	-1	1	1	1	1	-1	-1	-1	-1
18	-1	-1	-1	1	1	1	1	-1	-1	-1
19	-1	-1	-1	-1	1	1	1	1	-1	-1
20	-1	-1	-1	-1	-1	1	1	1	1	-1
21	-1	-1	-1	-1	-1	-1	1	1	1	1
22	1	-1	-1	-1	-1	-1	-1	1	1	1
23	1	1	-1	-1	-1	-1	-1	-1	1	1
24	1	1	1	-1	-1	-1	-1	-1	-1	1

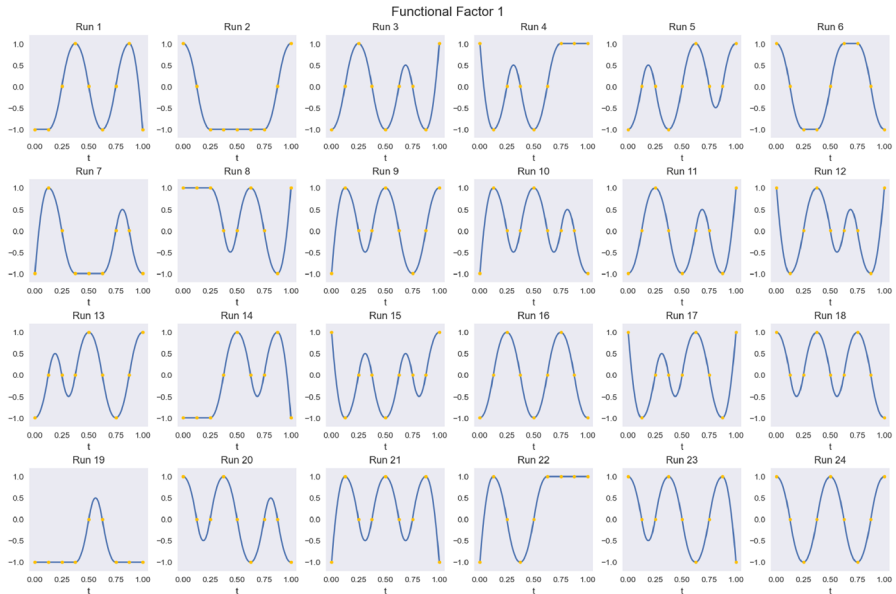
Each row corresponds to one experimental run, and each column  $X_k$  to the coded level of the factor at the  $k$ -th discretization point in  $s$

## A.2 Functional A-optimal designs

Figures 10 and 11 display the two functional A-optimal designs considered in the main paper, namely the step–step and spline–spline designs. In both cases, the input factor is a trajectory  $X(s)$  on  $[0, 1]$ , corresponding to the dosing program applied to the gravimetric feeder. Each curve in the figures represents one of the 24 runs in the corresponding design.



**Fig. 10** Twenty-four-run step–step A-optimal design. Each curve represents one dosing trajectory constructed using a step-function basis with ten segments over the normalized time domain  $s \in [0, 1]$



**Fig. 11** Twenty-four-run spline–spline A-optimal design. Each trajectory is represented using second-order B-splines with ten segments over  $s \in [0, 1]$

### A.2.1 Step–step design

In the step–step design, the input trajectories are constructed using a step-function basis with ten segments, reflecting a setting where the feeder operates at a finite number of discrete levels over successive time intervals. This basis aligns closely with realistic actuation patterns (on/off or few-level changes over time) and with the stepwise representation used when computing the A-optimal design in the factor space.

### A.2.2 Spline–spline design

In the spline–spline design, the input trajectories are represented using second-order B-splines with ten segments, yielding smoother dosing profiles. This setting is appropriate when the actuation can vary more gradually over time, and it complements the use of smooth B-spline bases for both the factor and response directions in the estimation of the coefficient surface  $\beta(s, t)$ .

**Acknowledgements** This work was supported by the Engineering and Physical Sciences Research Council, UK [grant number EP/T021624/1, *Multi-objective optimal design of experiments*]. The second author's research has been supported by the EPSRC project 2607359. The first author was supported also by the Italian national project PRIN 2022 [grant number 2022TRB44L]. The authors are grateful to the Associate Editor and the reviewers for their valuable comments and suggestions, which helped improve this paper. They would also like to thank Martin Otava for useful feedback regarding the motivating example.

**Funding** Open access funding provided by Università degli Studi del Piemonte Orientale Amedeo Avogadro within the CRUI-CARE Agreement.

**Data Availability** All the data used in the study were artificially created through simulations, and no real-world data were collected from experiments or observations. The optimal experimental designs are available via the supplementary material and the codes are available on request.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aletti G, May C, Tommasi C (2014) Optimal designs for linear models with functional responses. In: Bongiorno E, Salinelli E, Goia A, Vieu P (eds) Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna, pp 19–24
- Aletti G, May C, Tommasi C (2016a) Best estimation of functional linear models. [arXiv:1412.7332](https://arxiv.org/abs/1412.7332)
- Aletti G, May C, Tommasi C (2016) On applying optimal design of experiments when functional observations occur. In: Kunert J, Müller CH, Atkinson AC (eds) mODa 11: advances in model-oriented design and analysis. Springer, Cham, pp 1–9
- Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs, with SAS. Oxford statistical science series, vol 34. Oxford University Press, Oxford
- Blanco M, Coello J, Iturriaga H, Maspoch S, De La Pezuela C (1998) Near-infrared spectroscopy in the pharmaceutical industry. *Anal-Lond-Soc Public Anal R Soc Chem* 123:135R–150R

- Chen J, Mak S, Joseph VR, Zhang C (2021) Function-on-function kriging, with applications to three-dimensional printing of aortic tissues. *Technometrics* 63(3):384–395. <https://doi.org/10.1080/00401706.2020.1801255>
- Fedorov V (1972) *Theory of optimal experiments*. Cellular neurobiology. Academic Press, New York
- Fedorov V, Hackl P (1997) *Model-oriented design of experiments*. In: *Lecture notes in statistics*. Springer, New York
- Gertheiss J, Rügamer D, Liew BX, Greven S (2024) Functional data analysis: an introduction and recent developments. *Biom J* 66(7):e202300363
- Horváth L, Kokoszka P (2012) *Inference for functional data with applications*. Springer, New York
- Kao MH, Huang PH (2024) Hybrid exact-approximate design approach for sparse functional data. *Comput Stat Data Anal* 190:107850. <https://doi.org/10.1016/j.csda.2023.107850>
- Kokoszka P, Reimherr M (2017) *Introduction to functional data analysis*. Chapman and Hall/CRC, Boca Raton
- Magnus JR, Neudecker H (2019) *Matrix differential calculus with applications in statistics and econometrics*. Wiley series in probability and statistics, 3rd edn. Wiley, Oxford
- Malfait N, Ramsay JO (2003) The historical functional linear model. *Can J Stat* 31(2):115–128
- Meyer RK, Nachtsheim CJ (1995) The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37(1):60–69
- Michaelides D, Adamou M, Woods DC, Overstall AM (2026) Optimal designs for experiments for scalar-on-function linear models. *Biometrics* 82(1):169
- Morris MD (2014) Maximin distance optimal designs for computer experiments with time-varying inputs and outputs. *J Stat Plan Inference* 144:63–68
- Müller HG, Chiou JM, Leng X (2008) Inferring gene expression dynamics via functional regression analysis. *BMC Bioinform* 9:1–20
- Pigoli D, Ferraty F, Aston JAD, Mazumder A, Richards C, Hall MJR (2023) 03. Estimation of temperature-dependent growth profiles for the assessment of time of hatching in forensic entomology. *J R Stat Soc Ser C: Appl Stat* 72(2):231–253. <https://doi.org/10.1093/jrssc/qlac003>
- Pukelsheim F (2006) *Optimal design of experiments*. Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia
- Rahman R, Dhruva SR, Ghosh S, Pal R (2019) Functional random forest with applications in dose-response predictions. *Sci Rep* 9(1):1628
- Ramsay JO (1988) Monotone regression splines in action. *Stat Sci* 3(4):425–441
- Ramsay JO, Dalzell C (1991) Some tools for functional data analysis. *J R Stat Soc Ser B Stat Methodol* 53(3):539–561
- Ramsay JO, Silverman BW (2005) *Functional data analysis*. Springer, New York
- Rha H, Kao MH, Pan R (2020) Design optimal sampling plans for functional regression models. *Comput Stat Data Anal* 146:106925. <https://doi.org/10.1016/j.csda.2020.106925>
- SAS Institute Inc (2024) JMP®, Version 18. Cary, NC, pp 1989–2024
- Shi J, Wang B, Will E, West R (2012) Mixed-effects gaussian process functional regression models with application to dose-response curve prediction. *Stat Med* 31(26):3165–3177
- Silvey S (1980) *Optimal design: an introduction to the theory for parameter estimation*. Chapman and Hall, London
- Smith PW (1980) A practical guide to splines (carl de boor). *SIAM Rev* 22(4):520
- Van Rossum G, Drake FL (2009) *Python 3 reference manual*. CreateSpace, Scotts Valley
- Van Snick B, Kumar A, Verstraeten M, Pandelaere K, Dhondt J, Di Pretoro G, De Beer T, Vervae C, Vanhoorne V (2019) Impact of material properties and process variables on the residence time distribution in twin screw feeding equipment. *Int J Pharm* 556:200–216. <https://doi.org/10.1016/j.ijpharm.2018.11.076>
- Zhang X, Chiou JM, Ma Y (2018) Functional prediction through averaging estimated functional linear regression models. *Biometrika* 105(4):945–962