Optimal design of experiments for dynamic factors and functional responses

Caterina May

Università del Piemonte Orientale

Joint work with: Theodoros Ladas, Davide Pigoli, Kalliopi Mylona – King's College London

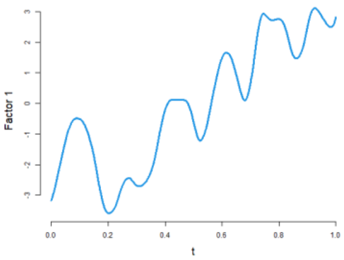
mODa14, June 15, 2026 - Drongen Abbey, Belgium

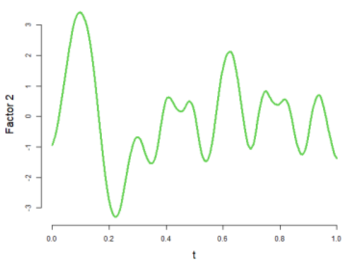
Outline

- Dynamic factors and functional linear models
- Optimal design of experiments for the reduced rank estimator of a fully-functional model
- Numerical examples (with B-splines)
- Proof-of-concept example and simulation
- Conclusions and further developments

Dynamic factors and functional linear models

Dynamic factors

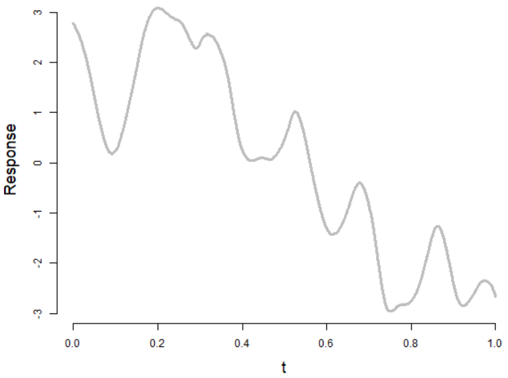




Numerical variables

Categorical variables

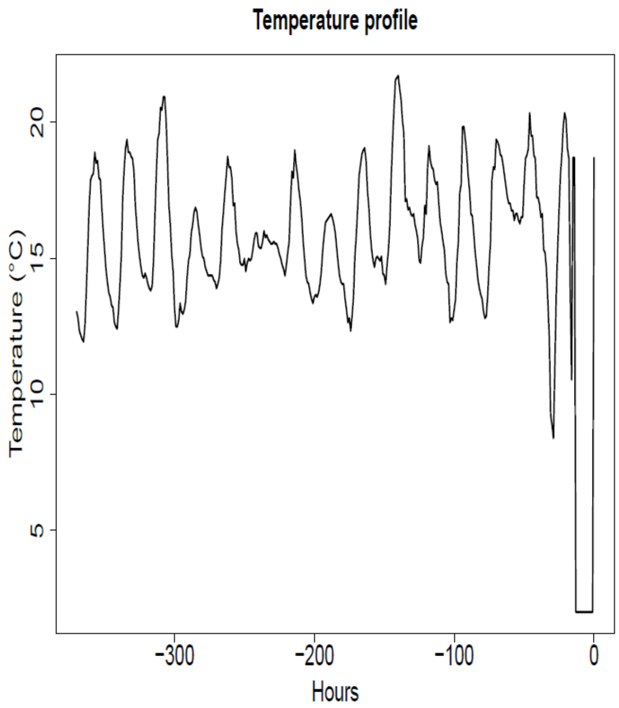
Statistical model



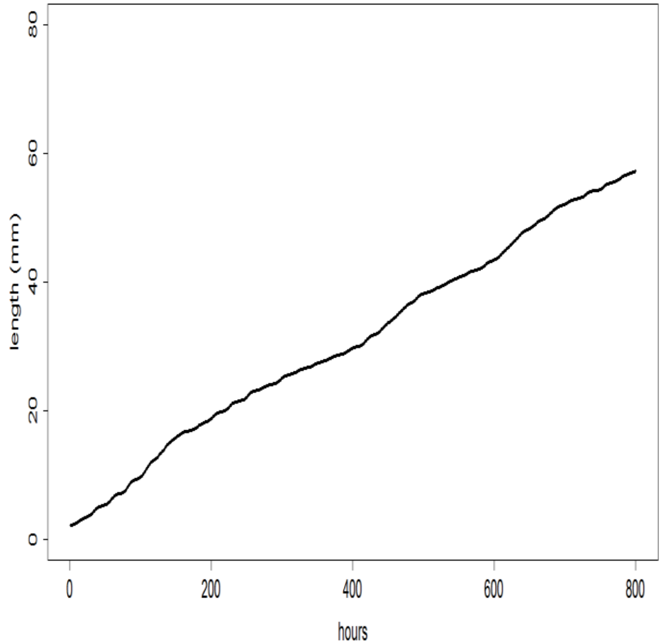
Motivating example

In entomology, there is an interest in the association between temperature curves and insects' growth curves

Dynamic factor = Temperature



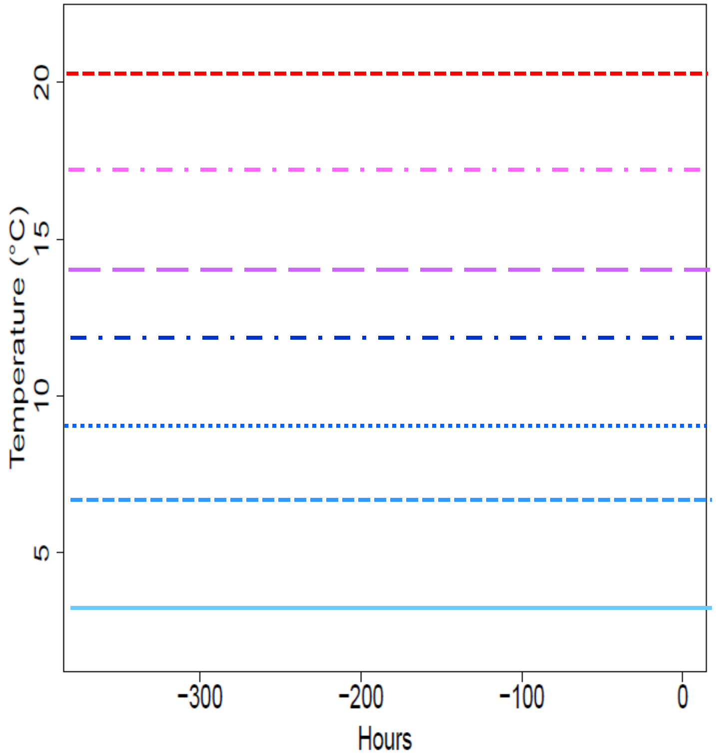
Response = Growth curve

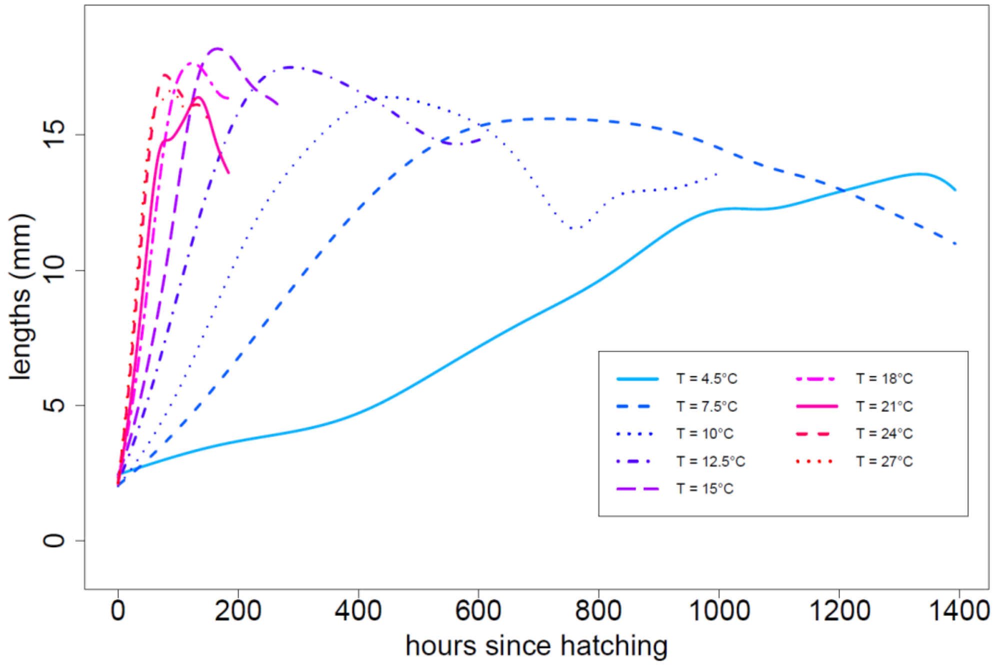


Motivating example

This is studied with experiments in incubators, this is an example from the Natural History Museum Lab in London:

Is there a more efficient way to design this experiment?





UNIVERSITÀ DEL PIEMONTE ORIENTALE

Function-on-function linear models

$$y_n(t) = \mu(t) + \sum_{i=1}^{p} f(\beta_i, x_{ni}) + \varepsilon_n(t)$$

Independent error processes

$$f(\beta, \mathrm{x}) = \begin{cases} \beta x(t) \\ \beta(t)x(t) \\ \int_0^{\mathrm{T}} \beta(s,t)x(s)\mathrm{d}s \end{cases}$$

$\beta x(t)$ — Proportional linear model

$\beta(t)x(t)$ — Concurrent linear model

$\int_0^{\mathrm{T}} \beta(s,t)x(s)\mathrm{d}s$ — Non concurrent linear model (fully functional) (historical if T= t)

Function-on-function linear model

GOAL: Estimation of the parameters in a fully function linear model:

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

where:

$$\mathbf{y}(t) = (y_1(t), \ldots, y_N(t))^T$$

$$\boldsymbol{\beta}(s,t) = (\beta_0(t), \beta_1(s,t), \ldots, \beta_p(s,t))^T$$

$$\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_N(t))^T$$

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

Design matrix

$$\mathcal{X}(s) = [\mathbf{1}_N \,|\, \mathrm{X}(s)]$$

Model matrix

$$\mathbf{1}_N = (1, \ldots, 1)^T$$

Estimation

Finite *basis expansion:*

Let us consider some sets of bases of functions:

Orthonormal basis

$$\{\theta_l(t), l \geq 1\} \ , \{\eta_k^1(s), k \geq 1\}, \ \ldots, \ \{\eta_k^p(s), k \geq 1\}$$

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

and let $\mu(t) = \sum_l b_l^0 \, \theta_l(t) = \mathbf{b}^{0^T} \boldsymbol{\theta}(t)$

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

and $\beta_i(s,t) = \sum_{k,l} b_{k,l}^i \, \eta_k^i(s) \, \theta_l(t) = \boldsymbol{\eta}^i(s)^T \mathrm{B}^i \, \boldsymbol{\theta}(t)$

Separable representation

Estimation

Finite *basis expansion* of the fully linear model:

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

$$\mathrm{H}(s)^T = \mathrm{diag}(1, \boldsymbol{\eta}^1(s)^T, \ldots, \boldsymbol{\eta}^p(s)^T)$$

$$\mathbf{y}(t) = \int \mathcal{X}(s)\mathrm{H}(s)^T\,\mathrm{B}\,\boldsymbol{\theta}(t)\,ds + \boldsymbol{\varepsilon}(t)$$

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

$$\mathrm{Z} = \int \mathcal{X}(s)\mathrm{H}(s)^T\,ds$$

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

Two possible approaches for estimation

The projection of the original model on these basis functions is:

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

Now we have two possible approaches for estimation of the matrix of coefficients B:

1. Choosing $K \ll N$, so that the resulting linear model can be estimated by Integrated SSE (*Reduced rank estimation*)

2. Choosing $K \approx N$ and then include some regularisation on the coefficients (*Penalised estimation*)

C. May et al. caterina.may@uniupo.it

Estimation of the parameters (reduced rank)

Estimation of the parameters in a fully function linear model: *reduced rank estimator* (minimizing ISSE)

$$\hat{\mathrm{B}} = \arg\min_{\mathrm{B}} \int \|\mathbf{y}(t) - \mathrm{Z}\,\mathrm{B}\,\boldsymbol{\theta}(t)\|^2 dt$$

Proposition (after Ramsay and Silverman, 2005): If Z is full rank and $\{\theta_l(t)\}$ is an orthonormal basis, then

$$\boxed{\hat{\mathrm{B}} = (\mathrm{Z}^T\mathrm{Z})^{-1}\mathrm{Z}^T \int \mathbf{y}(t)\,\boldsymbol{\theta}(t)^T\,dt}$$

Penalized estimation of the parameters

Estimation of the parameters in a fully function linear model: *penalised estimator* (minimizing penalized ISSE)

$$\mathrm{vec}\,\hat{\mathrm{B}}_\lambda = \mathrm{F}_\lambda^{-1}(I_L \otimes \mathrm{Z}^T)\,\mathrm{vec}\left(\int \mathbf{y}(t)\,\boldsymbol{\theta}(t)^T dt\right)$$

$$\mathrm{F}_\lambda = \mathrm{I}_L \otimes (\mathrm{Z}^T\mathrm{Z}) + \lambda_s(\mathrm{I}_L \otimes \mathrm{R}) + \lambda_t(\mathrm{S} \otimes \mathrm{J}_{HH})$$

$$\mathrm{S} = \int L_t\boldsymbol{\theta}(t)\,L_t\boldsymbol{\theta}(t)^T\,dt \qquad \mathrm{R} = \int L_s\mathrm{H}(s)\,L_s\mathrm{H}(s)^T\,ds$$

$$\mathrm{J}_{HH} = \int \mathrm{H}(s)\,\mathrm{H}(s)^T\,ds$$

Optimal design of experiments for the reduced rank estimator

C. May et al. caterina.may@uniupo.it

Optimal DoE for the reduced rank estimator

Result (May et al.):
If Z has maximum rank and $\{\theta_l(t)\}$ is an orthonormal basis, then

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

Vectorisation of the matrix of the coefficients' estimators, obtained by stacking the columns

Matrix of the covariances of the projections of the error on the basis functions, i.e.
$\Sigma_{lq} = \mathrm{cov}(\int \varepsilon(t)\theta_l(t)\mathrm{d}t\,, \int \varepsilon(t)\theta_q(t)\mathrm{d}t\,)$

$$\boxed{\mathrm{Z} = \int \mathcal{X}(s)\mathrm{H}(s)^T\,ds}$$

Optimal DoE for the reduced rank estimator

Result (May et al.):
If Z has maximum rank and $\{\theta_l(t)\}$ is an orthonormal basis, then

[A-optimality] The A-optimal design $X_A^*(s)$ satisfies

$$X_A^*(s) = \arg\min_{X(s)} \mathrm{tr}(\mathrm{Z}^T\mathrm{Z})^{-1}$$

[D-optimality] The D-optimal design $X_A^*(s)$ satisfies

$$X_D^*(s) = \arg\min_{X(s)} \frac{1}{\det \mathrm{Z}^T\mathrm{Z}}$$

A few remarks

Optimal DoE for the reduced rank estimator:

- The previous expressions are still based on the functional expression of the factors. In practice we need to use some bases of functions to express them, and **the optimisation will be w.r.t the basis coefficients**.

- We didn't put any requirements on this basis functions: the range of admissible functional shapes can be freely chosen based on the experimental constraints.

- We required the basis $\{\theta_l(t)\}$ to be orthonormal, but we don't need its expression to obtain the design.

UNIVERSITÀ DEL PIEMONTE ORIENTALE

Numerical examples

Numerical examples

When we express the factors in basis of functions we have:

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

Hence, the matrix Z can be rewritten as

$$\mathrm{Z} = \Gamma \int \mathrm{C}(s)\, \mathrm{H}(s)^T\, \mathrm{d}s = \Gamma\, \mathrm{J}_{CH}$$

- In the examples we use **various B-splines** for C(s) and H(s), that is:

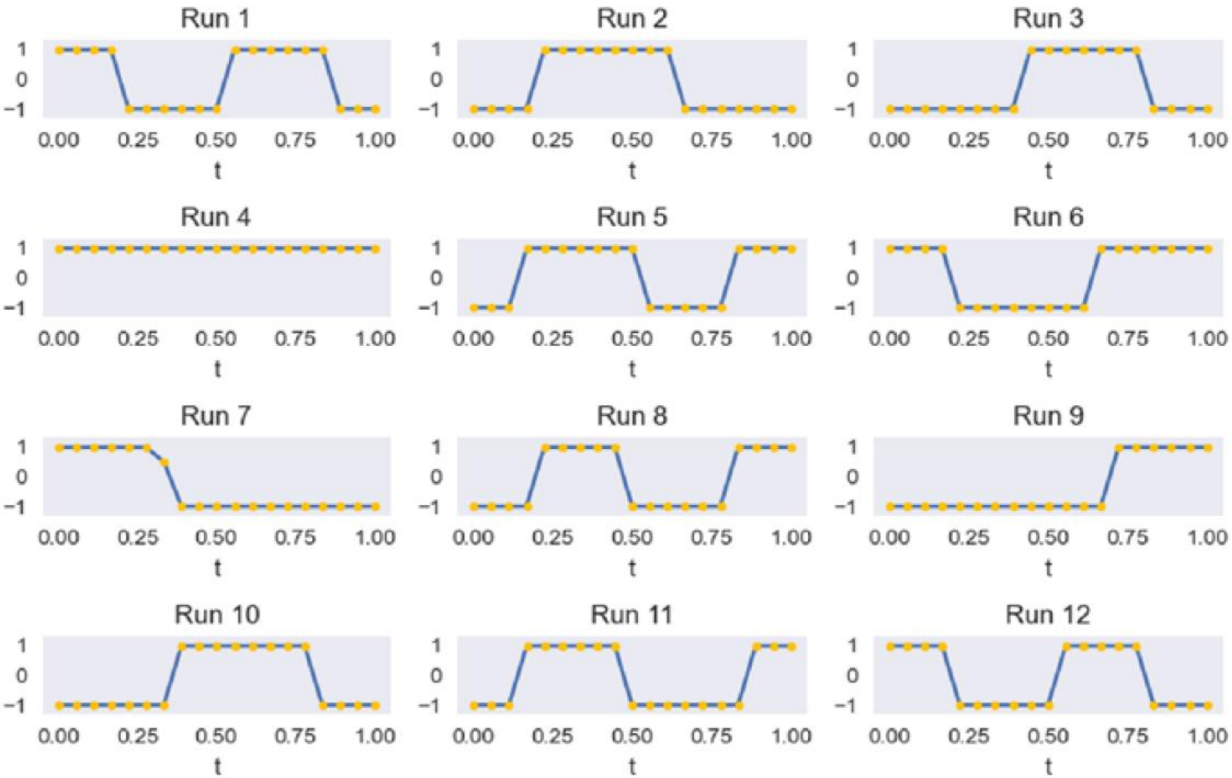
we expand X(s) and the functional coefficients $\beta(s, t)$ in the direction of s with B-splines of various degrees and various number of breakpoints K_X and K_β.

- Exact optimal designs are computed with a coordinate-exchange algorithm implemented in Python 3.10, using 1000 random starts and 100 refinement iterations.

Numerical example (A-optimality, N=12 runs)

The **Figure** displays the A-optimal design in the following scenario:

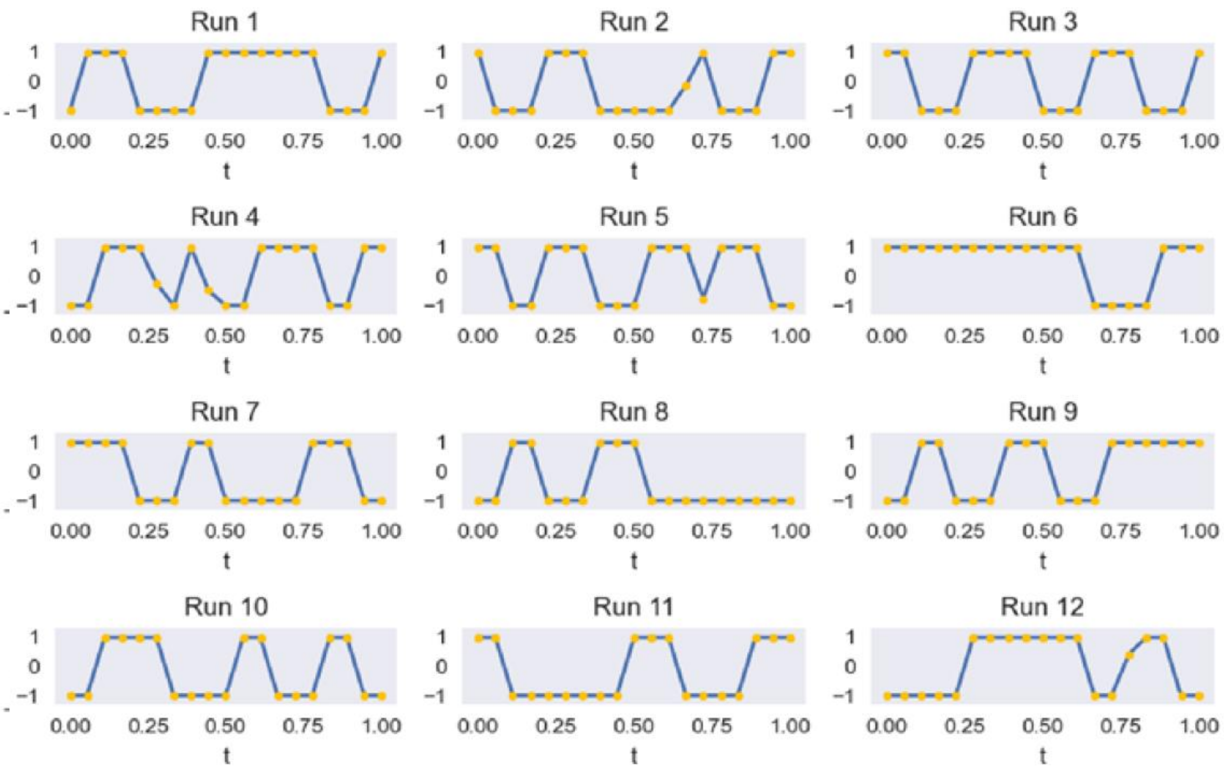
- First-degree B-spline basis to expand a single factor X(s), $K_X = 19$
- Second-degree B-spline basis to expand $\beta(s,t)$ in direction of *s*, $K_\beta = 3$



Numerical example (A-optimality, N=12 runs)

The **Figure** displays the A-optimal design in the following scenario:

- First-degree B-spline basis to expand a single factor X(s), $K_X = 19$
- Second-degree B-spline basis to expand $\beta(s,t)$ in direction of s, $K_\beta = 7$



Numerical examples

The **Table** shows the A-optimality values and the relative efficiencies (ratio between total variances) in the same scenario:

- First-degree B-spline basis to expand X(s);
- Second-degree B-spline basis to expand β(s,t) in direction of *s*.

The comparison is done with {5, 9, 15, 19, 29} breakpoints for X(s) and {3, 5, 7} breakpoints for β(s,t).

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

Insights

- A-optimal and D-optimal designs do not depend on the unknown covariance matrix Σ. This is crucial since we can generate optimal designs without the information from the data that will be available after the experiment.

- The optimal design is affected by the bases chosen to represent the factor and the functional coefficient (in the direction of s). It possible to compare designs based on different basis functions for the factor to see which is more efficient, given a basis for the functional coefficient.

- Increasing the complexity of the coefficient impacts the efficiency of the design and the precision of the estimation when combined with predictor flexibility.

Proof-of-concept example

Example: Pharmaceutical Manufacturing Process

Proof-of-concept example inspired by a real-world continuous pharmaceutical manufacturing line using a gravimetric feeder.

Dynamic System:

- **Input factor**: a mixture dosed into a gravimetric feeder $X_j(s)$
- **Output factor**: the outlet concentration monitored in-line downstream $Y_j(t)$

Due to physical transport delays and mixing inside the line, input changes propagate smoothly into the output after a time lag, creating a function-on-function regression problem:

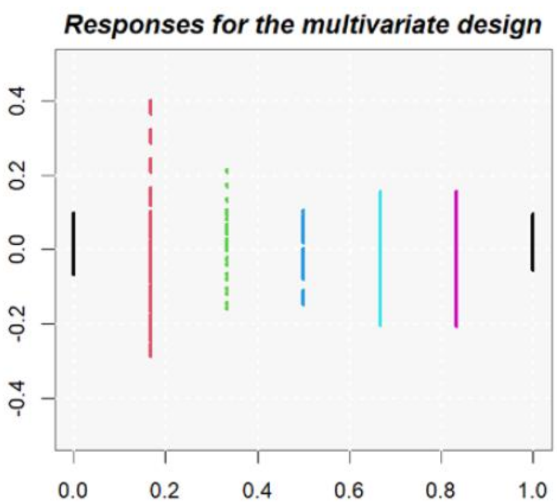
$$Y_j(t) = \int_0^1 X_j(s)\, \beta(s,t)\, ds + \varepsilon_j(t), \qquad j = 1, \dots, n,$$

Data generation:

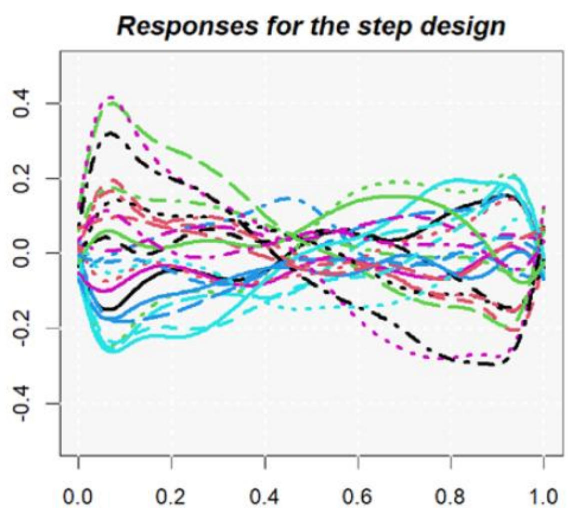
- The true coefficient is represented on a B-spline basis with cubic splines and random coefficients.
- The error is a Gaussian process with radial-basis-function kernel and small variance represented on a Fourier basis with 81 terms.
- We compare **three designs**.

Simulated responses (24-run designs)

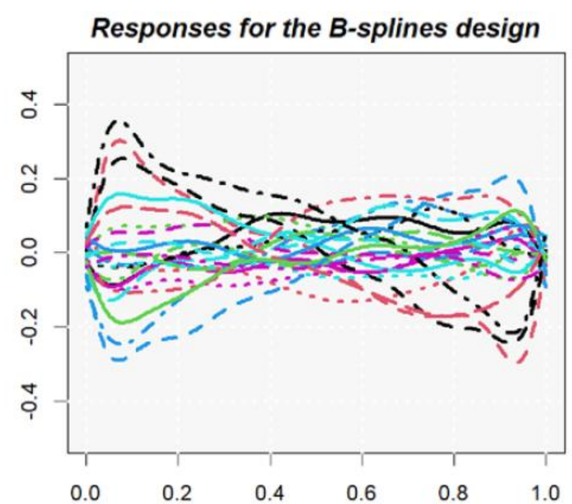
a. **Multivariate baseline (naïve approach):** we discretize the dosing profiles and apply a classical multivariate A-optimal design.
b. **Step–step (basis-aligned to physical actuation):** the dosing trajectories are represented on a degree-0 B-spline. The coefficient surface is expressed on the same degree-0 B-spline basis, and the response is represented on a 7-term Fourier basis.
c. **Spline–spline (smooth dosing):** the dosing trajectories are represented on a degree-2 B-splines. The coefficient surface is expressed using degree-3 B-splines.



(a) Multivariate baseline



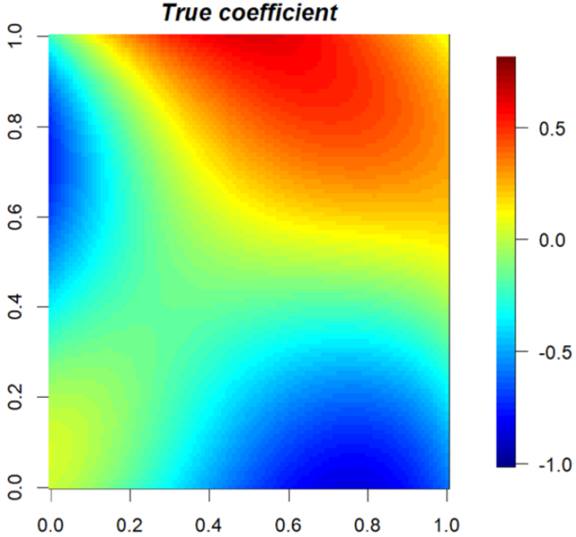
(b) Step-step design



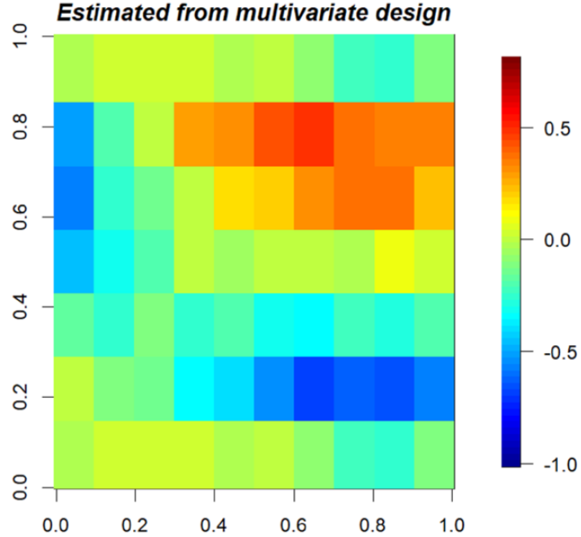
(c) Spline-spline design

C. May et al. caterina.may@uniupo.it

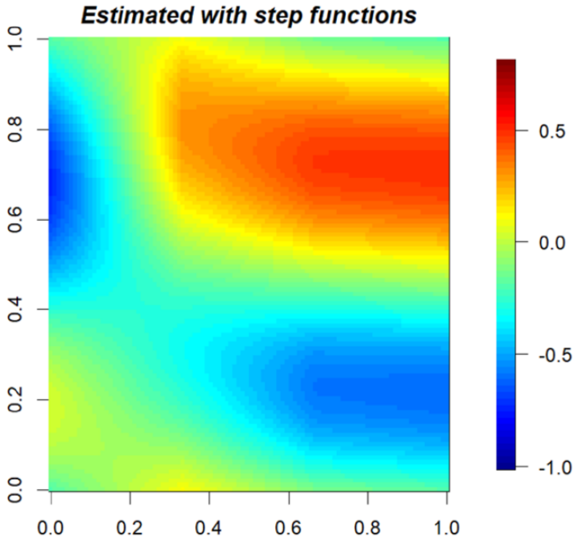
True and estimated coefficient surfaces under the three designs



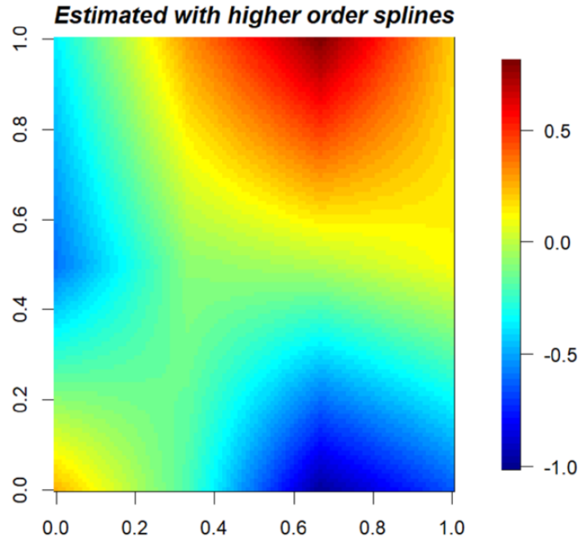
(a) True coefficient



(b) Estimated from multivariate design



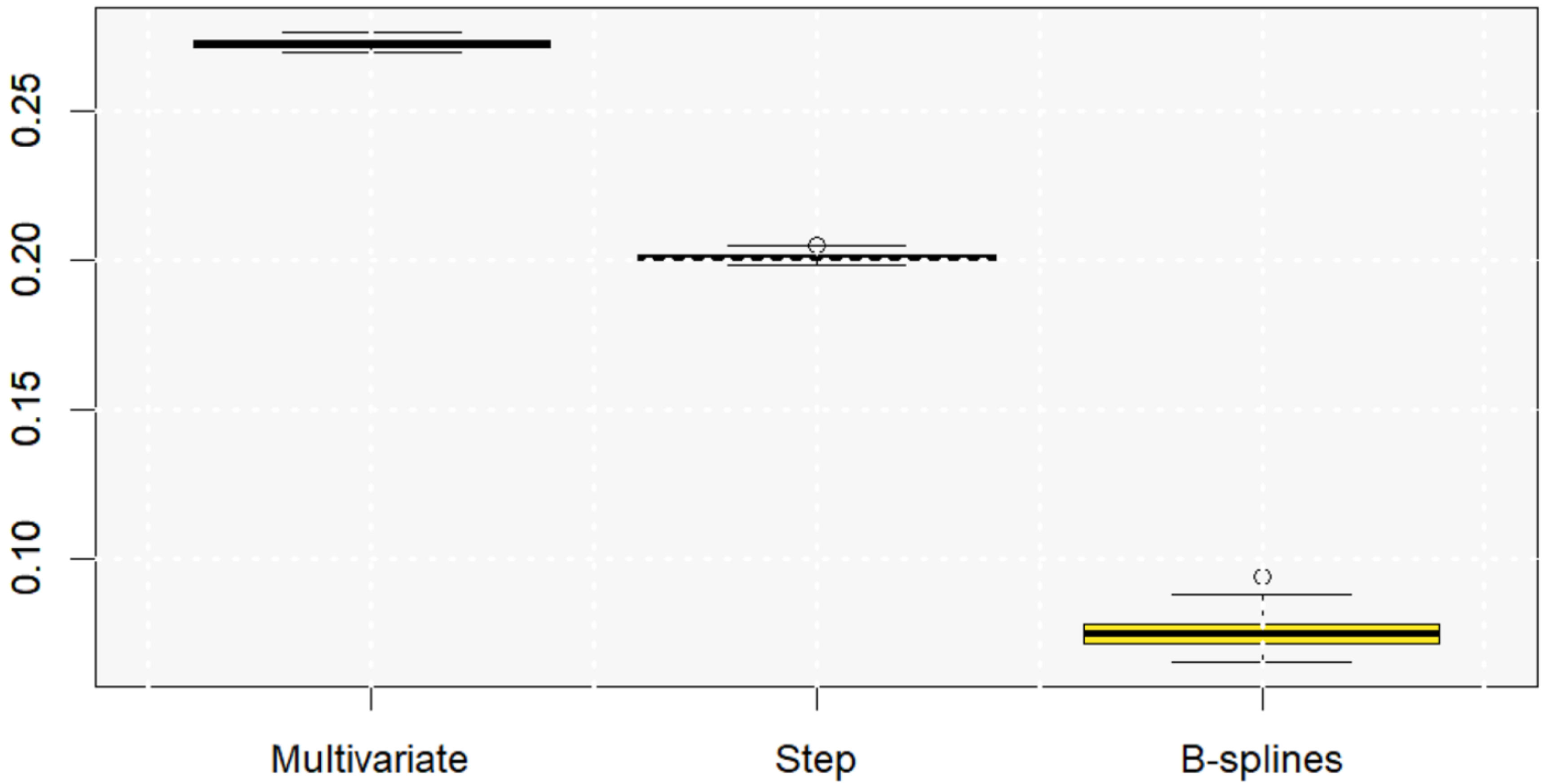
(c) Estimated from step-step design



(d) Estimated from spline-spline design

Performances of the three designs

RMSE on a 100 x 100 grid, over 100 simulated experiments:



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

Conclusions and further developments

Conclusions and further developments

- We developed a strategy for A-optimal and D-optimal design for reduced rank estimators in fully functional linear models.

- Interestingly, the designs do not depend on the stochastic distributions of the errors or on the bases of functions used for the responses.

- The basis representation of the factors can also be flexibly chosen based on the experimental requirements.

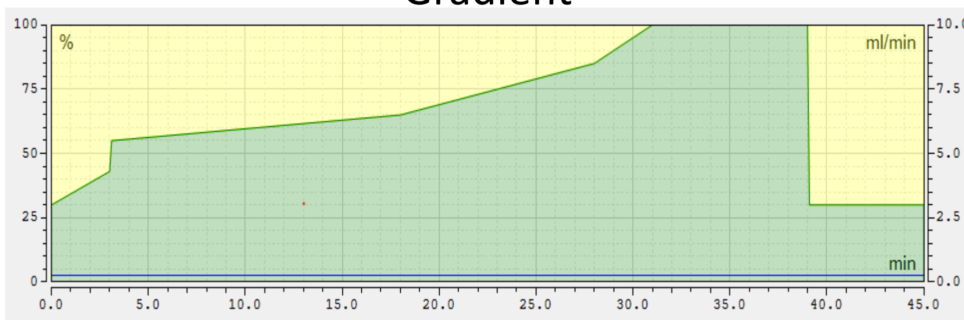
- The complexity of the coefficients in the direction of the factor is an important modelling choice that impact on the design.

- Things are more complicated for the penalized estimator.
  Penalization guarantees existence of the estimator when the unpenalized information matrix is singular, but the associated sampling covariance matrix remains rank-deficient in supersaturated regimes. We need to solve the design problem, and we develope a Bayesian approach.

Conclusions and further directions

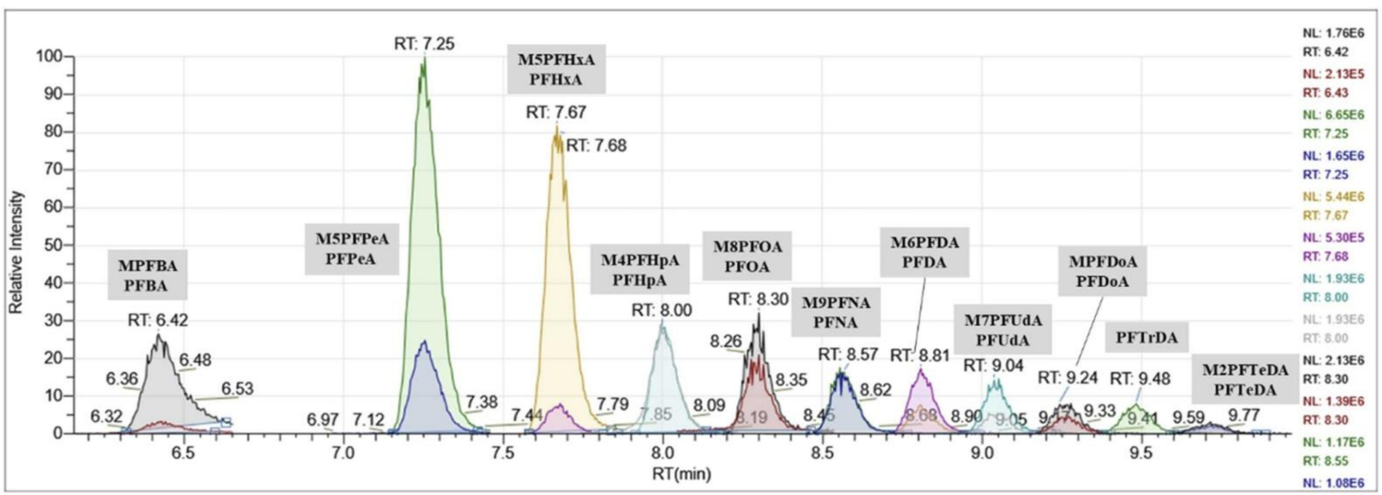
High-Performance Liquid Chromatography (HPLC)

Gradient



optimization

Chromatogram



C. May et al. caterina.may@uniupo.it

UNIVERSITÀ DEL PIEMONTE ORIENTALE

References

- Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, 2nd ed. Springer, New York.
- Aletti, G., May, C., Tommasi, C. (2016). On applying optimal design of experiments when functional observations occur. In: mODa 11 - Advances in Model-Oriented Design and Analysis, pp. 1–9. Springer, Cham.
- Pigoli, D., Ferraty, F., Aston, J.A.D., Mazumder, A., Richards, C., Hall, M.J.R. (2023). Estimation of temperature-dependent growth profiles for the assessment of time of hatching in forensic entomology. JRSS C 72(2), 231–253.
- Michaelides, D., Adamou, M., Woods, D.C., Overstall, A.M. (2026). Optimal design of dynamic experiments for scalar-on-function linear models with application to a biopharmaceutical study, BIOMETRICS, 82(1).

- **May, C., Ladas, T., Pigoli, D. and Mylona, K. (2024). Optimal design of experiments for functional linear models with dynamic factors. TEST (2026).**

- May, C., Ladas, T., Pigoli, D. and Mylona, K. (2025). A-optimal Designs of Experiments in Linear Models with Dynamic Factors and Functional Responses. Proceedings of IWFOS 2025, Springer.

- May, C., Ladas, T., Pigoli, D. and Mylona, K. (2024). Penalized optimum design of experiments for functional linear models with dynamic factors, *Working paper*.

Acknowledgments

THANK YOU!

Funded by:

- *European Union – NextGeneration EU through the Italian Ministry of University and Research under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice Prin 2022 - DD N. 104 del 2/2/2022, title Optimal and adaptive designs for modern medical experimentation, proposal code [2022TRB44L] - CUP [G53D23001600006].*
- *Multi-objective optimal design of experiments* funding by the EPSRC (Engineering and Physical Sciences Research Council, UK.
- *SSHINE 2025* funding of University of Eastern Piedmont.

caterina.may@uniupo.it