POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Bayesian Latent Factor Modeling for Multi-Target Inference

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Bruno Ursino**

Student ID: 939687
Advisor: Prof.ssa Ilenia Epifani
Co-advisors: Francesco Trovò PhD, Prof.ssa Manuela Antonelli, Beatrice Cantoni PhD
Academic Year: 2022-23

Abstract

Providing safe drinking water is essential for human health. To achieve this, analyzing multidimensional spectroscopic measurements of water, such as absorbance and fluorescence spectra, is crucial. However, measuring such parameters entails significant time and instrumental efforts. Therefore, it is essential to develop statistical tools to minimize the required lab analyses. We propose a bivariate functional data model, where each sampling unit comprises the bivariate target: absorbance and fluorescence. We employed a Bayesian bivariate functional latent factor model extending Montagna et al. (2012). In our water analyses application, interpretable posterior distributions of the latent factors are crucial, thus, we addressed their identifiability question by applying the Varimax-RSP algorithm. We developed a Python package, available on GitHub,implementing our model and an extensive toolbox for exploring its posterior and posterior predictive distribution, tailored to the use-case of applied functional data analysis. Stan Hamiltonian Monte Carlo No U-Turn Sampler was applied to sample the posterior distribution, employing efficient computations of the likelihood function. This was achieved by projecting the likelihood functions on a low-dimensional vector space, and observing the property of full conjugacy, of the residual of such projection, with a portion of the prior distribution. Finally, this methodology was validated using both simulated and real datasets, showcasing its applicability to functional data analysis problems, including the water data from the European Project SafeCREW.

Keywords: Basis Functions, Bayesian Latent Factor model, Bivariate Functional Data, Drinking Water NOM Characterization, Hamiltonian MCMC, Identifiability of Latent Factors, Varimax-RSP algorithm

Abstract in lingua italiana

Fornire acqua potabile sicura è essenziale per la salute umana. Per raggiungere questo obiettivo, è cruciale analizzare delle misurazioni spettroscopiche multidimensionali dell'acqua, come gli spettri di assorbanza e fluorescenza. Tuttavia, la misurazione di tali parametri comporta significativi sforzi in termini di tempo e strumentazione. Pertanto, è essenziale sviluppare strumenti statistici per minimizzare le analisi di laboratorio richieste. Proponiamo un modello per dati funzionali bivariati, dove ogni unità di campionamento comprende il target bivariato: assorbanza e fluorescenza. Abbiamo utilizzato un modello Bayesiano funzionale bivariato a fattori latenti. Nella nostra applicazione di analisi dell'acqua, è cruciale ottenere distribuzioni a posteriori interpretabili dei fattori latenti, quindi abbiamo affrontato la loro questione di identificabilità, applicando l'algoritmo Varimax-RSP. Abbiamo sviluppato un pacchetto Python, disponibile su GitHub, che implementa il nostro modello e un ampio toolbox per l'esplorazione della sua distribuzione a posteriori e predittiva a posteriori, su misura per il caso d'uso dell'analisi di dati funzionali. L'Hamiltonian Monte Carlo No U-Turn Sampler di Stan è stato applicato per campionare la distribuzione a posteriori, impiegando calcoli efficienti della funzione di verosimiglianza. Ciò è stato ottenuto proiettando le funzioni di verosimiglianza su uno spazio vettoriale a bassa dimensionalità, e osservando la proprietà di completa coniugazione del residuo di tale proiezione, con una parte della distribuzione a priori. Infine questa metodologia è stata validata, utilizzando sia dati simulati che reali, dimostrando la sua applicabilità a problemi di analisi di dati funzionali, incluso dati sull'acqua provenienti dal Progetto Europeo SafeCREW.

Parole chiave: Algoritmo Varimax-RSP, caratterizzazione della NOM in acqua, dati funzionali bivariati, funzioni di base, identificabilità dei fattori latenti, MCMC hamiltoniano, modello bayesiano con fattori latenti

Contents

Abstract **i**

Abstract in lingua italiana **iii**

Contents **v**

Introduction **1**

1 Motivating Example **5**

2 Bivariate Bayesian Latent Factor Model **9**

2.1 BBLF Model Specification 9

2.1.1 Likelihood Function 10

2.1.2 Prior Specification 13

2.2 An overview of Basis Matrices 16

2.2.1 Parametric Basis Functions 16

2.2.2 Basis Matrices induced by a Covariance Kernel 19

2.2.3 Basis Matrices for non-functional targets 22

3 Exploring the BBLF Model: Mathematical Properties and Tools for Posterior Analysis **23**

3.1 Alternative Representations of the Likelihood Function 23

3.1.1 Marginal Likelihood Function of the Target with respect to the Latent Factors 25

3.1.2 Marginal Likelihood Function of the target decomposed in the basis space and its residual 27

3.1.3 Other relations using the Basis Space-Residual decomposition . . . 29

3.2 Prior Predictive Statistical Moments 30

3.3 Interpretability of Latent Factors: the Varimax-RSP algorithm 32

4 Validation of the BBLF Model with Toy Datasets **37**
4.1 Dataset Generated from Prior Distribution: The 1D Case 37
4.2 Dataset Generated from Prior Distribution: The 2D case 41
4.3 Application to Function-to-Vector Regression:
The Biscuit Dataset 42

5 Posterior Inference on NOM Characterization **49**
5.1 Prediction of the Peak Picking 49
5.2 Prediction of Excitation-Emission matrix 51

6 Conclusions and future developments **57**
6.1 Conclusions 57
6.2 Future Developments 57

Bibliography **59**

A Prior Statistical Moments of the Latent Factors **63**

B Prior Predictive Moments **69**

C Calibration of Post-Varimax-RSP Latent Components **75**

D Clustering and Correlation Analysis between Absorbance and Fluorescence spectra **79**
D.1 Data Pre-processing 79
D.2 Hierarchical Clustering 80
D.2.1 Dissimilarity Measure 81
D.2.2 Clustering 82
D.2.3 Post-processing 83
D.2.4 Results 84

E Stan Code **89**

Acknowledgements **95**

Introduction

Natural organic matter (NOM) significantly influences aquatic ecosystems, affecting water quality through changes in taste, and odor, and the creation of harmful disinfection byproducts (DBPs) such as trihalomethanes and haloacetic acids. These DBPs form when NOM reacts with the chemical disinfectants used in water purification, posing health risks. Research has shown that the specific makeup of NOM can affect the levels and types of DBPs produced, highlighting the importance of accurately assessing NOM's concentration and composition. Such analysis is key in predicting DBP formation, refining water treatment methods, and safeguarding drinking water quality.

Analyzing NOM through UV-Visible spectroscopy involves examining multivariate functional data obtained from each water sample. Traditionally, NOM's presence in water is evaluated through absorbance spectra, which are relatively straightforward to obtain, and fluorescence spectra, which are more complex, time-intensive, and costly. The data from these spectroscopic methods are represented as functions over a two-dimensional domain, introducing additional complexity to the analysis.

This study aims to quantify the link between absorbance and fluorescence spectra under various conditions, identified by differing covariate values. For example, covariates can include specifics about the water sample's treatment process, such as the contact time with Activated Carbon. Establishing this connection is crucial for developing reliable methods to infer NOM characteristics using more easily obtainable measurements.

To address this, we introduce a novel bivariate Bayesian latent factor (BBLF) model that builds upon the Bayesian regression model for a single functional target proposed in [15]. This model accommodates a multivariate framework, allowing simultaneous analysis of different spectroscopic data types. It aims to identify latent factors common to both targets, associated with each water sample, to uncover underlying variables indicating how various solutes influence the spectra. This approach is similar to the Parallel Factor Analysis (PARAFAC) used in Fluorescence Spectroscopy, providing a familiar conceptual framework for those studying NOM.

The BBLF model offers a structured method for comprehensively exploring the relation-

ship between two functional targets. It facilitates the identification and examination of latent variables and tackles the complexities of function-on-function regression. This provides a robust mathematical tool integrating advanced statistical methods into environmental science. The proposed methodology has been implemented in a Python package, and, thanks to such a tool, the model's wide applicability to functional data analysis is demonstrated over synthetically generated and real-world data.

A novelty that is discussed in this thesis, consists in an alternative representation of the likelihood function of the model, allowing for significant improvement in the computational complexity of the likelihood computation step of the Hamiltonian Monte Carlo sampler we used. In particular we prove a decomposition of the likelihood function in two parts, one that projects the data onto a lower dimensional vector space, the other which allows for the choice of a conjugate prior for one of the parameters of the model.

The organization of this thesis is outlined as follows:

In Chapter 1, the problem motivating our investigation is explored in greater detail, introducing typical tools of fluorescence spectroscopy analysis.

Chapter 2 is devoted to a comprehensive analysis of the Bayesian Bivariate Latent Factor (BBLF) model, including model specification (likelihood function, prior distribution and the selection of basis matrices).

Chapter 3 is focused on a discussion on the mathematical properties of the BBLF model. In this chapter we present various useful mathematical tools for the exploration of the posterior and posterior predictive distribution, including a reformulation of the likelihood function and the Varimax-RSP algorithm, a tool useful for identifiability of the latent components.

Chapter 4 showcases applications of the model through three examples: the first two validate the BBLF model and its comprehensive toolkit by fitting datasets generated from the model's prior/prior predictive distribution; the third applies the model to a function-to-vector regression task with a real dataset, naturally partitioned into training and test sets.

In Chapter 5 the BBLF model is applied to absorbance and fluorescence spectroscopy data, showcasing its practical application.

Appendices A, B and C supplement the main text with additional results and proofs not included in the body of the thesis.

Appendix D provides a concise overview of the preliminary unsupervised analysis con-

ducted to investigate the relationship between absorbance and fluorescence spectra. The results have been incorporated into a deliverable of a completed task within the European project SafeCREW (Grant agreement ID: 101081980, [2]).

Finally, Appendix E presents the Stan code for the BBLF model.

1 | Motivating Example

Natural Organic Matter (NOM) constitutes a critical component of aquatic ecosystems, emerging from the biological decomposition of plants, animals, and microbial processes. This mixture of carbon-based substances encompasses a wide range of molecular structures, from humic and fulvic acids to carbohydrates, proteins, and various other organic molecules. The ubiquity and diversity of NOM play pivotal roles in water systems, not only by participating in essential biogeochemical cycles but also by influencing water quality through its interactions with microbial communities and water treatment chemicals [13].

The inherent complexity of NOM stems from its varied origins and the myriad of transformation processes it undergoes in the environment. These processes lead to a heterogeneous mixture of molecules with a broad spectrum of sizes, structures, and functional groups, each exhibiting different chemical and physical properties. The variability of NOM is further influenced by factors such as seasonal changes, watershed characteristics, and anthropogenic activities, which contribute to fluctuations in its concentration, composition, and reactivity in aquatic environments.

Within the realm of water treatment, NOM's tendency to react with disinfectants can lead to the creation of disinfection byproducts (DBPs). These compounds, some of which are harmful to human health, necessitate a careful balance in the treatment process to effectively remove pathogens without exacerbating DBP formation. The variability and complexity of NOM require sophisticated analytical techniques for its accurate characterization, enabling water treatment facilities to adapt their processes dynamically to the changing nature of NOM and minimize potential health risks.

Climate change introduces additional complexity, influencing aquatic ecosystems through changes in precipitation and temperatures, leading to significant variations in the concentration, composition, and reactivity of NOM. The warming climate intensifies the microbial processes that contribute to NOM formation, potentially increasing the presence of both NOM and associated micropollutants in water bodies. Extreme weather events, such as heavy rainfall and droughts, can concentrate or alter NOM and its constituents,

thereby influencing its interactions with water treatment chemicals [2, 5].

DBPs are generated when disinfectants such as chlorine, chloramines, ozone, and chlorine dioxide interact with NOM precursors, as well as other organic and inorganic matter present in the water. The formation of DBPs is influenced by a range of water treatment factors, including the type and concentration of disinfectant, water pH, temperature, and contact time between disinfectants and precursors. This leads to a wide variety of DBPs, with trihalomethanes (THMs) and haloacetic acids (HAAs) being among the most common and concerning for their potential health effects.

The health implications associated with exposure to certain DBPs include increased risks of cancer, as well as potential impacts on the liver, kidney, and central nervous system. Some DBPs have also been implicated in adverse reproductive and developmental outcomes, prompting regulatory agencies worldwide to establish guidelines and standards to limit DBP concentrations in drinking water. Despite these regulations, the sheer variety of DBPs, many of which are still not fully characterized or understood in terms of their health impacts, presents ongoing challenges for water treatment and public health protection [22, 30].

The complex nature of DBPs formation calls for sophisticated detection and analysis methods, as well as innovative treatment strategies to reduce their potential formation. Approaches to mitigate DBPs formation involve optimizing the use of alternative disinfectants with lower DBPs formation potential, enhancing the removal of DBPs precursors through advanced water treatment processes (such as coagulation, adsorption, and advanced oxidation), and fine-tuning treatment parameters to minimize conditions favorable for DBPs formation.

The analysis of NOM in aquatic settings has greatly advanced through the use of absorbance and fluorescence spectroscopy. These techniques enable a comprehensive analysis of NOM, shedding light on its concentration, composition, and interactions within water systems by examining patterns of light absorption and emission [14].

UV-Visible (UV-Vis) spectroscopy exploits the differential absorption of ultraviolet (UV) and visible light by various compounds, providing valuable insights into the molecular complexity and dynamics of NOM in water environments. By quantifying the light absorbed by NOM constituents at particular wavelengths, UV-Vis spectroscopy facilitates a nuanced understanding of the concentration, structural characteristics, and possible interactions of the organic matrix within natural waters.

NOM's absorption of UV light, particularly in the spectral range of 200 to 280 nm,

is chiefly attributed to the electronic transitions in aromatic structures and conjugated double bonds within the NOM. The characteristic absorbance at 254 nm, prominently linked to the presence of aromatic organic compounds, is routinely employed as a surrogate measure for assessing NOM concentration. This parameter is critical in water treatment contexts, where the efficient removal of NOM is paramount to prevent the formation of DBPs, underscoring the technique's utility in optimizing water purification strategies [14].

Fluorescence spectroscopy extends the analytical depth by detecting and characterizing the fluorescent fractions of NOM. Through excitation at various wavelengths and the measurement of the resultant emission, this technique uncovers the presence of specific NOM components, such as humic and fulvic acids or protein-like substances. The generation of Excitation-Emission Matrices (EEMs) through this method provides a comprehensive fluorescence signature of NOM, allowing for the identification of diverse organic fractions.

The data obtained from fluorescence spectroscopy are instrumental in understanding the interactions of NOM with environmental contaminants and treatment chemicals. For instance, identifying protein-like fluorescence signals can indicate increased microbial activity, while humic-like signals indicate terrestrial-derived NOM. Such insights are crucial for assessing water quality and guiding the optimization of water treatment strategies, particularly in reducing the formation of harmful DBPs.

In fluorescence spectroscopy data analysis, peak picking and region integration are critical for extracting meaningful information from complex spectra. Peak picking involves identifying significant peaks within the fluorescence emission-excitation matrix, which correspond to specific NOM components. This method requires an understanding of the spectral signatures of different organic materials and the ability to differentiate between overlapping signals.

Region integration complements peak picking by quantifying the total fluorescence within predefined spectral regions. This approach provides an aggregate measure of NOM content, useful for comparing water samples or tracking changes in NOM composition over time.

Parallel Factor Analysis (PARAFAC) is a powerful tool for interpreting fluorescence data, utilizing a three-dimensional tensor that captures excitation, emission, and sample dimensions. This method breaks down the tensor into trilinear components, each representing a unique factor, through a least-squares fitting process. These factors are linked to specific fluorescent substances within the NOM mix, facilitating an in-depth exploration of the complex interactions among NOM constituents. Experts analyze the derived bilinear EEMs factors by comparing them to EEM-PARAFACs in previous literature. They then

leverage the factor loadings to scrutinize the samples in question, considering the context of their treatment processes or the specifics of their collection, such as location and timing. To ensure accuracy, EEM-PARAFAC incorporates multiple validation strategies. The analysis often involves multiple iterations, adjusting the number of factors, checking for residual patterns, and identifying outliers. For substantial datasets, a validation step is included, typically featuring split-half tests to verify the robustness of the findings [7, 17].

This thesis introduces a Bayesian Latent Factor model that offers several advantages over traditional techniques. Firstly, unlike peak-picking or region integration methods, our model maintains a functional representation of EEMs, similar to that achieved by PARAFAC, allowing for a more nuanced analysis. Secondly, it enables the integration of additional functional data. While our primary focus was on exploring the relationship between absorbance spectra and EEMs, our model is also compatible with data from other methodologies commonly used in the NOM characterization, such as Liquid Chromatography – Organic Carbon Detection, facilitating a comprehensive analysis of functional datasets. Thirdly, the Bayesian approach inherently simplifies the validation process. By analyzing the posterior distribution, our model provides direct insights into the optimal number of latent factors, an essential aspect for accurately analyzing datasets with limited samples. This characteristic is particularly valuable, as it obviates the need for extensive external validation steps typically required by other models.

In Appendix D, we provide a concise overview of the preliminary unsupervised analysis conducted to investigate the relationship between absorbance and fluorescence spectra. The results have been incorporated into a deliverable of a completed task within the European project SafeCREW (Grant agreement ID: 101081980, [2]).

2 | Bivariate Bayesian Latent Factor Model

In this chapter, we focus on the Bayesian Bivariate Latent Factor (BBLF) model, a framework developed to analyze bivariate functional data through the lens of shared latent factors. This model extends the foundational work of Montagna et al., [15], and incorporates the advancements made by Jang et al., who generalized the model to bivariate targets with shared latent factors [11]. While their focus was on imputation of missing data, our approach aims to uncover the intricate relationship between two functional targets by leveraging a common set of latent factors.

Within the context of spectroscopic measurements, these factors are conceptualized as representations of various analytes or groups of covarying analytes, each characterized by distinct absorbance and fluorescence spectra. Our method bears similarities to the PARAFAC technique, a well-established method in Fluorescence Spectroscopy [14, 16], thus facilitating the interpretation of results within familiar conceptual frameworks for experts in the characterization of natural organic matter (NOM).

More in general, the BBLF model represents a structured methodology that allows for a comprehensive examination of the underlying relationship between two functional targets. This not only includes the identification and exploration of latent components but also addresses the complexities involved in function-on-function regression, offering a mathematically rigorous tool, that supports the integration of advanced statistical techniques into environmental science.

2.1. BBLF Model Specification

Let $\boldsymbol{Y} = (\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)})^\top$ be the two-target functionals measured on N sample units $s = 1, \ldots, N$ and modeled by the BBLF model. For each sample unit $s = 1, \ldots, N$, the measurement on target $\boldsymbol{Y}^{(\iota)}$ is taken at L_ι fixed locations $h_\iota = 1, \ldots, L_\iota$ for $\iota = 1, 2$, and the data from the s-th unit are encoded as vectors. Let $Y^{(\iota)}_{h_1 s}$ denote the measurement of

target $\iota = 1, 2$ at location h_1 for the sample unit s,

$$\boldsymbol{Y}_s^{(\iota)} = \begin{pmatrix} Y_{h_1 s}^{(\iota)} \\ \vdots \\ Y_{h_{L_\iota}, s}^{(\iota)} \end{pmatrix} \qquad \boldsymbol{Y}_s = \begin{pmatrix} \boldsymbol{Y}_s^{(1)} \\ \boldsymbol{Y}_s^{(2)} \end{pmatrix}$$

Then $\boldsymbol{Y}_s$ is the bivariate functional data vector of the sample unit s.

2.1.1. Likelihood Function

The likelihood function quantifies the probability of observing the data given the model parameters.
The likelihood function of the BBLF model is depicted schematically in Figure 2.1.

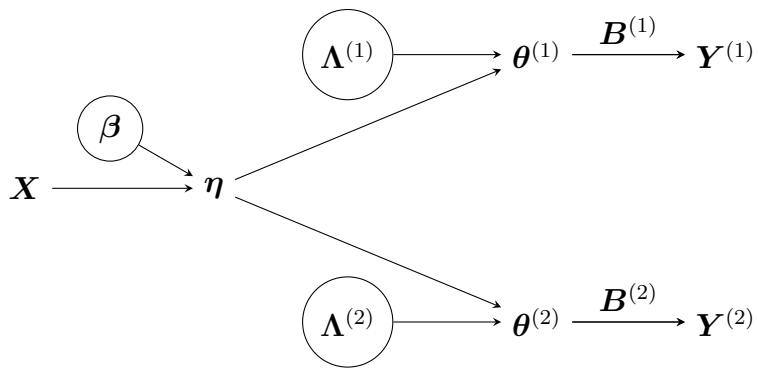


Figure 2.1: Schematic representation of the BBLF likelihood showing the progression from covariates $\boldsymbol{X}$ through latent factors $\boldsymbol{\eta}$ and their components $\boldsymbol{\Lambda}^{(1)}$ and $\boldsymbol{\Lambda}^{(2)}$, leading to the functional targets $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$.

At the core of the model, each target's data vector is conceptualized as a Gaussian random vector centered around a linear combination of ι-specific basis functions, represented by an $L_\iota \times p_\iota$ matrix $\boldsymbol{B}^{(\iota)}$; the coefficients, referred to as 'basis coefficients', are denoted by $\boldsymbol{\theta}_s^{(\iota)}$.

On each sample unit s, the basis coefficients are linearly regressed onto a set of k latent factor scores, represented as $\boldsymbol{\eta}_s = (\eta_{1,s}, \ldots, \eta_{k,s})$, by using the matrix of the target-specific loading factor $\boldsymbol{\Lambda}^{(\iota)}$, of dimensions $p_\iota \times k$ for $\iota = 1, 2$. Additionally, the latent factors $\boldsymbol{\eta}_s$, shared by both targets, are modeled as centered on a linear combination $\boldsymbol{\beta}\boldsymbol{X}_s$ of r covariates.

The basis matrices $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$, along with coefficients $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$, form the basis function representations for the functional targets. The design matrix $\boldsymbol{X}$ includes the

covariates, while $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ represent the target data matrices.

The large number of basis functions p_1, p_2 allows the BBLF model to flexibly accommodate a broad array of functional shapes, while the reduced number of latent factors k, implements a mechanism of regularization.

For simplicity, let us encompass all the model parameters in a comprehensive parameter set $\boldsymbol{\Theta}$ given by:

$$\boldsymbol{\Theta} = \left\{\boldsymbol{\beta}, \boldsymbol{\Lambda}^{(1)}, \boldsymbol{\Lambda}^{(2)}, \boldsymbol{\tau}_{\lambda}^{(1)}, \boldsymbol{\tau}_{\lambda}^{(2)} \boldsymbol{\tau}_{\theta}^{(1)}, \boldsymbol{\tau}_{\theta}^{(2)}, \psi^{(1)}, \psi^{(2)}\right\}, \tag{2.1}$$

where

- $\boldsymbol{\beta}$ is the $(k \times r)$ matrix of the regression coefficients linking r covariates to the k latent variables $\boldsymbol{\eta}$,
- $\boldsymbol{\Lambda}^{(\iota)}$ is the $(p_\iota \times k)$ matrix of the p_ι loading factors of the k latent variables $\boldsymbol{\eta}$ for each functional target $\iota = 1, 2$,
- $\boldsymbol{\tau}_{\lambda}^{(1)}$ and $\boldsymbol{\tau}_{\lambda}^{(2)}$ are the precision parameters for the latent factors (vectors of length k),
- $\boldsymbol{\tau}_{\theta}^{(1)}$ and $\boldsymbol{\tau}_{\theta}^{(2)}$ are vectors of length p_ι of the precision parameters for the basis function coefficients for $\iota = 1, 2$,
- $\psi^{(1)}$ and $\psi^{(2)}$ are the standard deviations of the observation noises around the functional targets.

In the context of our latent factor model, the likelihood of the N functional data

$$\{\boldsymbol{Y}_s\}_{s=1,\ldots,N} = \left\{\left(\boldsymbol{Y}_s^{(1)\top}, \boldsymbol{Y}_s^{(2)\top}\right)^\top\right\}_{s=1,\ldots,N}$$

is formulated under the assumption of independence between all sample units $s = 1, \ldots, N$, conditionally on the set of parameters. Therefore, the complete specification of the like-

lihood of the BBLF model in Figure 1.1 is:

$$
\begin{aligned}
Y^{(1)}_{h_1,s}\Big|\boldsymbol{\Theta},\theta^{(1)},\theta^{(2)},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\sum_{j=1}^{p_1} B^{(1)}_{h_1,j}\theta^{(1)}_{j,s},\ {\psi^{(1)}}^2\right)\\
Y^{(2)}_{h_2,s}\Big|\boldsymbol{\Theta},\theta^{(1)},\theta^{(2)},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\sum_{j=1}^{p_2} B^{(2)}_{h_2,j}\theta^{(2)}_{j,s},\ {\psi^{(2)}}^2\right)\\
\theta^{(1)}_{j_1,s}\Big|\boldsymbol{\Theta},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\sum_{\ell=1}^{k} \Lambda^{(1)}_{j_1,\ell}\eta_{\ell,s},\ 1/\tau^{(1)}_{\theta\, j_1}\right)\\
\theta^{(2)}_{j_2,s}\Big|\boldsymbol{\Theta},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\sum_{\ell=1}^{k} \Lambda^{(2)}_{j_2,\ell}\eta_{\ell,s},\ 1/\tau^{(2)}_{\theta\, j_2}\right)\\
\eta_{\ell,s}|\boldsymbol{\Theta} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\sum_{i=1}^{r} \beta_{\ell,i}X_{i,s},\ 1\right)
\end{aligned}
\tag{2.2}
$$

for:

$$
\begin{aligned}
h_1 &= 1,\dots,L_1, & h_2 &= 1,\dots,L_2,\\
j_1 &= 1,\dots,p_1, & j_2 &= 1,\dots,p_2,\\
\ell &= 1,\dots,k, & s &= 1,\dots,N.
\end{aligned}
$$

Equivalently, the BBLF model in Figure 1.1 can be expressed by means of affine transformations of Gaussian components:

$$
\begin{aligned}
\boldsymbol{\eta}_s &= \boldsymbol{\beta}\boldsymbol{X}_s &&+ \boldsymbol{\varepsilon}_{\eta,s}\\
\boldsymbol{\theta}^{(1)}_s &= \boldsymbol{\Lambda}^{(1)}\boldsymbol{\eta}_s &&+ \boldsymbol{\varepsilon}_{\theta_1,s}\\
\boldsymbol{\theta}^{(2)}_s &= \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}_s &&+ \boldsymbol{\varepsilon}_{\theta_2,s}\\
\boldsymbol{Y}^{(1)}_s &= \boldsymbol{B}^{(1)}\boldsymbol{\theta}^{(1)}_s &&+ \boldsymbol{\varepsilon}_{Y_1,s}\\
\boldsymbol{Y}^{(2)}_s &= \boldsymbol{B}^{(2)}\boldsymbol{\theta}^{(2)}_s &&+ \boldsymbol{\varepsilon}_{Y_2,s}
\end{aligned}
$$

Using matrix notation, the last formula is succinctly summarized as follows:

$$
\boldsymbol{Y}_s = \begin{bmatrix} \boldsymbol{Y}^{(1)}_s \\ \boldsymbol{Y}^{(2)}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{B}^{(1)} & \boldsymbol{0} & \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{B}^{(2)} & \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)} \end{bmatrix} \boldsymbol{\mathcal{E}}_s \tag{2.3}
$$

where

$$
\boldsymbol{\mathcal{E}}_s = \begin{pmatrix} \boldsymbol{\varepsilon}_{Y_1,s} \\ \boldsymbol{\varepsilon}_{Y_2,s} \\ \boldsymbol{\varepsilon}_{\theta_1,s} \\ \boldsymbol{\varepsilon}_{\theta_2,s} \\ \boldsymbol{\varepsilon}_{\eta,s} \end{pmatrix}
$$

and, conditionally on $\boldsymbol{\Theta}$, $\boldsymbol{\mathcal{E}}_s$ is normally distributed with a diagonal covariance structure, i.e.

$$\boldsymbol{\mathcal{E}}_s | \, \boldsymbol{\Theta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

with

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c|ccc|ccc|c}
{\psi^{(1)}}^2 \boldsymbol{I} & \mathbf{0} & & \mathbf{0} & & & \mathbf{0} & & \mathbf{0} \\ \hline
\mathbf{0} & {\psi^{(2)}}^2 \boldsymbol{I} & & \mathbf{0} & & & \mathbf{0} & & \mathbf{0} \\ \hline
 & & 1/\tau_{\theta\,1}^{(1)} & \cdots & 0 & & & & \\
\mathbf{0} & \mathbf{0} & \vdots & \ddots & \vdots & & \mathbf{0} & & \mathbf{0} \\
 & & 0 & \cdots & 1/\tau_{\theta\,p_1}^{(1)} & & & & \\ \hline
 & & & & & 1/\tau_{\theta\,1}^{(2)} & \cdots & 0 & \\
\mathbf{0} & \mathbf{0} & & \mathbf{0} & & \vdots & \ddots & \vdots & \mathbf{0} \\
 & & & & & 0 & \cdots & 1/\tau_{\theta\,p_2}^{(2)} & \\ \hline
\mathbf{0} & \mathbf{0} & & \mathbf{0} & & & \mathbf{0} & & \boldsymbol{I}
\end{array}\right] = \text{block_diag} \begin{bmatrix} {\psi^{(1)}}^2 \boldsymbol{I} \\ {\psi^{(2)}}^2 \boldsymbol{I} \\ \boldsymbol{\Delta}_\theta^{(1)} \\ \boldsymbol{\Delta}_\theta^{(2)} \\ \boldsymbol{I} \end{bmatrix}$$

2.1.2. Prior Specification

To finalize the Bayesian model, we carefully selected suitable marginal priors for each parameter in the likelihood, starting from those appearing in the last block to those in the first one in Equation (2.2), and assumed independence between blocks of parameters.

Prior Selection for Regression Coefficients $\boldsymbol{\beta}$. The regression coefficients $\boldsymbol{\beta}$ link the covariates to the latent variables. We employ a standard Student's t-distribution with χ degrees of freedom, allowing some coefficients to significantly diverge from the mean when supported by the data. In Chapter 4 on the implementation and validation of BBLF models on various datasets, $\chi = 4$ will be used, to obtain a vague prior for $\boldsymbol{\beta}$.

Prior Selection for Latent Factors $\boldsymbol{\Lambda}^{(\iota)}$ and hyperparameters $\boldsymbol{\tau}_\lambda^{(\iota)}$. The loadings $\boldsymbol{\Lambda}^{(1)}, \boldsymbol{\Lambda}^{(2)}$ are crucial for capturing the underlying structure of the functional targets. Following [15], we assume that a priori $\boldsymbol{\Lambda}^{(1)}, \boldsymbol{\Lambda}^{(2)}$ are independent and identically distributed *Multiplicative Gamma Processes Shrinkage* (MGPS) and write $\boldsymbol{\Lambda}^{(\iota)} \sim \text{MGPS}(\nu, \alpha_1, \alpha_2)$. This means that for $\iota = 1, 2$, $j = 1, \ldots, p_\iota$ and $\ell = 1, \ldots, k$, the loading factors $\Lambda_{j,\ell}^{(\iota)}$'s are indepedent and have a centered Student's t-distribution prior with precision parameter $\tau_{\lambda\,j}^{(\iota)}$, which acts as a global shrinkage parameter for the latent factor η_j with respect to target ι, i.e.

$$\Lambda_{j,\ell}^{(\iota)} \Big| \, \boldsymbol{\tau}_\lambda^{(\iota)} \overset{indep.}{\sim} \mathbf{t}_\nu \left(0, \frac{1}{\tau_{\lambda\ell}^{(\iota)}}\right)$$

$$\tau_{\lambda\ell}^{(\iota)} = \prod_{\gamma=1}^{\ell} \delta_{\gamma}^{(\iota)} \qquad \forall \ell = 1, \ldots, k$$

with

$$\delta_1^{(\iota)} \sim \mathrm{Ga}\left(\alpha_1, 1\right) \quad \text{and} \quad \delta_\ell^{(\iota)} \sim \mathrm{Ga}\left(\alpha_2, 1\right) \qquad \forall \ell = 2, \ldots, k.$$

Before moving on the other parameters, given the significant role the latent components play in this model, we pause to discuss prior means and variances of $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(2)}$ and $\boldsymbol{\tau}_\lambda^{(\iota)} = (\tau_{\lambda\ 1}^{(\iota)}, \ldots, \tau_{\lambda\ p_\iota}^{(\iota)})$, referring the reader to Appendix A for derivations and additional results. Indeed, the role of $\tau_{\lambda\ 1}^{(\iota)}$ as a global shrinkage parameter for the latent factor η_j with respect to target ι derives from them, with the hyperparameters α_1 and α_2 modulating the degree of regularization.

The conditional variance of the loading factor $\Lambda_{j,\ell}^{(\iota)}$ is given by:

$$\mathbb{V}\mathrm{ar}\left[\Lambda_{j,\ell}^{(\iota)} \middle| \boldsymbol{\tau}_\lambda^{(\iota)}\right] = \frac{\nu}{\nu - 2} \tau_{\lambda\ell}^{(\iota)^{-1}}$$

which leads an interest in the moments of the reciprocal of the precision for informed hyperparameter selection that are given by:

$$\mathbb{E}\left[\left(\tau_{\lambda\ \ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1},$$

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\ \ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_1 - 1)(\alpha_1 - 2)}\left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)}\right]^{\ell-1} \cdot \left[1 - \frac{\alpha_1 - 2}{\alpha_1 - 1}\left(\frac{\alpha_2 - 2}{\alpha_2 - 1}\right)^{\ell-1}\right]$$

for all $\ell = 1, \ldots, k$. If $1 < \alpha_1 < 2$, $1 < \alpha_2 < 2$, the expectation remains valid but the variance becomes undefined. The choice of $\alpha_2 > 2$ ensures diminishing effect sizes for higher indexed latent factors. Additionally, practical experience indicates difficulties in sampling using Hamiltonian Monte Carlo from distributions lacking variance, advocating for $\alpha_2 > 2$. Still, excessively large α_2 values, implying a limited number of latent factors with significant effect sizes, are not recommended.

Since α_1 influences only the 'initial condition' $\tau_{\lambda\ 1}^{(\iota)}$, other useful formulas are:

$$\mathbb{E}\left[\left(\frac{\tau_{\lambda\ \ell}^{(\iota)}}{\tau_{\lambda\ 1}^{(\iota)}}\right)^{-1}\right] = \left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

$$\mathbb{V}\mathrm{ar}\left[\left(\frac{\tau_{\lambda\ \ell}^{(\iota)}}{\tau_{\lambda\ 1}^{(\iota)}}\right)^{-1}\right] = \left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)}\right]^{\ell-1} \cdot \left[1 - \left(\frac{\alpha_2 - 2}{\alpha_2 - 1}\right)^{\ell-1}\right]$$

for all $\ell = 1, \ldots, k$.

Note that the number of latent factors k is unknown, and an optimal value can be determined by comparing models with varying k, using posterior predictive checks. The issue will be addressed in detail in Chapter 4 on the implementation and validation of BBLF models on various datasets.

Prior Selection for Precision Parameters of Basis Function Coefficients $\boldsymbol{\tau}_\theta^{(\iota)}$. For the p_ι precisions $\boldsymbol{\tau}_{\theta\,j}^{(\iota)}$ of each basis j and target ι, we assume a priori independence and the usual Gamma distribution: $\tau_{\theta\,j}^{(\iota)} \sim \mathrm{Ga}\left(\alpha_\sigma, \beta_\sigma\right)$. Individual $\boldsymbol{\tau}_{\theta\,j}^{(\iota)}$ for each basis function coefficient accommodates heteroscedasticity in the functional representation of the targets, providing a non-vanishing likelihood for residual functions that the latent factors failed to capture.

Prior Selection for Observation Noise $\psi^{(\iota)}$. The measurement noises $\boldsymbol{\varepsilon}_{Y_1,s}, \boldsymbol{\varepsilon}_{Y_2,s}$ of $\boldsymbol{Y}_1, \boldsymbol{Y}_2$, respectively are assumed to be homoscedastic, delegating the management of heteroscedasticity to the basis function coefficients' variance. In particular, the variances $\psi^{(1)}, \psi^{(2)}$ of the $\boldsymbol{\varepsilon}_{Y_1,s}, \boldsymbol{\varepsilon}_{Y_2,s}$ are a priori independent and Inv-Ga$(\alpha_\psi, \beta_\psi)$-distributed.

Complete Prior Specification. For $j_1 = 1, \ldots, p_1,\ j_2 = 1, \ldots, p_2,\ \ell = 1, \ldots, k,\ i = 1, \ldots, r$ we use:

$$
\begin{aligned}
\beta_{\ell,i} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\chi\left(0,\, 1\right) \\
\Lambda_{j_1,\ell}^{(1)} \Big| \ \tau_{\lambda\,\ell}^{(1)} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\nu\left(0,\, 1/\tau_{\lambda\,\ell}^{(1)}\right) \\
\tau_{\lambda\ell}^{(1)} = \prod_{\gamma=1}^{\ell} \delta_\gamma^{(1)}, \quad \delta_1^{(1)} &\sim \mathrm{Ga}\left(\alpha_1, 1\right),\ \delta_\ell^{(1)} \sim \mathrm{Ga}\left(\alpha_2, 1\right) \qquad \ell \geq 2 \\
\Lambda_{j_2,\ell}^{(2)} \Big| \ \tau_{\lambda\,\ell}^{(2)} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\nu\left(0,\, 1/\tau_{\lambda\,\ell}^{(2)}\right) \\
\tau_{\lambda\ell}^{(2)} = \prod_{\gamma=1}^{\ell} \delta_\gamma^{(2)}, \quad \delta_1^{(2)} &\sim \mathrm{Ga}\left(\alpha_1, 1\right),\ \delta_\ell^{(2)} \sim \mathrm{Ga}\left(\alpha_2, 1\right) \qquad \ell \geq 2 \\
\tau_{\theta\,j_1}^{(1)} &\overset{\text{i.i.d.}}{\sim} \mathrm{Ga}\left(\alpha_\sigma,\, \beta_\sigma\right) \\
\tau_{\theta\,j_2}^{(2)} &\overset{\text{i.i.d.}}{\sim} \mathrm{Ga}\left(\alpha_\sigma,\, \beta_\sigma\right) \\
{\psi^{(1)}}^2 &\overset{\text{i.i.d.}}{\sim} \text{Inv-Ga}\left(\alpha_\psi,\, \beta_\psi^{(1)}\right) \\
{\psi^{(2)}}^2 &\overset{\text{i.i.d.}}{\sim} \text{Inv-Ga}\left(\alpha_\psi,\, \beta_\psi^{(2)}\right).
\end{aligned}
$$

2.2. An overview of Basis Matrices

In the outlined model specification, the selection of appropriate basis matrices, denoted as $\boldsymbol{B}^{(\iota)}$, emerges as a pivotal step. This choice significantly impacts the model's ability to capture the variability observed in functional targets, allowing the statistician to customize the model to address the specific characteristics and challenges inherent in the analyzed data.

The basis matrices presented here are integrated into the Python package [25] that implements the BBLF model, demonstrating a commitment to providing a ready-to-use solution.

To illustrate the versatility of the BBLF model across diverse scenarios, we outline three categories of basis matrices. Firstly, we consider b-splines, Fourier basis functions, and radial basis functions, collectively referred to as 'parametric basis functions,' which offer a structured approach to modeling.

Next, we explore the category of 'covariance kernel-induced basis matrices.' This category differs from the first by focusing on constructing basis matrices through covariance kernels, a method driven by the modeler's intention to explicitly choose specific covariance structures a priori.

Lastly, we demonstrate how the BBLF model can be adapted for use with non-functional targets by appropriately selecting basis matrices and the hyperparameters of the prior distributions. This adaptation further exemplifies the model's broad applicability and flexibility in addressing a wide spectrum of data analytical challenges.

2.2.1. Parametric Basis Functions

Parametric basis functions are a cornerstone in the construction of models for approximating complex target functions across various domains. These functions are defined over the entire domain of the target function and are characterized by their dependence on a set of parameters and the location of measurement

Among the parametric basis functions, the b-splines, which are piecewise polynomials defined over a sequence of knots, ensure a certain regularity in the fitted curve depending on the degree of the polynomial considered. Initially defined in 1D, in the Python package made publicly accessible on GitHub [25], the 2D generalization is defined as the tensor product of the b-splines. A key advantage of this choice is the ability to select a non-uniform distribution of knots, providing more degrees of freedom in certain locations of

the domain or accommodating non-homogeneous sampling.

To define b-splines we can recur to the Cox-de Boor formula:

$$B_{i,0}(t) := \begin{cases} 1 & \text{if } t_i \leq t < t_{i+1} \\ 0 & \text{otherwise.} \end{cases}$$

$$B_{i,p}(t) := \frac{t - t_i}{t_{i+p} - t_i} B_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} B_{i+1,p-1}(t).$$

In the Python package developed in this thesis project, the construction of b-splines and Fourier bases is delegated to the scikit-FDA package [20]. In the BBLF model, one should not only choose the basis matrix based on the desired functional forms but also consider the amplitude of the target functions. To simplify this choice, in [25], the basis functions are renormalized to have an L^2 norm equal to 1:

$$\widetilde{B}_{i,p}(t) = \frac{B_{i,p}(t)}{\|B_{i,p}\|_2}.$$

During the prior elicitation step, the statistician should multiply the basis matrix by an appropriate multiplicative constant depending on the expected magnitude of the measurements at hand. Figure 2.2 illustrates some b-splines on the $(0, 1)$ domain. The Fourier

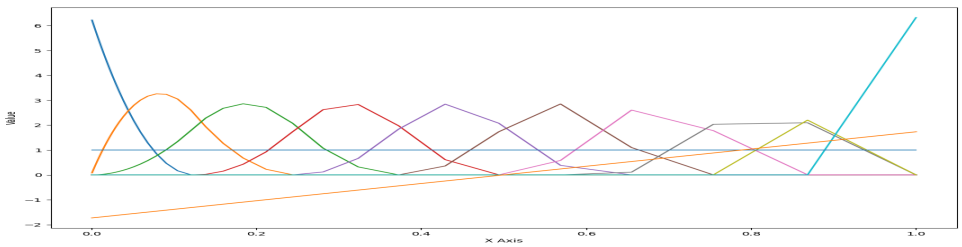


Figure 2.2: B-splines on the $(0, 1)$ domain, with power-law distributed knots. Constant and linear terms are also depicted.

basis functions are essential in decomposing signals into their sinusoidal components. This decomposition is effective in capturing oscillatory behavior and trends across different scales, making it a powerful tool for analyzing cyclical and seasonal phenomena. The implementation of the Fourier basis also relies on the Scikit-FDA package, and by definition, they are already normalized:

$$\phi_0(t) = \frac{1}{\sqrt{T}}, \quad \phi_{2n-1}(t) = \frac{\sin\left(\frac{2\pi n}{T} t\right)}{\sqrt{\frac{T}{2}}}, \quad \phi_{2n}(t) = \frac{\cos\left(\frac{2\pi n}{T} t\right)}{\sqrt{\frac{T}{2}}}$$

In this case as well, when dealing with 2D domains, we rely on tensor products. Figure 2.3 illustrates a sample of columns of the basis matrix of the Fourier in 2D. Another versatile

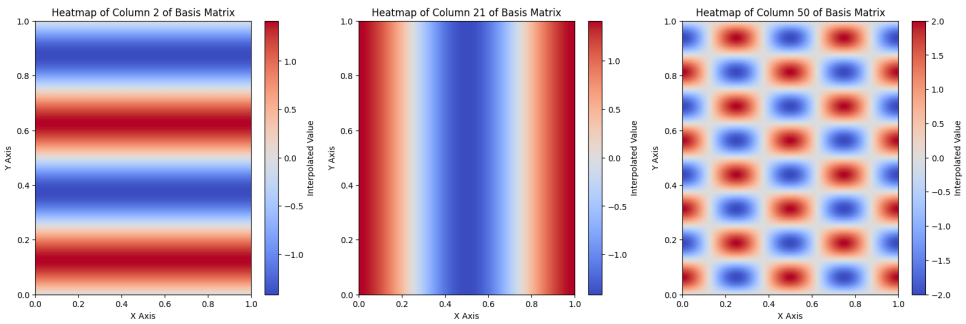


Figure 2.3: Sample of columns of the basis matrix of the Fourier in 2D.

tool is provided by the radial basis functions, defined by a set of 'centers' $\boldsymbol{x}_j$ and a 'radial' function $f : \mathbb{R}^+ \to \mathbb{R}$. In [15] the Latent Factor model was introduced with Gaussian radial basis functions $\exp\{-r^2/(2\rho^2)\}$. Here, we propose the radial bump functions given by

$$\exp\left\{-\frac{1}{1-(r/\rho)^2}\right\} \text{ for } r < \rho, \text{ and } 0 \text{ elsewhere.}$$

Both have the merit of being defined in domains of any dimension, and by appropriately positioning the centers $\boldsymbol{x}_j$, they can cover irregular domains, such as those appearing in fluorescence spectra. The advantage of bump functions is that, similarly to b-splines, they are local. The characterization of the radial bump function we implement in [25] is slightly different from the definition presented here, as each basis function is normalized by its L^2 norm. Examples of equispaced radial bumps functions in the $(0, 1)$ domain are shown in Figure 2.4.

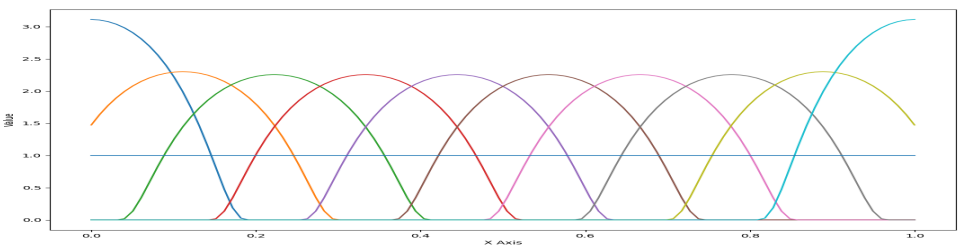


Figure 2.4: Equispaced radial bumps functions in the $(0, 1)$ domain. The normalization causes the bumps hitting the boundary to have higher peaks.

In all cases, our package allows adding constant, linear, and bilinear (in the 2D case)

functions to enrich the basis matrix's ability to describe the data:

$$
\begin{aligned}
\text{Constant}(x) &= \frac{1}{\sqrt{L}}\ , \\
\text{Constant}(x,y) &= \frac{1}{\sqrt{L_x \cdot L_y}} \\
\text{Linear}(x) &= \frac{x - \frac{1}{2}(a+b)}{\sqrt{12}\,(b-a)^{3/2}}\ , \\
\text{Linear}_x(x,y) &= \frac{x - \frac{1}{2}(a+b)}{\sqrt{12}\,(b-a)^{3/2}}\ , \\
\text{Linear}_y(x,y) &= \frac{y - \frac{1}{2}(c+d)}{\sqrt{12}\,(d-c)^{3/2}}\ , \\
\text{Bilinear}(x,y) &= \text{Linear}_x(x,y)\ \cdot\ \text{Linear}_y(x,y).
\end{aligned}
$$

Linear and bilinear entries of a basis matrix in 2D are depicted in Figure 2.5.

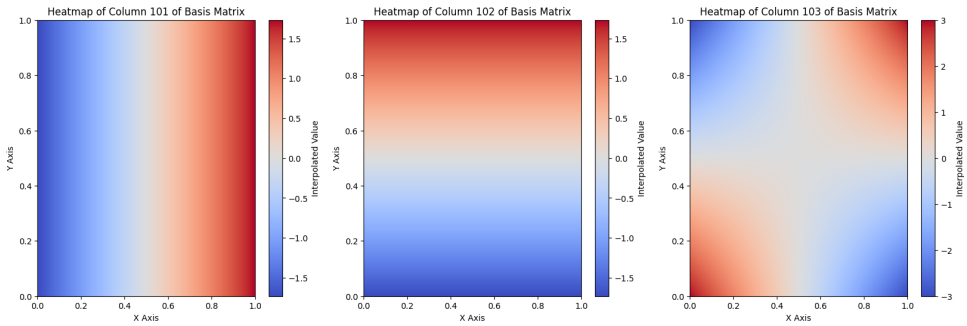


Figure 2.5: Linear and bilinear entries of basis matrix in 2D.

2.2.2. Basis Matrices induced by a Covariance Kernel

Section 3.2 on prior predictive statistical moments will reveal that $\boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)\top}$, augmented by a nugget describing the noise term $\boldsymbol{\varepsilon}_{Y_{\iota,s}}$, forms the prior covariance structure of the BBLF model. This implies that if we aim to establish a specific covariance matrix $\boldsymbol{\Sigma}$ as the a priori covariance structure of our data, we require an effective method to construct a matrix $\boldsymbol{B}$ such that $\boldsymbol{B}\boldsymbol{B}^\top$ closely approximates $\boldsymbol{\Sigma}$.

Initially, Cholesky decomposition was considered; however, it poses an issue: Cholesky decomposition yields a square matrix, which is undesirable as in most practical applications $L_\iota \gg 1$, making the complexity of fitting that many basis coefficients intractable. Downsampling the Cholesky decomposition is not straightforward, as its entries have varying effects across the domain, with the first column affecting all sampling locations and the last column affecting only a single location. If we reduce the dimensionality from the

last column upwards, the resulting basis matrix would have inhomogeneous abilities to describe the variations in the data.

The implemented solution involves applying spectral decomposition to obtain a representation of the form $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\top$, where $\boldsymbol{S}$ is a diagonal matrix with positive, decreasing entries, and $\boldsymbol{U}$ is an orthogonal matrix. This approach is always feasible, as $\boldsymbol{\Sigma}$ is positive definite, as demonstrated in [10].

By selecting:

$$\boldsymbol{B} = \boldsymbol{U}\sqrt{\boldsymbol{S}}$$

we achieve $\boldsymbol{B}\boldsymbol{B}^\top = \boldsymbol{\Sigma}$. Differently from the Cholesky decomposition, it is possible to define a diagonal matrix $\widetilde{\boldsymbol{S}}$ containing the first p_ι entries of $\boldsymbol{S}$, and $\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{S}}$, where $\widetilde{\boldsymbol{U}}$ is generated by the first p_ι columns of $\boldsymbol{U}$. The effect of this truncation is dispersed across the domain, minimally affecting the overall data representation capabilities. Choosing p_ι large enough induces $\boldsymbol{B}\boldsymbol{B}^\top \approx \boldsymbol{\Sigma}$. In practice, the choice can be based on capturing a certain percentage of total variability or by examining the distribution of $\mathrm{diag}(\boldsymbol{S})$, similarly to the approach used in PCA.

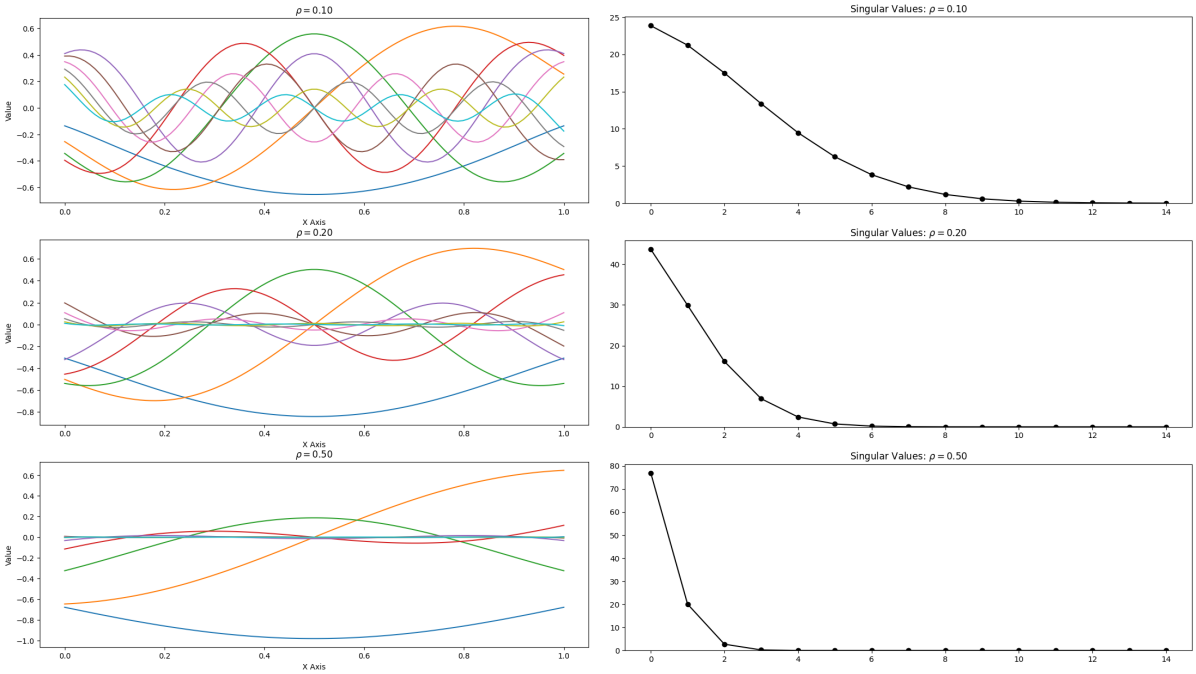
This methodology is particularly versatile, enabling approximation of covariance structures generated by various kernel functions like the Exponential Quadratic kernel and Matérn 3/2 or 5/2 kernels, which are prevalent in Gaussian Process modeling:

$$\begin{aligned}
\mathrm{exp_quad_cov}(x, y; \rho, \sigma_d) &= \sigma_d^2 \exp\left\{-\frac{\|x-y\|_2^2}{2\rho^2}\right\} \\
\mathrm{matern_3/2}(x, y; \rho, \sigma_d) &= \sigma_d^2 \left(1 + \frac{\sqrt{3}\|x-y\|_2}{\rho}\right) \exp\left\{-\frac{\sqrt{3}\|x-y\|_2}{\rho}\right\}
\end{aligned}$$

Figure 2.6 gives a picture of 10 basis functions derived from exponentiated quadratic covariance, whereas Figure 2.7 gives a picture of 10 basis functions derived from Matérn 3/2 covariance kernel. Comparing Figures 2.6 and 2.7, we can notice a slower decrease of the singular values of the Matérn kernel with respect to those of the exponentiated quadratic covariance kernel, due to the lower regularity constraints imposed by the covariance kernel.

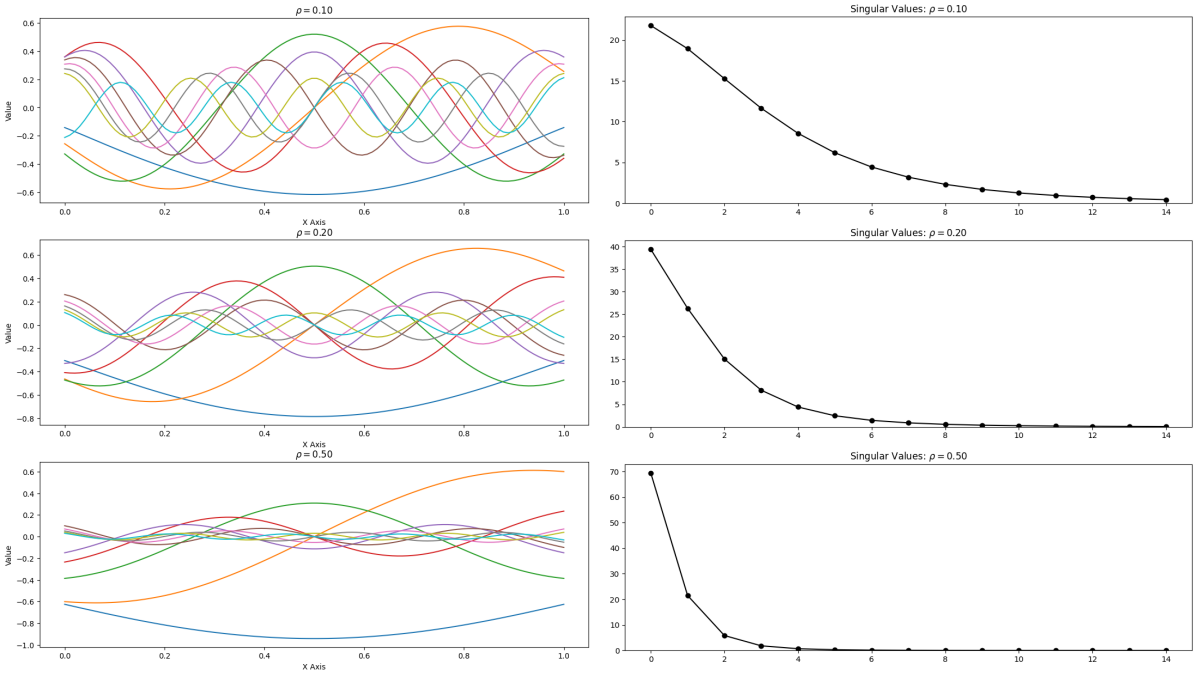
When deciding the appropriate covariance kernel to employ, considerations should mirror those applied in selecting covariance kernels for Gaussian Process fitting. The reader is directed to consult [21] for an in-depth discussion regarding the properties of each covariance function.

An advantage of this technique is that it is not constrained by the locations of measure-



(a) (b)

Figure 2.6: a) 10 basis functions derived from exponentiated quadratic covariance kernel with $\sigma_d = 1.0$ and $\rho = 0.1, 0.2, 0.5$. b) Corresponding distributions of the singular values.



(a) (b)

Figure 2.7: a) 10 basis functions derived from Matérn 3/2 covariance kernel with $\sigma_d = 1.0$ and $\rho = 0.1, 0.2, 0.5$. b) Corresponding distributions of the singular values.

ments, automatically addressing situations with irregular domains; on the other hand, since the approach originates from a known covariance matrix, extrapolation beyond the sampling points without considering future extrapolation locations during the basis matrix construction is not feasible.

In [25], these strategies can be implemented. Examples proving the effectiveness of this approach will be presented in Chapter 4.

2.2.3. Basis Matrices for non-functional targets

The straightforward approach would be using an identity matrix and setting the rate parameter β_ψ, relating to the noise term $\boldsymbol{\varepsilon}_\psi$, equal to zero. However, given that isotropic priors are employed, any set of orthonormal vectors that span the entire target vector space are equally valid from the model's perspective, as they do not impact the target's likelihood. For this reason, we recommend utilizing all components derived from a Principal Component Analysis (PCA). This methodology allows the basis coefficients, θ, to represent the scores of the principal components, while the product $B\theta$ will describe the predictions, giving the statistician an additional tool to interpret the results effectively.

For longitudinal datasets, there might be a temptation to truncate the principal components at a certain threshold of explained variability, selecting a basis matrix comprised of the leading p principal components. Such an approach leads to an empirical Bayes method, as the prior predictive distribution is modified by the data. Despite potential reservations, in this model, the noise term $\boldsymbol{\varepsilon}_\psi$ assumes the role of capturing the residual variability not explained by the principal components. This necessitates careful selection of the hyperparameters α_ψ and β_ψ.

3 | Exploring the BBLF Model: Mathematical Properties and Tools for Posterior Analysis

This chapter delves into exploring the theoretical foundations underpinning the Bivariate Bayesian Latent Factor model. Section 3.1 discusses alternative representations of the likelihood function, showing, first, a formulation that highlights the covariance structure of the targets under the BBLF model, then extracting a computationally efficient expression of the same, showing that a portion of the likelihood function, together with the inverse-gamma prior for the variance of the measurement noise, form a conjugate pair.

During this process, we derive various tools that are useful in exploring the posterior predictive distribution for practical goals, like function-to-function regression and latent factor space projection.

In Section 3.2, the prior predictive statistical moments are discussed, with their derivation deferred to appendix B, illustrating the practical significance of each component. The techniques outlined in Section 3.2 will be revisited and discussed in more practical terms in Section 4.3, elucidating a general approach to prior elicitation for the BBLF model.

3.1. Alternative Representations of the Likelihood Function

We start this section with a brief review of useful tools for matrix manipulation.

From now on, we work under the explicit assumption that the number of $L_\iota \geq p_\iota$ for $\iota = 1, 2$. We indicate with $\boldsymbol{B}^{(\iota)^\dagger}$ the Moore-Penrose pseudo-inverse, and with $\boldsymbol{\mathcal{M}}^{(\iota)}$ the projection matrix onto the orthogonal space to the basis-space $\boldsymbol{I} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}$.

Proposition 3.1. *If the basis matrix $\boldsymbol{B}^{(\iota)}$ has full column-rank, then the following properties hold:*

1. $\boldsymbol{B}^{(\iota)^\dagger} = \left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-1}\boldsymbol{B}^{(\iota)^\top}$
2. $\boldsymbol{B}^{(\iota)^\dagger}\boldsymbol{B}^{(\iota)} = \boldsymbol{I}$
3. $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)} = \boldsymbol{0}$
4. $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)^{\dagger T}} = \boldsymbol{0}$
5. $\boldsymbol{\mathcal{M}}^{(\iota)}$ *is idempotent*
6. $\boldsymbol{\mathcal{M}}^{(\iota)}$ *is symmetric*

Proof.
For proof of 1. the reader is referred to [10].
2. $\boldsymbol{B}^{(\iota)^\dagger}\boldsymbol{B}^{(\iota)} = \left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-1}\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)} = \left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-1}\left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right) = \boldsymbol{I}$.
3. $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)} = \left(\boldsymbol{I} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}\right)\boldsymbol{B}^{(\iota)} = \boldsymbol{B}^{(\iota)} - \boldsymbol{B}^{(\iota)}\boldsymbol{I}$.
4. $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)^{\dagger T}} = \boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)}\left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-T} = \boldsymbol{0}$
5.
$$\begin{aligned}\boldsymbol{\mathcal{M}}^{(\iota)^2} &= \boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\mathcal{M}}^{(\iota)} = \boldsymbol{\mathcal{M}}^{(\iota)}\left(\boldsymbol{I} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}\right) = \boldsymbol{\mathcal{M}}^{(\iota)} - \left(\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{B}^{(\iota)}\right)\boldsymbol{B}^{(\iota)^\dagger} = \\ &= \boldsymbol{\mathcal{M}}^{(\iota)} - \boldsymbol{0}\boldsymbol{B}^{(\iota)^\dagger} = \boldsymbol{\mathcal{M}}^{(\iota)}\end{aligned}$$
6.
$$\begin{aligned}\boldsymbol{\mathcal{M}}^{(\iota)^\top} &= \boldsymbol{I} - \left(\boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}\right)^\top = \boldsymbol{I} - \left(\left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-1}\boldsymbol{B}^{(\iota)^\top}\right)^\top\boldsymbol{B}^{(\iota)^\top} = \\ &= \boldsymbol{I} - \boldsymbol{B}^{(\iota)}\left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-T}\boldsymbol{B}^{(\iota)^\top} = \\ &= \boldsymbol{I} - \boldsymbol{B}^{(\iota)}\left(\boldsymbol{B}^{(\iota)^\top}\boldsymbol{B}^{(\iota)}\right)^{-1}\boldsymbol{B}^{(\iota)^\top} = \boldsymbol{I} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger} = \boldsymbol{\mathcal{M}}^{(\iota)}\end{aligned}$$

□

Proposition 3.2. *The following representations of target ι are equivalent:*

$$\boldsymbol{Y}^{(\iota)} \Longleftrightarrow \left(\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)},\ \boldsymbol{B}^{(\iota)^\dagger}\boldsymbol{Y}^{(\iota)}\right)$$

Proof. The left-hand side can be derived from the right-hand side as follows:

$$\left(\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)}\right) + \boldsymbol{B}^{(\iota)}\left(\boldsymbol{B}^{(\iota)^\dagger}\boldsymbol{Y}^{(\iota)}\right) = \left(\boldsymbol{I} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}\right)\boldsymbol{Y}^{(\iota)} + \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)^\dagger}\boldsymbol{Y}^{(\iota)} = \boldsymbol{Y}^{(\iota)}$$

□

Remark 3.2.1. $\boldsymbol{B}^{(\iota)\dagger}\boldsymbol{Y}^{(\iota)}$ describes the least-squares fit of the basis coefficients of target ι, and $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)}$ is the residual.

The proof shows we can move from one representation to the other by applying a simple linear transformation:

$$\boldsymbol{Y}_s^{(\iota)} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{B}^{(\iota)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}_s^{(\iota)} \\ \boldsymbol{B}^{(\iota)\dagger}\boldsymbol{Y}_s^{(\iota)} \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}_s^{(\iota)} \\ \boldsymbol{B}^{(\iota)\dagger}\boldsymbol{Y}_s^{(\iota)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(\iota)} \\ \boldsymbol{B}^{(\iota)\dagger} \end{bmatrix} \boldsymbol{Y}_s^{(\iota)} \tag{3.1}$$

3.1.1. Marginal Likelihood Function of the Target with respect to the Latent Factors

A pivotal aspect of the BBLF model is its ability to derive a marginalized likelihood expression by integrating out the latent factors $\boldsymbol{\eta}$ and the basis coefficients $\boldsymbol{\theta}$. This marginalized framework, denoted by the likelihood of $\boldsymbol{Y}_s$ given the parameter set $\boldsymbol{\Theta}$, represents the conditional distribution of the observed data as a multivariate normal distribution, whose mean and covariance structure are directly influenced by the basis functions, loading factors, and covariate effects.

This result is derived by recalling the formulation of the likelihood function as an affine transformation of centered normal distributions, given in Equation (2.3), and applying the well-known formula for the expectation and covariance matrix of an affine transformation of a multivariate normal distribution [1]:

$$\boldsymbol{Y}_s | \boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \qquad s = 1, \dots, N \tag{3.2}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{\iota\iota} &= \psi^{(\iota)2}\boldsymbol{I} + \boldsymbol{B}^{(\iota)}\left[\boldsymbol{\Delta}_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)\top}\right]\boldsymbol{B}^{(\iota)\top} \\ \boldsymbol{\Sigma}_{ij} &= \boldsymbol{B}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\Lambda}^{(j)\top}\boldsymbol{B}^{(j)\top} \\ \boldsymbol{\Delta}_\theta^{(\iota)} &= \operatorname{diag}\left(1/\tau_{\theta,1}^{(\iota)}, \dots, 1/\tau_{\theta,p_i}^{(\iota)}\right) \end{aligned}$$

This result is of great both practical and theoretical importance as it allows for the computation of conditional distributions like (without loss of generality):

$$\boldsymbol{Y}_s^{(2)} \big| \boldsymbol{\Theta}, \boldsymbol{Y}_s^{(1)} \sim \mathcal{N}\left(\boldsymbol{\mu}_{X_s}^{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left[\boldsymbol{Y}_s^{(1)} - \boldsymbol{\mu}_{X_s}^{(1)}\right], \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right) \tag{3.3}$$

with $\boldsymbol{\mu}_{X_s}^{(\iota)} := \boldsymbol{B}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}_s$.

We are interested in $\boldsymbol{Y}_s^{(2)}\big|\boldsymbol{Y}_s^{(1)}$, which is obtained using the Monte Carlo Markov samples from the posterior distribution. For instance, the mean will be approximated by:

$$\mathbb{E}\left[\boldsymbol{Y}_s^{(2)}|\boldsymbol{Y}_s^{(1)}\right] \approx \sum_{w=1}^{W} \mathbb{E}\left[\boldsymbol{Y}_s^{(2)}|\boldsymbol{\Theta}_w, \boldsymbol{Y}_s^{(1)}\right]$$

where $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_W$ is a sample from the posterior distribution of $\boldsymbol{\Theta}$.

If we consider the total variances $\mathbf{tr}\left(\boldsymbol{\Sigma}_{\iota\iota}\right)$, it follows immediately that it makes sense to define the R^2 conditional on $\boldsymbol{\Theta}$:

$$R^2_{1\to 2}\big|\boldsymbol{\Theta} := \frac{\mathbf{tr}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|\boldsymbol{\Theta}\right)}{\mathbf{tr}\left(\boldsymbol{\Sigma}_{22}|\boldsymbol{\Theta}\right)} = 1 - \frac{\mathbf{tr}\left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|\boldsymbol{\Theta}\right)}{\mathbf{tr}\left(\boldsymbol{\Sigma}_{22}|\boldsymbol{\Theta}\right)} \tag{3.4}$$

When using this formula, it is important to remember that we are comparing the model to the one that uses only the covariates to predict the target $\boldsymbol{Y}_s^{(2)}$, which means we are computing the fraction of the variance unexplained by the covariates which is explained by $\boldsymbol{Y}_s^{(1)}$.

Marginalizing only the basis coefficients. Applying similar reasoning to the derivation of Equation (3.3), we obtain the 'joint likelihood' of the s-th observation and s-th latent factor $(\boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}, \boldsymbol{\eta}_s)$ given $\boldsymbol{\Theta}$, $s = 1, \ldots, N$:

$$\begin{pmatrix} \boldsymbol{Y}_s^{(1)} \\ \boldsymbol{Y}_s^{(2)} \\ \boldsymbol{\eta}_s \end{pmatrix} \Bigg| \boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{pmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s \\ \boldsymbol{\beta}\boldsymbol{X}_s \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)} \\ \boldsymbol{\Lambda}^{(1)\top}\boldsymbol{B}^{(1)\top} & \boldsymbol{\Lambda}^{(2)\top}\boldsymbol{B}^{(2)\top} & \boldsymbol{I} \end{bmatrix} \right). \tag{3.5}$$

This expression finds applications in projecting new targets on the fitted latent factor space:

$$\boldsymbol{\eta}_s \Big| \boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}, \boldsymbol{\Theta} \sim \text{MultivariateNormal} \tag{3.6}$$

with conditional moments:

$$\mathbb{E}\left[\boldsymbol{\eta}_s \middle| \boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}, \boldsymbol{\Theta}\right] = \boldsymbol{\beta}\boldsymbol{X}_s + \begin{bmatrix} \boldsymbol{\Lambda}^{(1)\top}\boldsymbol{B}^{(1)\top} & \boldsymbol{\Lambda}^{(2)\top}\boldsymbol{B}^{(2)\top} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Y}_s^{(1)} - \boldsymbol{\mu}_{X_s}^{(1)} \\ \boldsymbol{Y}_s^{(2)} - \boldsymbol{\mu}_{X_s}^{(2)} \end{bmatrix}$$

$$\mathbb{C}\text{ov}\left[\boldsymbol{\eta}_s \middle| \boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}, \boldsymbol{\Theta}\right] = \boldsymbol{I} - \begin{bmatrix} \boldsymbol{\Lambda}^{(1)\top}\boldsymbol{B}^{(1)\top} & \boldsymbol{\Lambda}^{(2)\top}\boldsymbol{B}^{(2)\top} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)} \end{bmatrix}$$

3.1.2. Marginal Likelihood Function of the target decomposed in the basis space and its residual

Up to now, we have not discussed computational complexity at all, but most of the questions one might want to ask this model, involve heavy computational lifting: an example is the formula we derived for the conditional distribution $\boldsymbol{Y}_s^{(2)}\big|\boldsymbol{\Theta}, \boldsymbol{Y}_s^{(1)}$, which necessitates the inversion of an $L_\iota \times L_\iota$ square matrix (one per each Monte Carlo sample!).

Similar problems arise in using the Hamiltonian Monte Carlo method implemented in Stan, both in terms of computational and memory complexity.

In the following, we show, that under the assumption of linear independence of the columns of the basis matrices, applying equation 3.1 allows us to move the problem to the space of the basis coefficients, reducing its dimensionality from L_ι to p_ι, which in practice is often taken to be much smaller than L_ι.

Proposition 3.3. *If $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$ are full column rank:*

$$\begin{bmatrix} \boldsymbol{\mathcal{M}}^{(1)}\boldsymbol{Y}_s^{(1)} \\ \boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)} \\ \boldsymbol{\mathcal{M}}^{(2)}\boldsymbol{Y}_s^{(2)} \\ \boldsymbol{B}^{(2)\dagger}\boldsymbol{Y}_s^{(2)} \end{bmatrix} \Bigg| \boldsymbol{\Theta} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s \\ \boldsymbol{0} \\ \boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s \end{bmatrix}, \begin{bmatrix} {\psi^{(1)}}^2\boldsymbol{\mathcal{M}}^{(1)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{\Sigma}}_{11} & \boldsymbol{0} & \boldsymbol{\Lambda}^{(1)}{\boldsymbol{\Lambda}^{(2)}}^\top \\ \boldsymbol{0} & \boldsymbol{0} & {\psi^{(2)}}^2\boldsymbol{\mathcal{M}}^{(2)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}^{(2)}{\boldsymbol{\Lambda}^{(1)}}^\top & \boldsymbol{0} & \widetilde{\boldsymbol{\Sigma}}_{22} \end{bmatrix} \right)$$

where

$$\widetilde{\boldsymbol{\Sigma}}_{\iota\iota} = {\psi^{(\iota)}}^2 \left({\boldsymbol{B}^{(i)}}^\top \boldsymbol{B}^{(i)} \right)^{-1} + \boldsymbol{\Delta}_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(i)}{\boldsymbol{\Lambda}^{(i)}}^\top$$

Proof. From 3.1

$$\begin{bmatrix} \boldsymbol{\mathcal{M}}^{(1)}\boldsymbol{Y}_s^{(1)} \\ \boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)} \\ \boldsymbol{\mathcal{M}}^{(2)}\boldsymbol{Y}_s^{(2)} \\ \boldsymbol{B}^{(2)\dagger}\boldsymbol{Y}_s^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(1)} & \boldsymbol{0} \\ \boldsymbol{B}^{(1)\dagger} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{M}}^{(2)} \\ \boldsymbol{0} & \boldsymbol{B}^{(2)\dagger} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_s^{(1)} \\ \boldsymbol{Y}_s^{(2)} \end{bmatrix}$$

So we apply the formula for affine transformations of multivariate normal distributions to Equation (3.2).

The expression for the expected value is obtained from properties 2. and 3. of the Moore-Penrose pseudo-inverse when $\boldsymbol{B}^{(\iota)}$ has full column rank (Proposition 3.1).

In the computation of the covariance matrix, two kinds of terms appear:

$$\begin{bmatrix}\boldsymbol{\mathcal{M}}^{(\iota)}\\ \boldsymbol{B}^{(i)\dagger}\end{bmatrix}\boldsymbol{\Sigma}_{\iota\iota}\begin{bmatrix}\boldsymbol{\mathcal{M}}^{(\iota)\top} & \boldsymbol{B}^{(i)\dagger\top}\end{bmatrix}, \qquad \begin{bmatrix}\boldsymbol{\mathcal{M}}^{(\iota)}\\ \boldsymbol{B}^{(i)\dagger}\end{bmatrix}\boldsymbol{\Sigma}_{ij}\begin{bmatrix}\boldsymbol{\mathcal{M}}^{(j)\top} & \boldsymbol{B}^{(j)\dagger\top}\end{bmatrix}$$

We now discuss each term separately:

- $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\Sigma}_{\iota\iota}\boldsymbol{\mathcal{M}}^{(\iota)\top} \overset{(3)}{=} \psi^{(\iota)2}\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\mathcal{M}}^{(\iota)\top} \overset{(5)}{=} \psi^{(\iota)2}\boldsymbol{\mathcal{M}}^{(\iota)}$
- $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\Sigma}_{\iota\iota}\boldsymbol{B}^{(i)\dagger\top} = \boldsymbol{B}^{(i)\dagger}\boldsymbol{\Sigma}_{\iota\iota}\boldsymbol{\mathcal{M}}^{(\iota)\top} = \boldsymbol{0}$ by properties 3. and 4.
- $\begin{aligned}\boldsymbol{B}^{(i)\dagger}\boldsymbol{\Sigma}_{\iota\iota}\boldsymbol{B}^{(i)\dagger\top} &\overset{(2)}{=} \psi^{(\iota)2}\boldsymbol{B}^{(i)\dagger}\boldsymbol{B}^{(i)\dagger\top} + \left[\boldsymbol{\Delta}_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(i)}\boldsymbol{\Lambda}^{(i)\top}\right]\\ &= \psi^{(\iota)2}\left(\boldsymbol{B}^{(i)\top}\boldsymbol{B}^{(i)}\right)^{-1} + \boldsymbol{\Delta}_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(i)}\boldsymbol{\Lambda}^{(i)\top}\end{aligned}$
- $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\Sigma}_{ij}\boldsymbol{\mathcal{M}}^{(j)} = \boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{\Sigma}_{ij}\boldsymbol{B}^{(j)\dagger\top} = \boldsymbol{B}^{(i)\dagger}\boldsymbol{\Sigma}_{ij}\boldsymbol{\mathcal{M}}^{(j)\top} = \boldsymbol{0}$ by 3.
- $\boldsymbol{B}^{(i)\dagger}\boldsymbol{\Sigma}_{ij}\boldsymbol{B}^{(j)\dagger} = \boldsymbol{\Lambda}^{(i)}\boldsymbol{\Lambda}^{(j)\top}$

□

Remark 3.3.1. The likelihood of $\boldsymbol{Y}$ can, therefore, be decomposed in the two independent terms:

$$\begin{bmatrix}\boldsymbol{\mathcal{M}}^{(1)}\boldsymbol{Y}_s^{(1)}\\ \boldsymbol{\mathcal{M}}^{(2)}\boldsymbol{Y}_s^{(2)}\end{bmatrix}\Bigg|\boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\psi^{(1)2}\boldsymbol{\mathcal{M}}^{(1)} & \boldsymbol{0}\\ \boldsymbol{0} & \psi^{(2)2}\boldsymbol{\mathcal{M}}^{(2)}\end{bmatrix}\right) \tag{3.7}$$

and

$$\begin{bmatrix}\boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}\\ \boldsymbol{B}^{(2)\dagger}\boldsymbol{Y}_s^{(2)}\end{bmatrix}\Bigg|\boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s\\ \boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s\end{bmatrix}, \boldsymbol{\Sigma}^\star\right) \tag{3.8}$$

with

$$\boldsymbol{\Sigma}^\star = \text{block_diag}\begin{bmatrix}\psi^{(1)2}\left(\boldsymbol{B}^{(1)\top}\boldsymbol{B}^{(1)}\right)^{-1} + \boldsymbol{\Delta}_\theta^{(1)}\\ \psi^{(2)2}\left(\boldsymbol{B}^{(2)\top}\boldsymbol{B}^{(2)}\right)^{-1} + \boldsymbol{\Delta}_\theta^{(2)}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Lambda}^{(1)}\\ \boldsymbol{\Lambda}^{(2)}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}^{(1)\top} & \boldsymbol{\Lambda}^{(2)\top}\end{bmatrix}.$$

Here 3.7 should be considered as a degenerate normal distribution; the proof of Proposition 3.4 clarifies it.

Remark 3.3.2. Let $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$ be full column rank, and let $\boldsymbol{Q}^{(\iota)}, \boldsymbol{R}^{(\iota)}$ be the thin-QR decomposition of $\boldsymbol{B}^{(\iota)}$ (so $\boldsymbol{R}^{(\iota)}$ is an invertible square matrix). Then:

$$\begin{aligned}\boldsymbol{B}^{(\iota)\dagger}\boldsymbol{Y}_s^{(\iota)} &= \boldsymbol{\theta}_s^{(\iota)} + \psi^{(\iota)}\boldsymbol{R}^{(\iota)^{-1}}\widetilde{\boldsymbol{\varepsilon}}^{(\iota)}\\ \widetilde{\boldsymbol{\varepsilon}}^{(\iota)}|\boldsymbol{\Theta}, \boldsymbol{\theta}_s^{(\iota)}, \boldsymbol{\eta}_s &\sim \mathcal{N}(0,\,1), \qquad \text{for } \iota = 1, 2\end{aligned}$$

is an equivalent formulation of the likelihood of $\left(\boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)},\ \boldsymbol{B}^{(2)\dagger}\boldsymbol{Y}_s^{(2)}\right)$.

Proposition 3.4. *If $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$ are full column rank:*

$$\mathrm{RSS}_\theta^{(1)} := \sum_{s=1}^{N}\sum_{h=1}^{L_1}\left(\boldsymbol{\mathcal{M}}^{(1)}\boldsymbol{Y}^{(1)}\right)^2_{h,\,s}, \quad \mathrm{RSS}_\theta^{(2)} := \sum_{s=1}^{N}\sum_{h=1}^{L_2}\left(\boldsymbol{\mathcal{M}}^{(2)}\boldsymbol{Y}^{(2)}\right)^2_{h,\,s}$$

are sufficient statistics for $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)}$ in the BBLF model.

Proof. Once fixed the matter of $\boldsymbol{\mathcal{M}}^{(\iota)}$, we are in the case of independently identically distributed normals with known mean (0) and unknown variance [9].

$\boldsymbol{\mathcal{M}}^{(\iota)}$ is nilpotent, and, in particular, it has $L_\iota - p_\iota$ eigenvalues equal to one, and p_ι equal to zero (only possiblities for a nilpotent matrix [10]). That $\dim\left\{\mathrm{Ker}\,\boldsymbol{\mathcal{M}}^{(\iota)}\right\} \geq p_\iota$ is proven by 3. To prove the equality, let $\mathbf{0} \neq \boldsymbol{v} \perp \boldsymbol{B}^{(\iota)}$, then $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{B}^{(\iota)}\boldsymbol{B}^{(\iota)\dagger}\boldsymbol{v} = \boldsymbol{v}$ by 1. and the orthogonality of $\boldsymbol{v}$.

Notice also that $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}_s^{(\iota)}$ is in the image of $\boldsymbol{\mathcal{M}}^{(\iota)}$, which, for symmetric matrices, corresponds to the vector space orthogonal to the kernel. In our case the eigen-space associated to eigenvalue $\lambda = 1$. □

Remark 3.4.1. *The inverse-gamma priors for ${\psi^{(\iota)}}^2$ are conjugate to $\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)}$:*

$${\psi^{(\iota)}}^2|\boldsymbol{\mathcal{M}}^{(\iota)}\boldsymbol{Y}^{(\iota)} \sim \text{Inv-Ga}\left(\alpha_\psi + \frac{N\cdot(L_\iota - p_\iota)}{2},\ \beta_\psi^{(\iota)} + \frac{RSS_\theta^{(\iota)}}{2}\right)$$

3.1.3. Other relations using the Basis Space-Residual decomposition

In this section, we show how the mathematical framework we have devised, particularly for characterizing likelihood within the vector space of basis coefficients, facilitates the computation of quantities previously outlined in Section 3.1.1.

Conditional Distribution. Decomposing only target 1:

$$\begin{bmatrix}\boldsymbol{\mathcal{M}}^{(1)}\boldsymbol{Y}_s^{(1)}\\ \boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}\\ \boldsymbol{Y}_s^{(2)}\end{bmatrix}\Bigg|\boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s\\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s\end{bmatrix}, \begin{bmatrix}{\psi^{(1)}}^2\boldsymbol{\mathcal{M}}^{(1)} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \widetilde{\boldsymbol{\Sigma}}_{11} & \boldsymbol{\Lambda}^{(1)}{\boldsymbol{\Lambda}^{(2)}}^\top{\boldsymbol{B}^{(2)}}^\top\\ \mathbf{0} & \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}{\boldsymbol{\Lambda}^{(1)}}^\top & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

therefore the prediction of $\boldsymbol{Y}_s^{(2)}$ depends only on $\boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}$, $\boldsymbol{Y}_s^{(2)}|\boldsymbol{\Theta}, \boldsymbol{Y}_s^{(1)}$ is multivariate normal, and:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{Y}_s^{(2)}\middle|\,\boldsymbol{\Theta},\, \boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}\right] &= \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\left[\boldsymbol{\beta}\boldsymbol{X}_s + \boldsymbol{\Lambda}^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\left(\boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)} - \boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s\right)\right] \\
\mathbb{C}\,\mathrm{ov}\left[\boldsymbol{Y}_s^{(2)}\middle|\,\boldsymbol{\Theta},\, \boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}\right] &= \boldsymbol{\Sigma}_{22} - \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Lambda}^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Lambda}^{(2)\top}\boldsymbol{B}^{(2)\top}
\end{aligned}$$

We see the advantage of the decomposition, as $\widetilde{\boldsymbol{\Sigma}}_{11}$ is an easier to invert $p_\iota \times p_\iota$ matrix.

Coefficient of Determination. Without loss of generality, we plug into:

$$R^2_{1\to 2}|\boldsymbol{\Theta} := \frac{\mathbf{tr}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|\boldsymbol{\Theta}\right)}{\mathbf{tr}\left(\boldsymbol{\Sigma}_{22}|\boldsymbol{\Theta}\right)}$$

the following substitutions:

$$\begin{aligned}
\mathbf{tr}\left(\boldsymbol{\Sigma}_{22}|\boldsymbol{\Theta}\right) &= \mathbf{tr}\left({\psi^{(2)}}^2\boldsymbol{\mathcal{M}}^{(2)}\right) + \mathbf{tr}\left(\boldsymbol{B}^{(2)}\widetilde{\boldsymbol{\Sigma}}_{22}\boldsymbol{B}^{(2)\top}|\boldsymbol{\Theta}\right) = \qquad (3.9)\\
&= (L_2 - p_2){\psi^{(2)}}^2 + \mathbf{tr}\left(\boldsymbol{B}^{(2)}\widetilde{\boldsymbol{\Sigma}}_{22}\boldsymbol{B}^{(2)\top}|\boldsymbol{\Theta}\right)\\
\mathbf{tr}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|\boldsymbol{\Theta}\right) &= \mathbf{tr}\left(\boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Lambda}^{(1)\top}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Lambda}^{(2)\top}\boldsymbol{B}^{(2)\top}|\boldsymbol{\Theta}\right) \qquad (3.10)
\end{aligned}$$

Projection on Latent Factor Space. We already know the relevant components will be the projections on the basis space, while the residuals won't have any informative content. Similarly to 3.8:

$$\begin{bmatrix}\boldsymbol{B}^{(1)\dagger}\boldsymbol{Y}_s^{(1)}\\ \boldsymbol{B}^{(2)\dagger}\boldsymbol{Y}_s^{(2)}\\ \boldsymbol{\eta}_s\end{bmatrix}\Bigg|\boldsymbol{\Theta} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}\boldsymbol{X}_s\\ \boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}\boldsymbol{X}_s\\ \boldsymbol{\beta}\boldsymbol{X}_s\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11}^\star & \boldsymbol{\Sigma}_{12}^\star & \boldsymbol{\Lambda}^{(1)}\\ \boldsymbol{\Sigma}_{21}^\star & \boldsymbol{\Sigma}_{22}^\star & \boldsymbol{\Lambda}^{(2)}\\ \boldsymbol{\Lambda}^{(1)\top} & \boldsymbol{\Lambda}^{(2)\top} & \boldsymbol{I}\end{bmatrix}\right) \qquad (3.11)$$

from which $\boldsymbol{\eta}_s|\boldsymbol{\Theta}, \boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}$ is multivariate normal, with formulas for expectation and covariance matrix derived the same way as before.

3.2. Prior Predictive Statistical Moments

The proofs of all the results in this subsection are postponed in Appendix B.

Otherwise explicitly stated, the presented results in this section hold under the assumption of $\alpha_2, \nu, \chi > 2$ and $\alpha_1, \alpha_\sigma, \alpha_\psi > 1$.

Expected Value. As all quantities that enter in the likelihood model are centered, the expected value of the prior predictive is 0:

$$\mathbb{E}\left[\boldsymbol{Y}_s\right] = \mathbf{0}$$

Total Variance. We define the total variance Σ_T^k as the trace of the covariance matrix $\mathbb{C}\mathrm{ov}\left[\boldsymbol{Y}_s\right]$, where the number of latent factors k is explicitly indicated, while Σ_T^∞ indicates the limit as $k \to \infty$ which is shown to exist under the assumptions of this section.

The total variance can be expressed as:

$$\Sigma_{T,s}^k = \sigma_{\mathcal{WN}}^2 + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^\top\right)\left[\sigma_{\mathcal{I}}^2 + \rho_\Lambda^{(k)}\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}\left\|\boldsymbol{X}_{\cdot,s}\right\|_2^2\right)\right]$$

We also introduce the Total Variance explained by the latent factors:

$$\Sigma_{T\Lambda,s}^k = \rho_\Lambda^{(k)}\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^\top\right)\left[\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}\left\|\boldsymbol{X}_{\cdot,s}\right\|_2^2\right)\right]$$

where we introduced the following quantities:

- the Prior Observation Variance $\sigma_{\mathcal{WN}}^2 = \frac{L_1\beta_\psi^{(1)} + L_2\beta_\psi^{(2)}}{\alpha_\psi - 1}$,
- the Trace of the Hat Matrix $\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^\top\right)$,
- the Idiosyncratic Functional Variance $\sigma_{\mathcal{I}}^2 = \frac{\beta_\sigma}{\alpha_\sigma - 1}$,
- $\rho_\Lambda^{(k)} = \Sigma_{T\Lambda}^k / \Sigma_{T\Lambda}^\infty = 1 - \left(\frac{1}{\alpha_2 - 1}\right)^k$ the portion of latent factor's variance, explained by the first k latent factors,
- the variance of the latent factors $\sigma_\Lambda^2 = \frac{\nu}{\nu-2}\frac{1}{\alpha_1-1}\frac{\alpha_2-1}{\alpha_2-2}$,

To complete the analysis, we assume the covariates are whitened:

$$\boldsymbol{X}_{\cdot,s} \sim \mathcal{N}\left(0, V_X\boldsymbol{I}\right)$$

$$\Sigma_T^k := \lim_{N\to\infty}\frac{1}{N}\sum_{s=1}^N \Sigma_{T,s}^k = \sigma_{\mathcal{WN}}^2 + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^\top\right)\left[\sigma_{\mathcal{I}}^2 + \rho_\Lambda^{(k)}\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}rV_X\right)\right]$$

$$\Sigma_{T\Lambda}^k := \lim_{N\to\infty}\frac{1}{N}\sum_{s=1}^N \Sigma_{T\Lambda,s}^k = \rho_\Lambda^{(k)}\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^\top\right)\left[\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}rV_X\right)\right]$$

Covariance Matrix. The expression of the Covariance matrix is greatly simplified by defining

$$\sigma_{\mathcal{F}}^2 := \sigma_{\mathcal{I}}^2 + \rho_{\Lambda}^{(k)} \sigma_{\Lambda}^2 \left(1 + \frac{\chi}{\chi - 2} \left\| \boldsymbol{X}_{\cdot,s} \right\|_2^2 \right)$$

which allows for the following matrix block representation:

$$\boldsymbol{\Sigma}_s^k = \frac{1}{\alpha_\psi - 1} \begin{bmatrix} \beta_\psi^{(1)} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \beta_\psi^{(2)} \boldsymbol{I} \end{bmatrix} + \sigma_{\mathcal{F}}^2 \begin{bmatrix} \boldsymbol{B}^{(1)} \boldsymbol{B}^{(1)\top} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{(2)} \boldsymbol{B}^{(2)\top} \end{bmatrix}$$

where the dependence on the covariates of the sample is hidden inside $\sigma_{\mathcal{F}}^2$.

We define, for $h_1 \neq h_2$, the correlation:

$$\mathbb{C}\mathrm{orr}_{h_1,h_2}^{(\iota)}(s) = \frac{\mathbb{C}\mathrm{ov}\left[\boldsymbol{Y}_{h_1,s}^{(\iota)}, \boldsymbol{Y}_{h_2,s}^{(\iota)}\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h_1,s}^{(\iota)}\right] \mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h_2,s}^{(\iota)}\right]}} =$$

$$= \left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_1,h_2} \left[\frac{\frac{\beta_\psi}{\alpha_\psi - 1}}{\sigma_{\mathcal{F}}^2} + \left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_1,h_1}\right]^{-1/2} \left[\frac{\frac{\beta_\psi}{\alpha_\psi - 1}}{\sigma_{\mathcal{F}}^2} + \left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_2,h_2}\right]^{-1/2}$$

Under the assumption of small white noise, the effect of the covariates on the correlation is vanishingly small:

$$\frac{\beta_\psi}{\alpha_\psi - 1} \ll \sigma_{\mathcal{F}}^2 \quad \Longrightarrow \quad \mathbb{C}\mathrm{orr}_{h_1,h_2}^{(\iota)}(s) \approx \frac{\left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_1,h_2}}{\sqrt{\left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_1,h_1} \left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_2,h_2}}}$$

Between-sample Covariance.
Let $s_1 \neq s_2$,

$$\mathbb{C}\mathrm{ov}\left[\boldsymbol{Y}_{h_1,s_1}^{(\iota_1)}, \boldsymbol{Y}_{h_2,s_2}^{(\iota_2)}\right] = \begin{cases} \frac{\chi}{\chi-2} \rho_{\Lambda}^{(k)} \sigma_{\Lambda}^2 \left(\boldsymbol{X}_{\cdot,s_1}^\top \boldsymbol{X}_{\cdot,s_2}\right) \left(\boldsymbol{B}^{(\iota)} \boldsymbol{B}^{(\iota)\top}\right)_{h_1,h_2} & \text{if } \iota := \iota_1 = \iota_2 \\ 0 & \text{if } \iota_1 \neq \iota_2 \end{cases}$$

3.3. Interpretability of Latent Factors: the Varimax-RSP algorithm

In this section, we discuss the implementation details of the algorithms used to explore the posterior distribution of the latent components of the proposed Bayesian Bivariate

Latent Factor model.

The Bivariate Bayesian Latent Factor model has issues identifying the latent components due to the equivalence of the likelihood functions up to a rotation plus a signed permutation. Indeed, let $\boldsymbol{R}$ be a rotation matrix and $\boldsymbol{P}_S$ a signed permutation matrix, then:

$$\begin{aligned}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta} &= \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{R}\boldsymbol{R}^\top\boldsymbol{\beta} = \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{R}\boldsymbol{P}_S\boldsymbol{P}_S^\top\boldsymbol{R}^\top\boldsymbol{\beta} = \widetilde{\boldsymbol{\Lambda}}^{(\iota)}\widetilde{\boldsymbol{\beta}} \\ \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)^\top} &= \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{R}\boldsymbol{R}^\top\boldsymbol{\Lambda}^{(\iota)^\top} = \boldsymbol{\Lambda}^{(\iota)}\boldsymbol{R}\boldsymbol{P}_S\boldsymbol{P}_S^\top\boldsymbol{R}^\top\boldsymbol{\Lambda}^{(\iota)^\top} = \widetilde{\boldsymbol{\Lambda}}^{(\iota)}\widetilde{\boldsymbol{\Lambda}}^{(\boldsymbol{\iota})^\top}\end{aligned}$$

This issue was solved by Papastamoulis and Ntzoufras in [18]. Here we describe the algorithm highlighting how it is applied to the BBLF model.

The Varimax criterion consists of rotating the latent factors to optimize the Varimax objective function:

$$\text{maximize} \quad \frac{1}{4}\sum_{\ell=1}^{k}\left[\left(\sum_{i=1}^{p}\widetilde{\Lambda}_{i\ell}^4\right) - \frac{1}{p}\left(\sum_{i=1}^{p}\widetilde{\Lambda}_{i\ell}^2\right)^2\right]$$

where $\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\boldsymbol{R}$ and $\boldsymbol{R}$ is a rotation matrix. In the context of the Bayesian Bivariate Latent Factor model, we stack the latent component's matrices $\boldsymbol{\Lambda} = \begin{bmatrix}\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{\Lambda}^{(2)}\end{bmatrix}$ and define $p = p_1 + p_2$.

Denote $\widetilde{\boldsymbol{\Lambda}}_w$ for $w = 1, \ldots, W$ the Monte Carlo Markov Chain samples of the latent components after the application of the Varimax criterion to each of them. The result of this operation positions all samples about the same mode, modulo a signed permutation.

To align all samples on the same mode the algorithm continues by applying an appropriate signed permutation to each Monte Carlo sample. The correct signed permutation $\boldsymbol{P}_{Sw}$ can be derived by means of the iterative algorithm.

Such a procedure is presented in Algorithm 3.1. The signed permutations of all the Monte Carlo samples are initialized to the identity matrix, then the average latent component is computed $\boldsymbol{\Lambda}^\star = \frac{1}{W}\sum_{w=1}^{W}\widetilde{\boldsymbol{\Lambda}}_w\boldsymbol{S}_w\boldsymbol{P}_w$. The entire algorithm revolves around the iterative change of signs and permutations of the latent components to minimize the distance from the average latent component. For each Monte Carlo sample, the for-cycle in Lines $11-16$ computes the best permutation per each of the possible 2^k choice of the signs, then stores the one with the minimal cost function. The signed permutations obtained this way are then used to update the value of $\boldsymbol{\Lambda}^\star$. The algorithm repeats until convergence is achieved.

For the Varimax problem, multiple solutions are provided in the literature. For the devel-

Algorithm 3.1 Iterative Signed Permutation Optimizer

$\boldsymbol{S} \leftarrow$ list of W identity matrices eye(k)
$\boldsymbol{P} \leftarrow$ list of W identity matrices eye(k)
$\boldsymbol{\Lambda}^\star \leftarrow \frac{1}{W}\sum_{w=1}^{W} \widetilde{\boldsymbol{\Lambda}}_w \boldsymbol{S}_w \boldsymbol{P}_w$
$C_0 \leftarrow +\infty$
$C \leftarrow 0$
while $C_0 \neq C$ **do**
  $C_0 \leftarrow C$
  $C \leftarrow 0$
  for $w \leftarrow 1, \dots, W$ **do**
    $\overline{\boldsymbol{S}} \leftarrow$ eye(k), $\overline{\boldsymbol{P}} \leftarrow$ eye(k), $\overline{C} \leftarrow +\infty$
    for $\boldsymbol{S}^\star \leftarrow$ all possible diagonal matrices of $\{-1, +1\}$ **do**
      $C^\star, \boldsymbol{P}^\star \leftarrow$ optimal_permutation $\left(\|\widetilde{\boldsymbol{\Lambda}}_w \boldsymbol{S}^\star \boldsymbol{P} - \boldsymbol{\Lambda}^\star\|^2\right)$
      if $C^\star < \overline{C}$ **then**
        $\overline{\boldsymbol{S}} \leftarrow \boldsymbol{S}^\star$, $\overline{\boldsymbol{P}} \leftarrow \boldsymbol{P}^\star$, $\overline{C} \leftarrow C^\star$
      end if
    end for
    $\boldsymbol{S}_w \leftarrow \overline{\boldsymbol{S}}$, $\boldsymbol{P}_w \leftarrow \overline{\boldsymbol{P}}$, $C \leftarrow C + \overline{C}$
  end for
  $\boldsymbol{\Lambda}^\star \leftarrow \frac{1}{W}\sum_{w=1}^{W} \widetilde{\boldsymbol{\Lambda}}_w \boldsymbol{S}_w \boldsymbol{P}_w$
end while
return $\boldsymbol{S}_w, \boldsymbol{P}_w$

oped Python package [25] we rely on the implementation provided by the Python package sklearn [19]. Similarly, for the optimal permutation problem used in Algorithm 3.1, an analytic solution is known, we leverage the implementation provided in the Python package SciPy [28], specifically with the procedure `scipy.optimize.linear_sum_assignement`.

In the following chapter, we consider datasets generated by sampling from the prior distribution itself with the goal of validating the ability of the model to fit the data. To compare the *true* latent components to the posterior distribution after the application of Varimax, we need to apply Varimax to the *true* latent components too, and then align the result to the Monte Carlo samples obtained using the Varimax-RSP algorithm. To do so, we compute $\boldsymbol{\Lambda}^\star = \frac{1}{W}\sum_{w=1}^{W} \widetilde{\boldsymbol{\Lambda}}_w \boldsymbol{S}_w \boldsymbol{P}_w$ and then apply the same algorithm as before to the *true* latent components $\overline{\boldsymbol{\Lambda}}$. The difference in this context is that we do not need to iterate the algorithm as a single iteration yields the correct answer immediately. Algorithm 3.2 shows these processes in more detail.

Due to the complexity of the Varimax-RSP algorithm, a check on the calibration of the resulting Monte Carlo draws has been performed. The results are reported in Appendix C.

Algorithm 3.2 Varimax-RSP for *true* Latent Factors

$\boldsymbol{\Lambda}^\star \leftarrow \frac{1}{W}\sum_{w=1}^{W} \widetilde{\boldsymbol{\Lambda}}_w \boldsymbol{S}_w \boldsymbol{P}_w$
$\overline{\boldsymbol{\Lambda}}^\star \leftarrow$ Varimax optimization of $\overline{\boldsymbol{\Lambda}}$
$\overline{\boldsymbol{S}} \leftarrow \mathrm{eye}(k)$, $\overline{\boldsymbol{P}} \leftarrow \mathrm{eye}(k)$, $\overline{C} \leftarrow +\infty$
for $\boldsymbol{S}^\star \leftarrow$ all possible diagonal matrices of $\{-1, +1\}$ **do**
  $C^\star, \boldsymbol{P}^\star \leftarrow \mathrm{optimal_permutation}\left(\|\overline{\boldsymbol{\Lambda}}^\star \boldsymbol{S}^\star \boldsymbol{P} - \boldsymbol{\Lambda}^\star\|^2\right)$
  if $C^\star < \overline{C}$ **then**
    $\overline{\boldsymbol{S}} \leftarrow \boldsymbol{S}^\star$, $\overline{\boldsymbol{P}} \leftarrow \boldsymbol{P}^\star$, $\overline{C} \leftarrow C^\star$
  end if
end for
return $\overline{\boldsymbol{S}}, \overline{\boldsymbol{P}}$

4 | Validation of the BBLF Model with Toy Datasets

In this chapter, we delve into the empirical validation of the Bivariate Bayesian Latent Factor model. The investigation is structured around distinct sections, each tailored to showcase the model's versatility and robustness across various data types and regression scenarios.

First, in Sections 4.1 and 4.2, we explore the model's performance when applied to datasets simulated from the model's prior distribution, examining both one-dimensional and two-dimensional cases. This self-validation technique illustrates the model's ability to recover underlying structures from the data it is designed to generate.

Next, we focus on a more practical application, employing the model in a function-to-vector regression task. Section 4.3 presents the Biscuit Dataset from the fds package in CRAN [23], serving as a benchmark to evaluate the model's effectiveness in extracting meaningful insights from complex functional data.

In [25], the Jupyter Notebooks used to produce the examples discussed here are provided for replicability.

4.1. Dataset Generated from Prior Distribution: The 1D Case

We simulated a dataset by sampling from the prior distribution of the BBLF model. The choice for the basis function was that of 10 cubic b-splines for $\boldsymbol{Y}^{(1)}$, and Fourier basis truncated to the third harmonic, plus a linear trend for $\boldsymbol{Y}^{(2)}$, as discussed in Section 2.2.1. We set $k = 4$ latent factors; 5 covariates per data point were drawn independently from a normal distribution with a standard deviation of 0.3.

We sampled $N = 150$ data points $\boldsymbol{Y}$, split into a training set of 100 samples and a test set containing the remaining 50. Observations for $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ were made on $L_1 = L_2 = 40$

evenly spaced points over the intervals $(-1, 1)$ and $(0, 5)$, respectively. We set $\alpha_1 = 2.1$, $\alpha_2 = 2.5$, $\nu = \chi = 4$.

The posterior Monte Carlo Markov Chain samples are taken using Hamiltonian Monte Carlo No U-turn sampler [24], with 4 chains, a burn-in of 500 samples, maximum tree depth of 12. 1000 samples were drawn per chain.

Figure 4.1 depicts the posterior predictive distribution conditioned on $\boldsymbol{Y}^{(1)}$, showing a smaller uncertainty (shaded area) compared to $\boldsymbol{Y}^{(2)}|\boldsymbol{Y}$ on the left. Additionally, the mean more accurately reflects the observed data. The Root Mean Square Error provided by the BBLF model estimated on the test set is:

$$\begin{aligned}
&\mathrm{RMSE}\big[\tilde{\boldsymbol{Y}}^{(2)}\big] = \sigma\big[\tilde{\boldsymbol{Y}}^{(2)}\big] \simeq 3.9,\\
&\mathrm{RMSE}\big[\tilde{\boldsymbol{Y}}^{(2)}\big|\tilde{\boldsymbol{X}}\big] \simeq 2.93,\\
&\mathrm{RMSE}\big[\tilde{\boldsymbol{Y}}^{(2)}\big|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}^{(1)}\big] \simeq 2.38.
\end{aligned}$$

where we used the approximation symbol to denote that we are considering an estimate of the quantity resulting from a Monte Carlo computation. These observations and the reported RMSE values confirm that predictions improve as additional information is integrated.

As discussed in Section 3.3, for datasets generated from the prior distribution, we have a

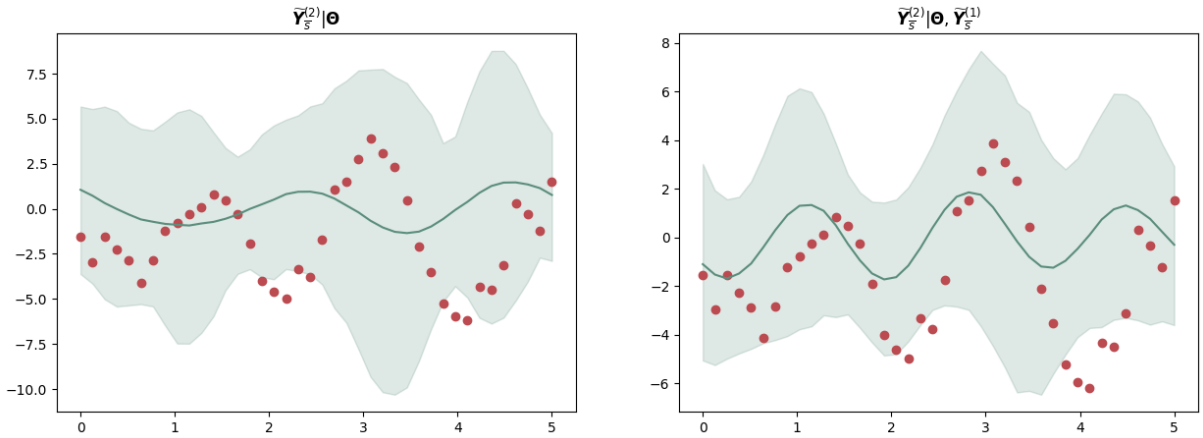


Figure 4.1: The solid lines and the shaded areas represent the mean and the 90% percentile intervals of the posterior predictive distribution for a sample $\bar{s}$ of the test set. Posterior predictive distribution (left), and conditional on $\widetilde{\boldsymbol{Y}}_{\bar{s}}^{(1)}$ (right). The solid dots indicate the observed values of $\widetilde{\boldsymbol{Y}}_{s}^{(2)}$.

true latent component, modulo a rotation, and a signed permutation. Using the Varimax-RSP algorithm, we can compare the *true* latent components with those fitted by the BBLF

model. The results are shown in Figure 4.2. Most of the true values lie within the respec-

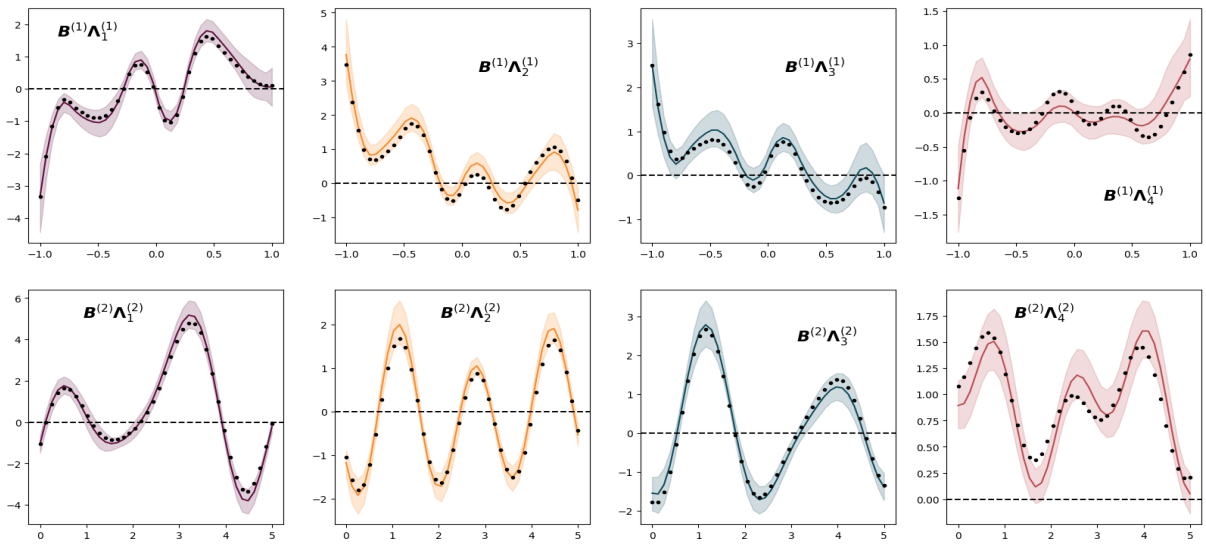


Figure 4.2: The solid lines and the shaded areas represent the mean and the 90% percentile intervals of the latent components sampled from the posterior distribution, after the application of the Varimax-RSP algorithm. The dots are the values of the 'true' latent components, once rotated and permuted to align to the Monte Carlo samples. To simplify the interpretation, the components have been multiplied by the respective basis matrices. The top graphs relate to the first target, bottom graphs to the second.

tive 90% credibility regions, corroborating the idea that the Varimax-RSP method can identify the true latent components.

Figure 4.3 illustrates the technique's robustness with respect to the number of components k. Indeed, we see in Figure 4.3 that two of the latent components have been given low statistical significance, and the remaining components (modulo a signed permutation) fit the correct latent components.

A different approach to performing inference on the number of loading factors is using posterior predictive checks. We propose the use of the Pareto Smoothed Importance Sampling Leave-One-Out (PSIS-LOO) method [27]. The estimated Expected Log pointwise Predictive Density (ELPD) confirmed $k = 4$ latent factors as optimal, consistent with the true model configuration. Nonetheless, it is important to be mindful of the fact that, in many experiments, the resulting Pareto shape parameter $\hat{k}$ appeared above the threshold of 0.7 for some samples (above which PSIS has a significant increase in complexity [26]). This typically happens to those with fitted latent factors distant from the origin. Figure 4.4 shows that this happens in this dataset too.

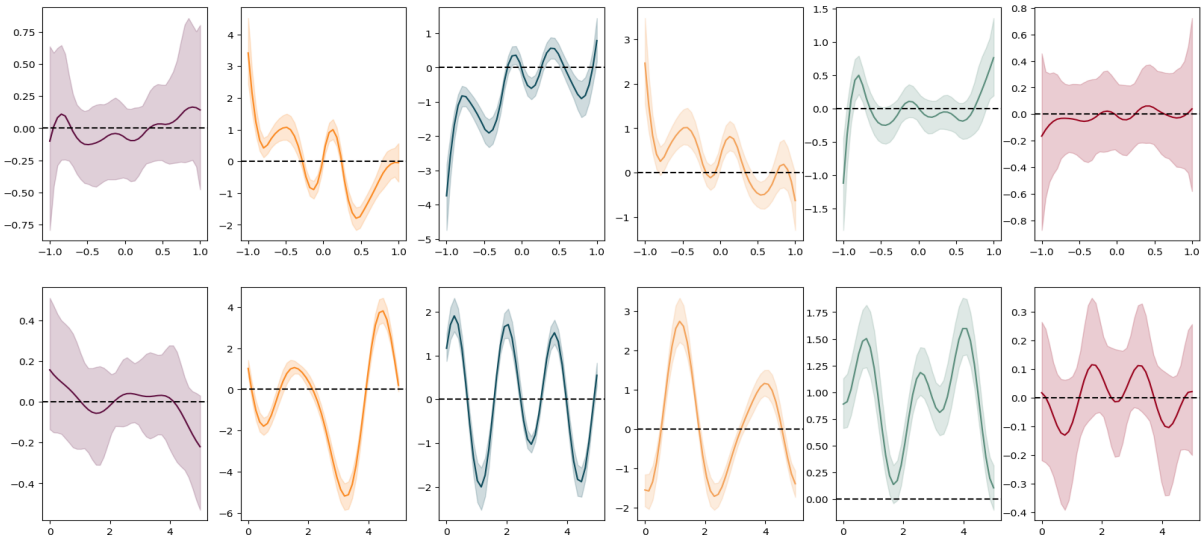


Figure 4.3: The solid lines and the shaded areas represent the mean and the 90% percentile intervals of the latent components (in increasing order, from left to right) sampled from the posterior distribution, after the application of the Varimax-RSP algorithm, when the number of Latent Factors has been fixed to $k = 6$. Two latent factors have been given credible intervals containing the 0 line. The top row refers to the latent components associated with the first target $\boldsymbol{Y}^{(1)}$, and the bottom row to the second $\boldsymbol{Y}^{(2)}$.

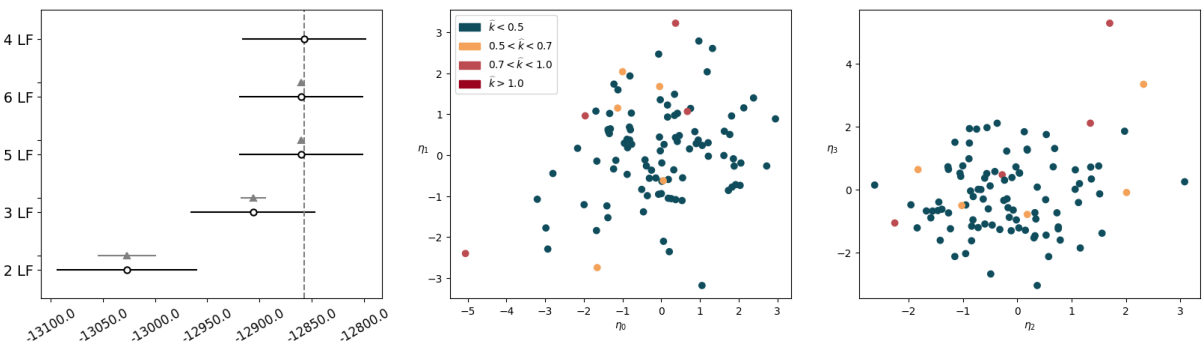


Figure 4.4: ELPD and its uncertainty for varying values of k (displayed in black) on the left; the difference in ELPD relative to the best-fitting model and its associated uncertainty are shown in gray. The model achieving the highest ELPD (and thus considered the best) corresponds to $k = 4$. Models with a lower number of latent factors incur greater penalties compared to those with a higher number of latent factors. The center figure illustrates the expectation of the posterior latent factors of the training set samples (first and second), while the figure on the right refers to the third and fourth latent factors. The points are color-coded according to the value of the Pareto shape parameter.

Validation of the Coefficient of Determination. We sampled 100 additional samples from the same prior and computed the posterior distribution for each using HMC. However, we used a test set of size $N_T = 2000$, while keeping the training set unchanged at a size of $N = 100$.

For each sample w from the posterior distribution, we applied the formula to compute $R^2_{1\to 2}|\boldsymbol{\Theta}_w$. Then, we calculated the mean and the 90% Highest Density Interval of $R^2_{1\to 2}|\boldsymbol{Y}$. On the other hand, computing the predictions, both unconditional and conditional of $\widetilde{\boldsymbol{Y}}_s^{(2)}$ for each s in the test set, one can estimate the value of $R^2_{1\to 2}$, as follows:

$$R^{2\star}_{1\to 2} \simeq 1 - \frac{\sum_{s=1}^{N_T}\left(\widetilde{\boldsymbol{Y}}_s^{(2)} - \mathbb{E}\left[\widetilde{\boldsymbol{Y}}_s^{(2)}\middle|\,\widetilde{\boldsymbol{Y}}_s^{(1)}, \boldsymbol{Y}\right]\right)^2}{\sum_{s=1}^{N_T}\left(\widetilde{\boldsymbol{Y}}_s^{(2)} - \mathbb{E}\left[\widetilde{\boldsymbol{Y}}_s^{(2)}\middle|\,\boldsymbol{Y}\right]\right)^2}. \tag{4.1}$$

Out of 100 samples, 91 of the computed $R^{2\star}_{1\to 2}$ resulted within the 90% Highest Density Interval of $R^2_{1\to 2}|\boldsymbol{Y}$ (binomial test p-value $\simeq 0.87$), proving good calibration of the sampling algorithm. Figure 4.5 illustrates that the estimated density is centered in zero, proving that $R^2_{1\to 2}|\boldsymbol{Y}$ presents an unbiased estimate of $R^{2\star}_{1\to 2}$.

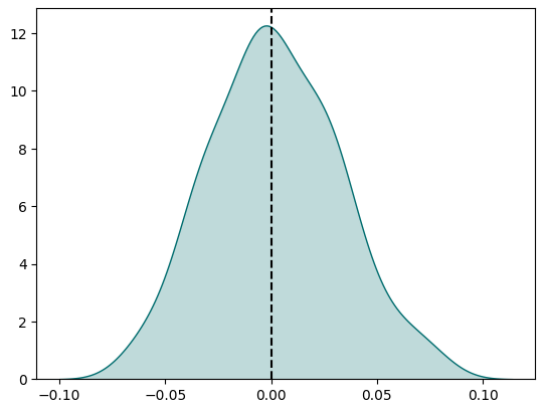


Figure 4.5: Kernel Density Estimation of the difference between $\mathbb{E}\left[R^2_{1\to 2}|\,\boldsymbol{Y}\right]$ and $R^{2\star}_{1\to 2}$ computed on the Test Set, related to the example in Section 4.1.

4.2. Dataset Generated from Prior Distribution: The 2D case

We simulated a dataset by sampling from the prior distribution of the BBLF model. For $\boldsymbol{Y}^{(1)}$ we employed 20 fourth order b-splines on the interval $(-1, 1)$ measured on $L_1 = 50$ evenly spaced points. $\boldsymbol{Y}^{(2)}$ is, instead, taken bidimensional, with basis matrix generated

from the exponential quadratic covariance kernel on the $(0,1)^2$ domain, with lengthscale $\rho = 0.3$, truncating the singular values decomposition to $p_2 = 30$ columns.

The number of latent factors was set to $k = 4$. We sampled 150 points, of which 100 formed the training set, and the remaining the test set. 5 covariates per data point were drawn independently from a normal distribution with a standard deviation fixed to 0.1. We set $\alpha_1 = 2.1$, $\alpha_2 = 2.5$, $\nu = \chi = 4$.

The posterior Monte Carlo Markov Chain samples are taken using Hamiltonian Monte Carlo No U-turn sampler [24], with 4 chains, a burn-in of 500 samples, maximum tree depth of 12. 1000 samples were drawn per chain.

The results obtained for this second example are depicted in Figures 4.6, 4.7 and 4.8.

4.3. Application to Function-to-Vector Regression: The Biscuit Dataset

The Biscuit dataset from [23] contains the result of two experiments, conducted on separate occasions, in which biscuit dough with varying composition was subject to measurements of near-infrared reflectance spectra (NIR) in the spectral range $1100 - 2498$nm, with steps of 2nm. For each dough, the NIR spectrum and the composition of the dough in terms of fat, sucrose, flour, and water, all expressed in percentage, were recorded. The percentage almost adds up to 100, not precisely because some other ingredients are in small quantities.

The interest in this dataset comes partly from its natural split in two sets, allowing one for training and the other for testing. Additionally, its measurements bear similarity to the motivating example of this thesis work. The training set contains $N = 40$ samples, while the testing set contains 32 samples. Both the sets are illustrated in Figure 4.9.

Prior Elicitation. The most crucial choices refer to the selection of basis function. For the vector target, we applied PCA and find that, as the sum of all features nearly totals 100, the last principal component explains an extremely low portion of variance. Therefore, we selected only the first three principal components. For the spectral target, we opted for a covariance kernel-induced basis matrix. After directly inspecting the shape of the functions, a Matérn 5/2 kernel seemed reasonable. Upon inspecting the empirical variogram in Figure 4.10, we set the length scale parameter to $\rho = 200.0$. In contrast, the variogram suggested a variance of 0.05 for σ_d, a quantity adjusted by the model. Thus, in the prior selection, we chose to represent the order of magnitude as $\sigma_d = 0.1$. The

latent components have a student-t prior. Choosing these degrees of freedom sufficiently small allows the model to capture more variability than what would be captured by a Gaussian Process, thus eliminating the need to be precise with the length scale parameter. Moreover, it is generally better to be conservative and allow more variability by choosing a smaller ρ. We approximated the covariance matrix with $p_1 = 25$ basis functions. We chose degrees of freedom $\nu = 4$ for the latent components, while $\alpha_\sigma = 5$. With $\alpha_1 = 2.1$ and $\alpha_2 = 2.5$, and $k = 4$ latent factors, the prior predictive portion of variance explained by the latent factors is around 80% of the total variance explainable by the latent factors. β_σ strikes the prior predictive balance between the variance $\rho_\Lambda^{(k)}\sigma_\Lambda^2 \simeq 4.4$ explained by the latent factors and that explained by the noise on the basis coefficients: $\beta_\sigma/(\alpha_\sigma - 1)$. We chose $\beta_\sigma = 4$.

In this particular example, there are no covariates. However, we can still apply the BBLF model to this scenario by treating it as a special case. This involves passing a single vector of covariates constantly equal to 0 and assigning the student-t prior to the regression coefficients with $\chi = 100$ degrees of freedom. While this parameter does not affect the rest of the model significantly, it may slightly slow down the sampling algorithm. We set $\alpha_\psi = 5$ and $\beta_\psi^{(2)} = 4 \times 10^{-6}$ to reflect our hypothesis that all variability is already captured by the principal components. Furthermore, we choose $\beta_\psi^{(1)} = 4 \times 0.05$ as the prior predictive white noise variance, ensuring it is of the same order of magnitude as the target.

Results. Figure 4.11 shows that all four latent components are significant, but only two affect the vector target too. The choice of $k = 4$ is validated by refitting the model with $k = 6$, which shows that the additional components are given extremely low statistical significance, as shown in Figure 4.12. Finally, thanks to the partition into training and test set, we also check the R^2 estimate in Figure 4.13, which shows the empirical estimate falls within the body of the posterior distribution of $R^2_{1\to 2}$. As no covariates are present, this amounts to the true portion of the explained variance.

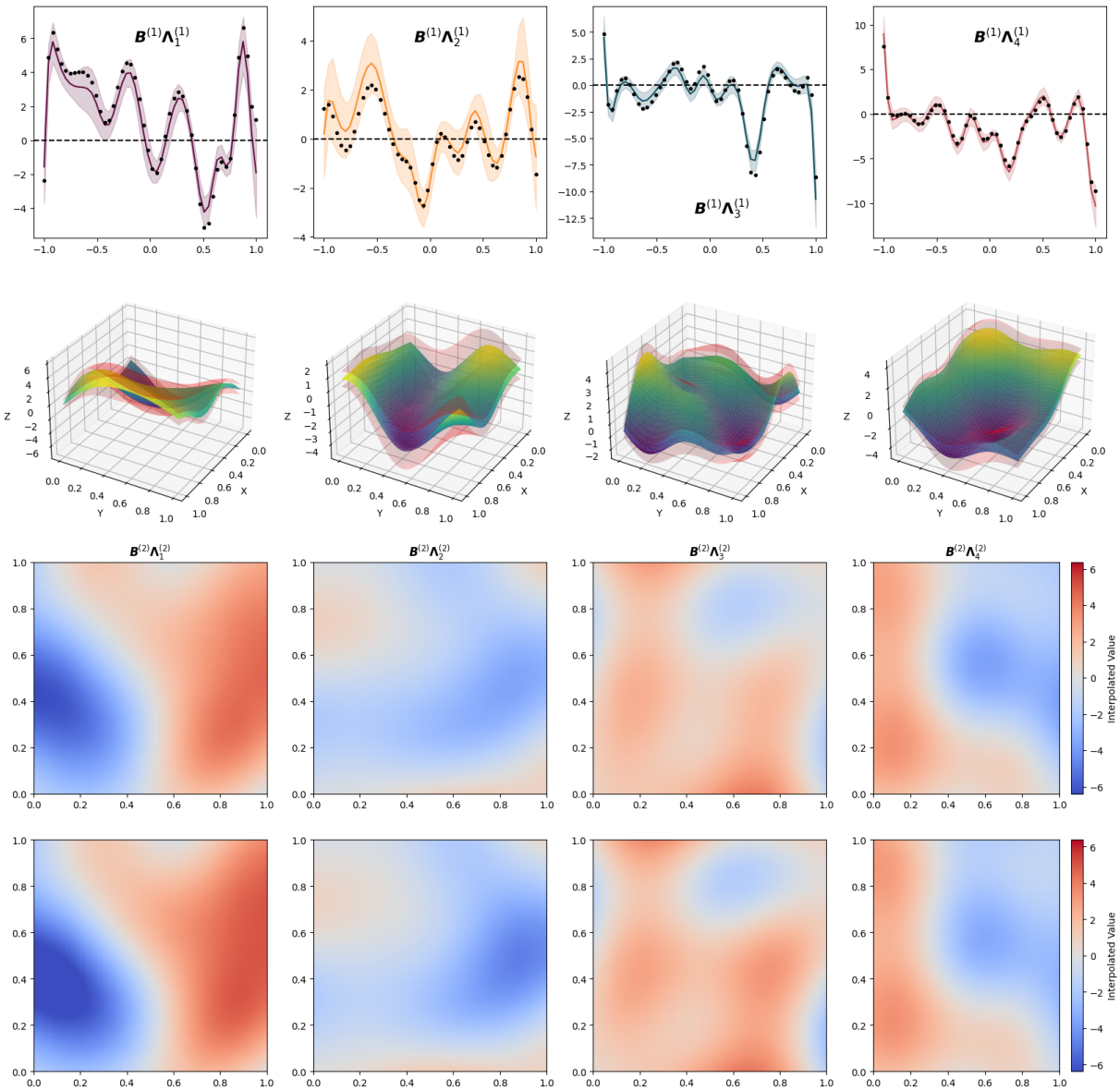


Figure 4.6: Top row: the solid lines and the shaded areas represent the mean and the 90% percentile intervals of the latent components sampled from the posterior distribution, after the application of the Varimax-RSP algorithm. The dots are the values of the *true* latent components, once rotated and permuted to align to the Monte Carlo samples. To simplify the interpretation, the components have been multiplied by the respective basis matrices. Second and third row show $\boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}$ estimated with the Varimax-RSP algorithm. The second row shows the 90% percentile intervals, too, while in the third row, only the means are shown. The fourth row shows the *true* latent components.

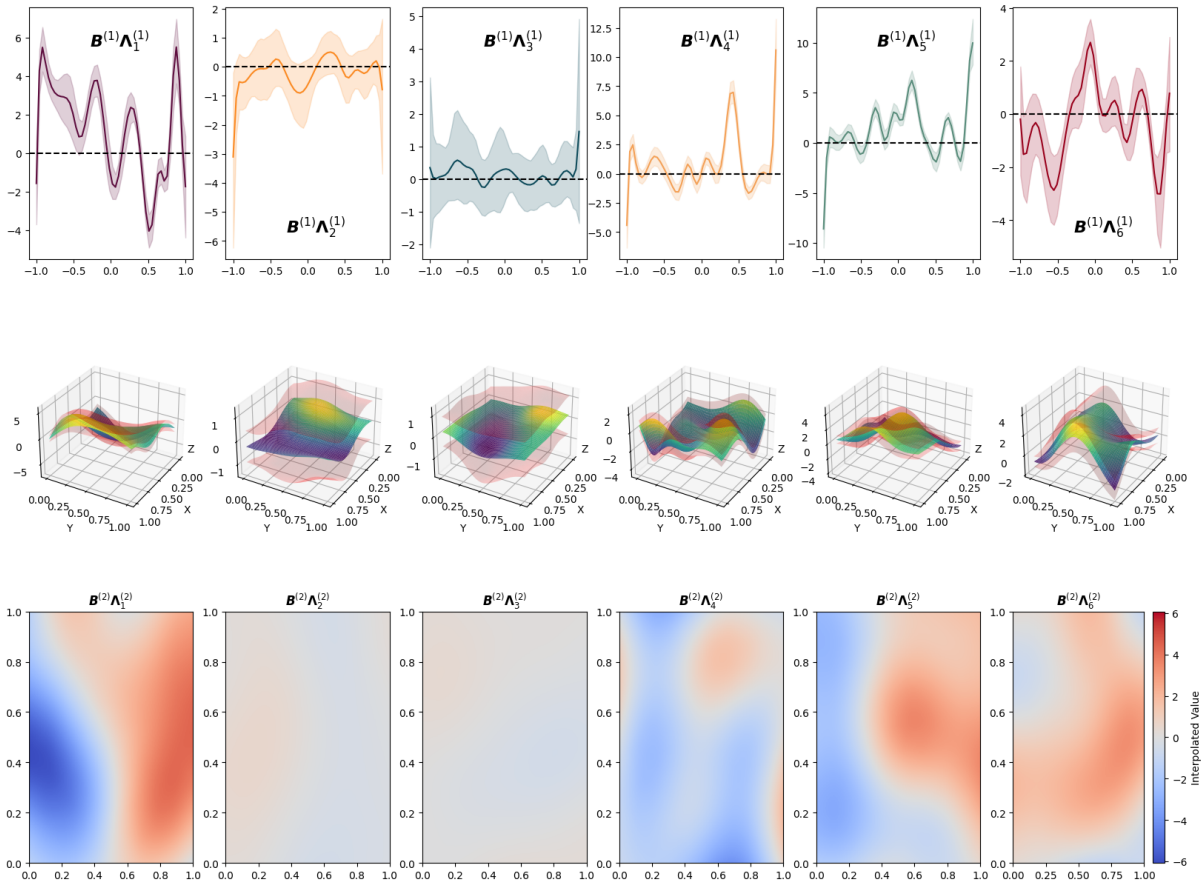


Figure 4.7: Mean and the 90% percentile intervals of the latent components sampled from the posterior distribution, after the application of the Varimax-RSP algorithm, when the number of Latent Factors has been fixed to $k = 6$. Two latent factors have been correctly given credible intervals containing the 0 line/plane.

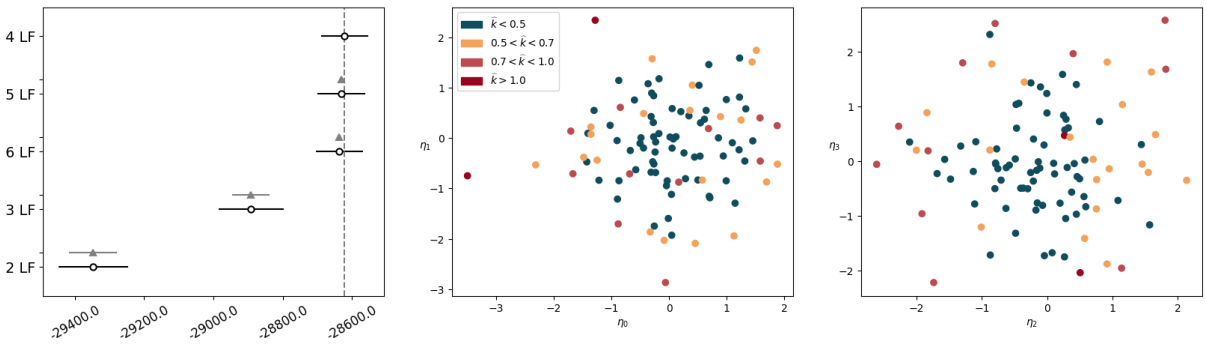


Figure 4.8: ELPD (left) and its uncertainty, for varying k (in black); the ELPD difference with respect to the best fitting model, and its uncertainty, in gray. The model with $k = 4$ is the model with the highest ELPD (best). Estimated latent factors for the samples in the training set (center and tight), color-coded by the value of the Pareto shape parameter.

Figure 4.9: Training set (left), and test set (right). The top row refers to the original data, and the bottom row is obtained by removing the mean of the training set to both datasets and downsampling a factor of 3.

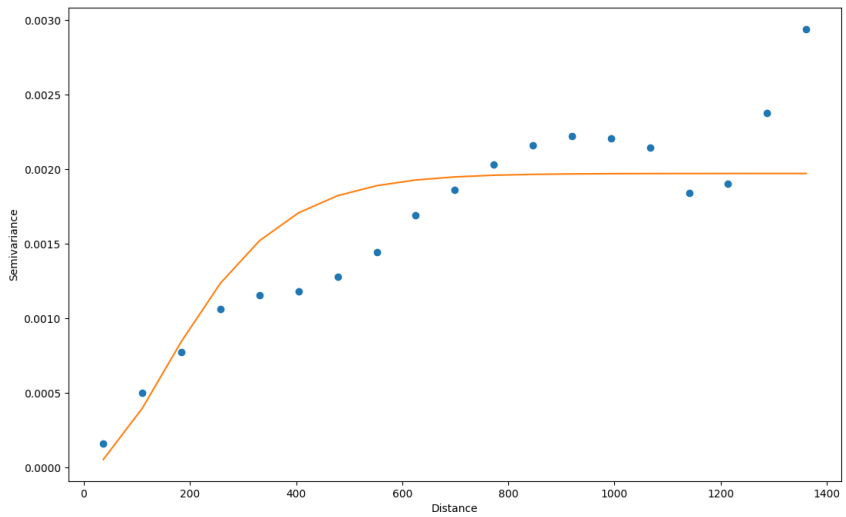


Figure 4.10: Blue dots represent the Empirical semivariogram for the training Set of the biscuit dataset. The solid line shows the semivariogram of the Matérn 5/2 kernel with $\rho = 200.0$, $\sigma_d = 0.045$.

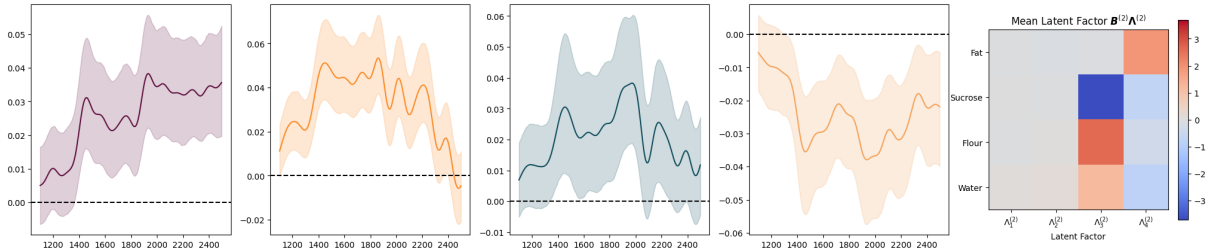


Figure 4.11: Solid and shaded lines depict $\boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}$ per each latent factor, the heatmap on the right, depicts $\mathbb{E}\left[\boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\middle|\boldsymbol{Y}\right]$ the posterior expectation of the regression from the latent factors to the vector target. Each column represents a different latent factor, each row a different component of the vector target; only the last two latent factors are effectively used for the regression.

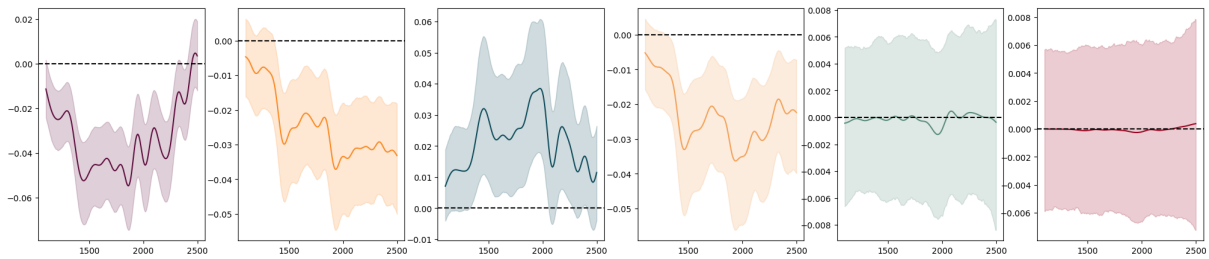


Figure 4.12: Solid and shaded lines depict $\boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}$ per each latent factor. The last two latent components have credibility regions encompassing the 0 line.

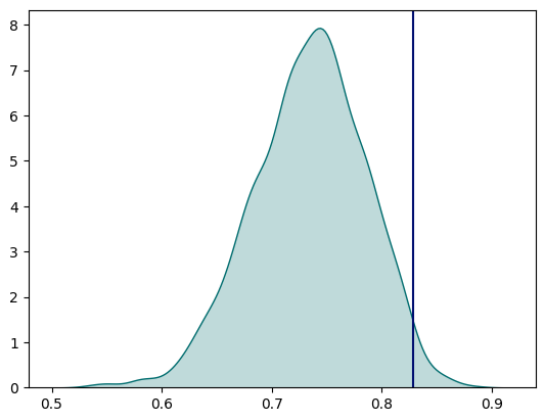


Figure 4.13: The shaded region illustrates the posterior density of the coefficient of determination $R^2_{1\to 2}|\boldsymbol{Y}$, as determined through Kernel Density Estimation. The vertical line marks the empirical $R^{2\star}_{1\to 2}$ as computed on the test set, in accordance with Equation (4.1).

5 | Posterior Inference on NOM Characterization

This chapter applies the BBLF model to real-life absorbance and fluorescence spectra.

The data presented originate from the water treatment facility located in Northern Italy, specifically along the treatment pathway of a river. The focus is on examining the dataset during the ozonation phase, a critical step downstream of a comprehensive treatment process that includes lagooning, clariflocculation, and sand filtration [3].

The dataset consists of $N = 113$ samples. Before analysis, the absorbance and fluorescence spectra data are centered and standardized. Each sample is subject to ozonation and activated carbon to various degrees, with this information incorporated in the form of covariates. A few samples have undergone different treatment processes, and if datum for either the activated carbon or ozonation is missing, this is written in a separate boolean covariate after imputing the value of 0 to the original covariate.

We will consider two applications, first a function-on-vector approach, where target $Y^{(1)}$ consists of the absorbance spectra and $Y^{(2)}$ the fluorescence spectra elaborated through the peak-picking technique. The second is a complete bivariate functional problem: target $Y^{(1)}$ is the absorbance, and $Y^{(2)}$ consists of the Excitation-Emission matrices of the fluorescence spectra.

5.1. Prediction of the Peak Picking

The hyperparameters of the prior are chosen as follows: $k = 6$, $\nu = 4$, $\chi = 5$, $\alpha_\psi = 5$, $\beta_\psi^{(1)} = 4$, $\beta_\psi^{(2)} = 0$, $\alpha_\sigma = 5$, $\beta_\sigma = 4$, $\alpha_1 = 2.1$, and $\alpha_2 = 2.5$. The absorbance spectra after centering and normalization are shown in Figure 5.1. The basis matrix is determined as a covariance-kernel-induced basis matrix. In particular, the Matérn 5/2 covariance kernel was chosen. The choice of the parameters is informed by the variogram shown in Figure 5.2, and fixed to $\rho = 25.0$, and $\sigma_d = 1.0$. The number of singular values considered is $p_1 = 20$.

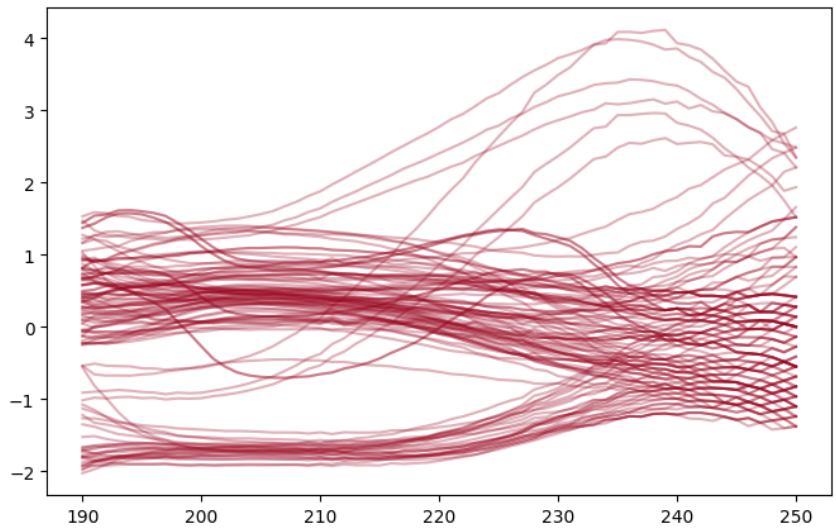


Figure 5.1: Centered and standardized absorbance spectra.

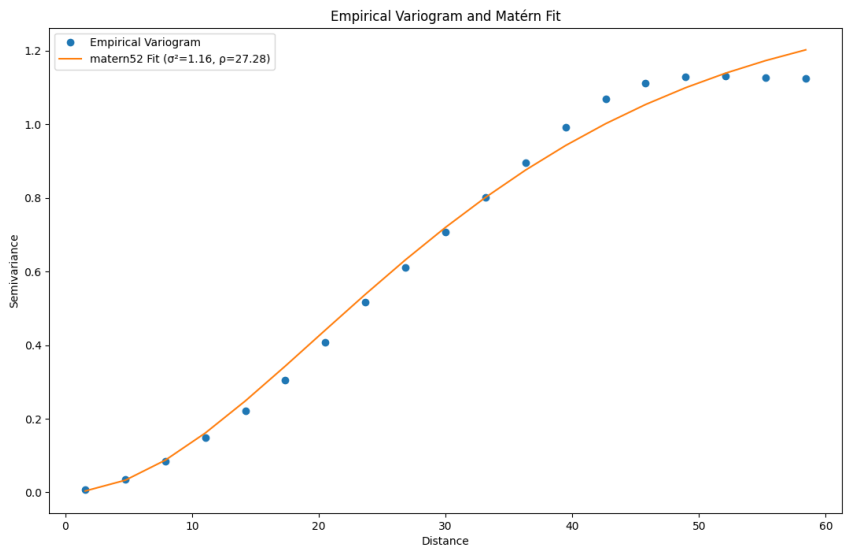


Figure 5.2: The blue dots represent the empirical semivariogram of the standardized absorbance spectra, while the orange line represents the best fit Matérn 5/2 model.

The basis matrix for the vector target consisting of the results of the peak-picking technique, encompassing the peaks known in the literature as A, M, C, D,and N [31], is chosen using PCA and picking all five components, depicted in Figure 5.3. Figures 5.4 and 5.5 show the latent components that BBLT model fit for this dataset. The components relating to the third and sixth factors have low statistical significance for both targets, indicating that $k = 4$ was sufficient. Latent component 4 is responsible for the effect of the first principal component, explaining almost all variability and describing the overall value of the fluorescence spectrum. Latent components 1 and 5 have similar interpreta-

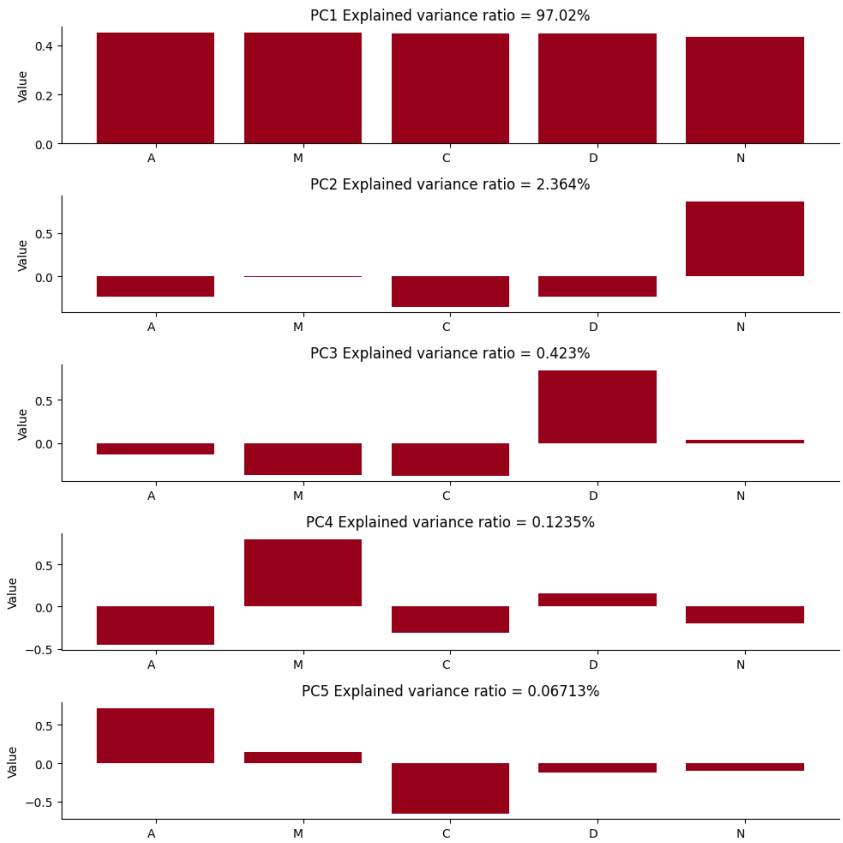


Figure 5.3: Result of PCA applied to the output of the Peak Picking technique.

tions, in that they are responsible for estimating the difference of peak N with respect to the others.

5.2. Prediction of Excitation-Emission matrix

The choices for the hyperparameters of the prior are $k = 8$, $\nu = 4$, $\chi = 5$, $\alpha_\psi = 5$, $\beta_\psi^{(1)} = 4$, $\beta_\psi^{(2)} = 0$, $\alpha_\sigma = 5$, $\beta_\sigma = 4$, $\alpha_1 = 2.1$, $\alpha_2 = 2.5$. The absorbance is treated like the previous example, except $p_1 = 10$.

The choice for the basis matrix for the fluorescence spectra is again a Matérn kernel 5/2, with parameters informed by the variogram in Figure 5.6 fixed to $\rho = 130.0$, $\sigma_d = 1.0$, taking $p_2 = 50$ singular values.

Figure 5.7 and 5.8 show the latent components. With respect to the peak-picking technique, it is possible to recognise typical peaks [4]. Latent components 5 and 6 exhibit part of the A peak, representative of analytes with humic-like structure with two aromatic rings, and C, with two or more aromatic rings. Latent components 1, 2, and 4 show part

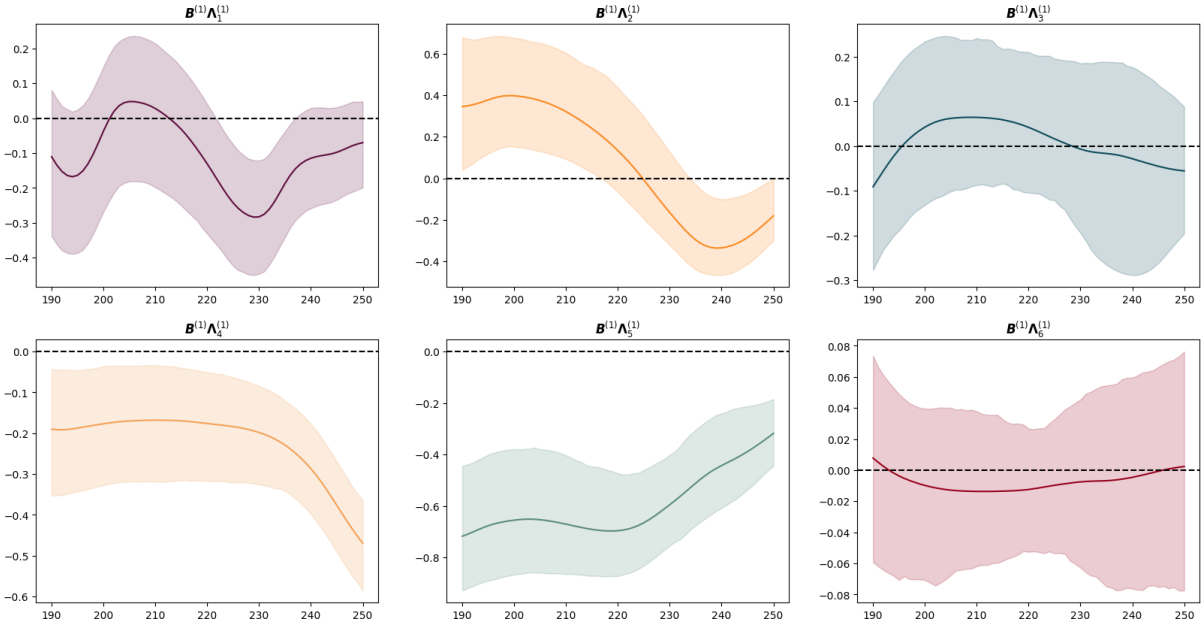


Figure 5.4: Latent components relating to the absorbance spectra.

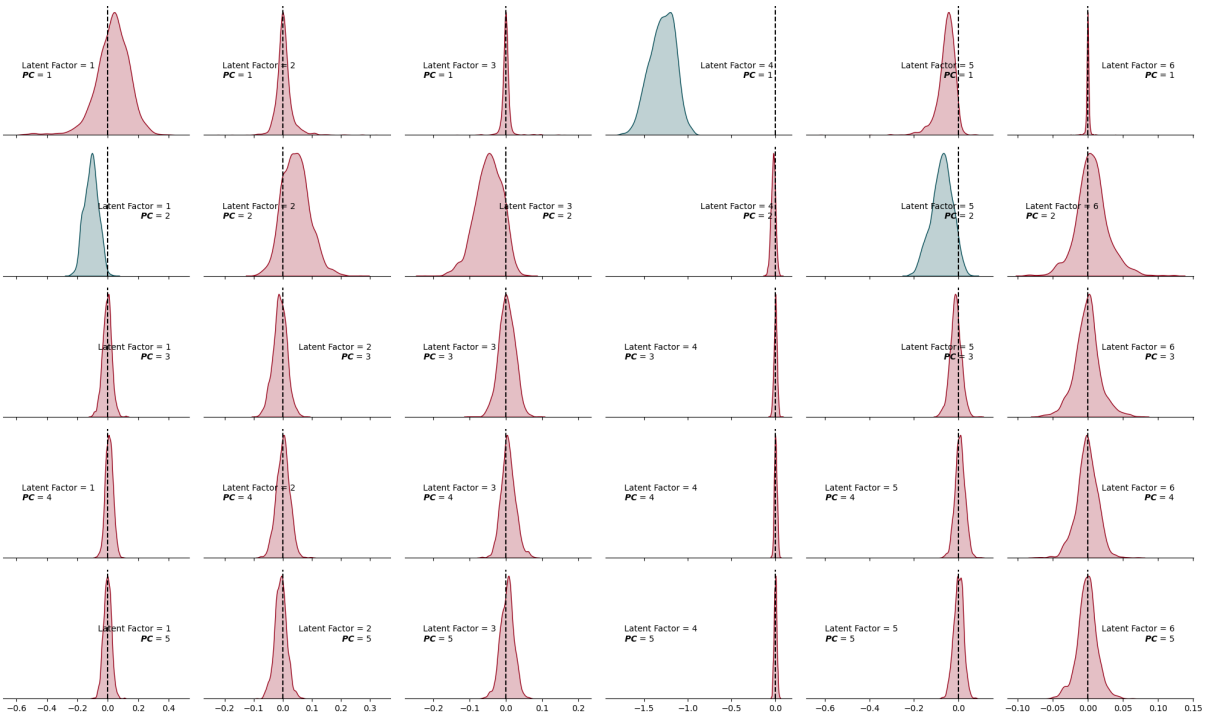


Figure 5.5: Kernel Density Estimation of the latent components related to the peak-picking target. Red if the 90% Highest Density Interval contains the 0, green otherwise.

of the T peak, indicative of analytes characterised by a tryptophan-like structure with at least one aromatic ring. Latent component 7 showing peak T and A. Latent component 3 has low statistical significance for both targets. Finally, latent component 8 recognizes

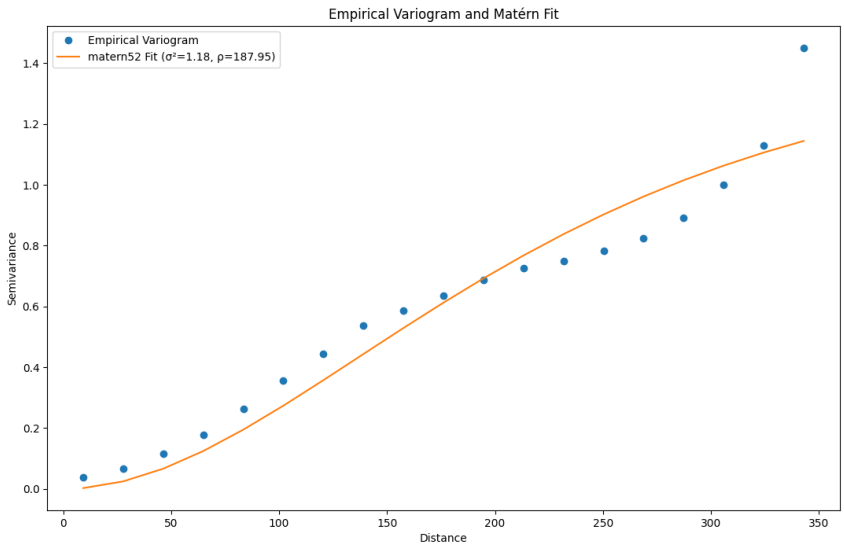


Figure 5.6: The blue dots represent the empirical semivariogram of the standardized fluorescence spectra, while the orange line represents the best fit Matérn 5/2 model.

a peak in the location between Rayleigh and Raman scattering, which, by inspection of the projection on the latent components in Figure 5.9, seems to happen for only $3-4$ samples in the dataset, describing what probably is an outlier.

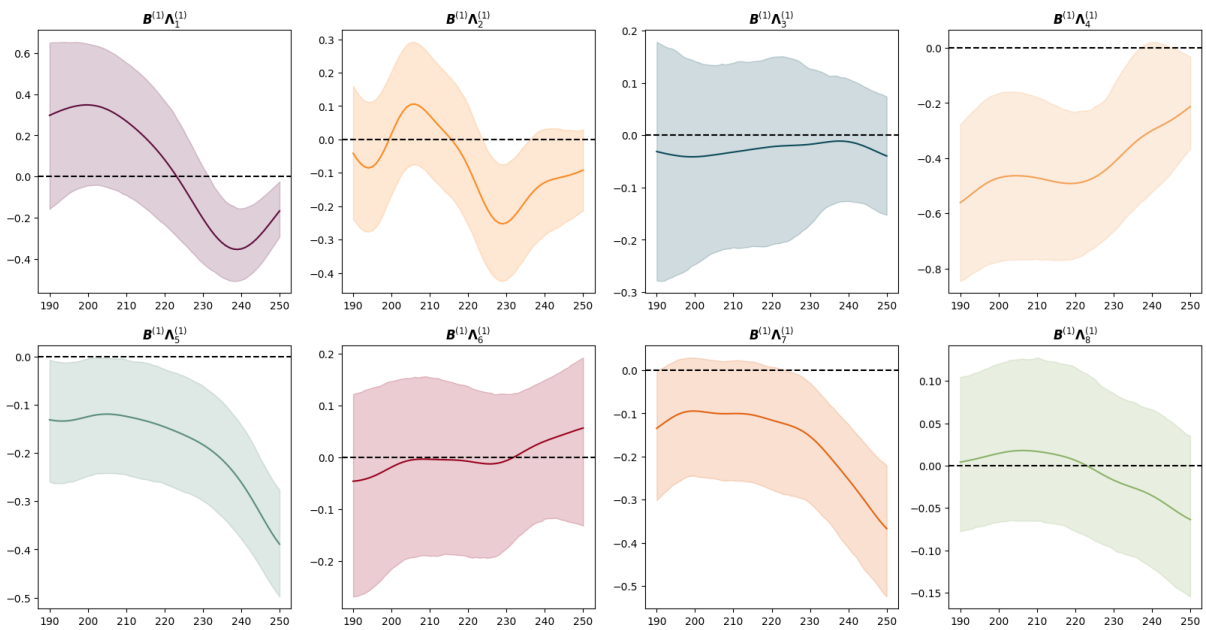


Figure 5.7: Latent components relating to the absorbance spectra.

Finally, we report in Figure 5.10 the posterior distribution of the coefficient of determination $R^2_{1\to 2}|\boldsymbol{Y}$ of the fluorescence spectra with respect to the absorbance spectra,

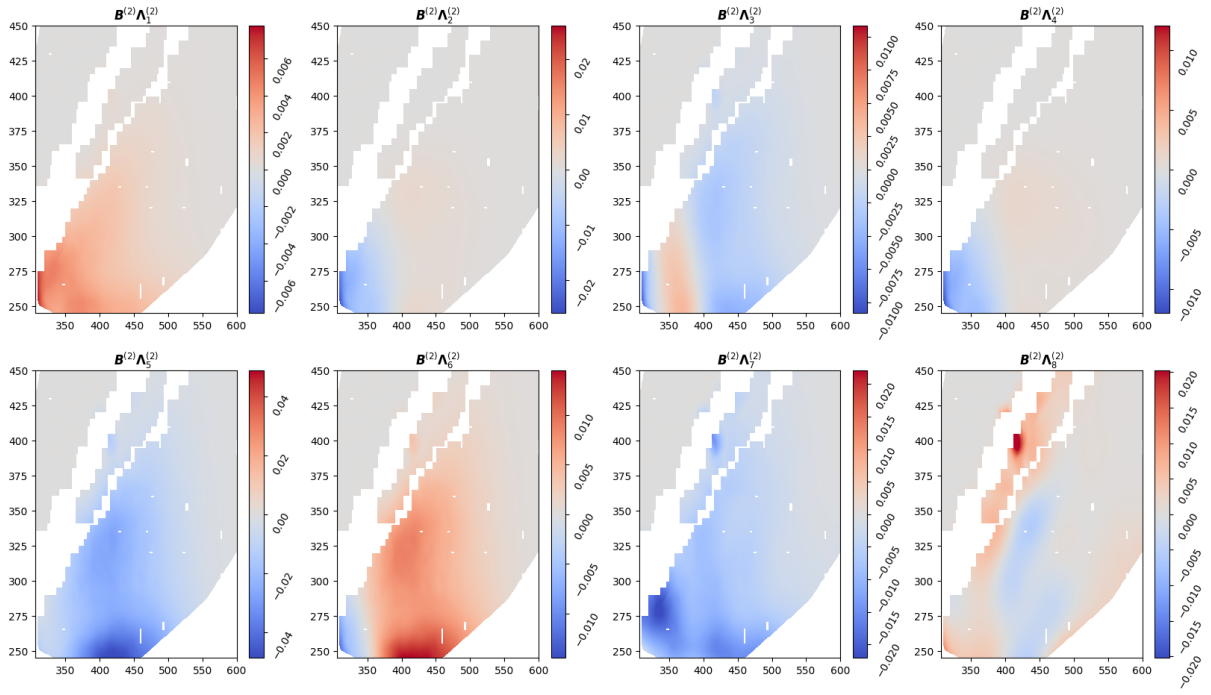


Figure 5.8: Latent components relating to the fluorescence spectra, the normalization was removed to simplify interpretation with respect to classical techniques for fluorescence spectra interpretation.

recalling that it refers to the portion of the variance of the fluorescence, explained by the absorbance, over that not explained by the covariates, showing a weak ability of the absorbance in predicting the fluorescence spectra.

Remark. It should be noted that these are preliminary results. Future investigations will determine a standardized approach to applying this toolset to the problem of NOM characterization, fine-tuning the technique to meet the expert's needs. For example, while the prior considered in this chapter are weakly informative, a data-driven approach to constructing the basis matrix could be explored.

Secondly, the centering and standardization of absorbance and fluorescence data considered here constitute a form of empirical Bayes. Alternatively, a different approach could involve estimating functions for the mean and standard deviation from a richer dataset instead of relying solely on the current dataset for estimation.

Furthermore, as of the time of writing, not all algorithms in GitHub have been updated to the version based on projection onto the vector space of basis coefficients. Therefore, due to memory constraints, it was necessary to downsample the fluorescence spectra.

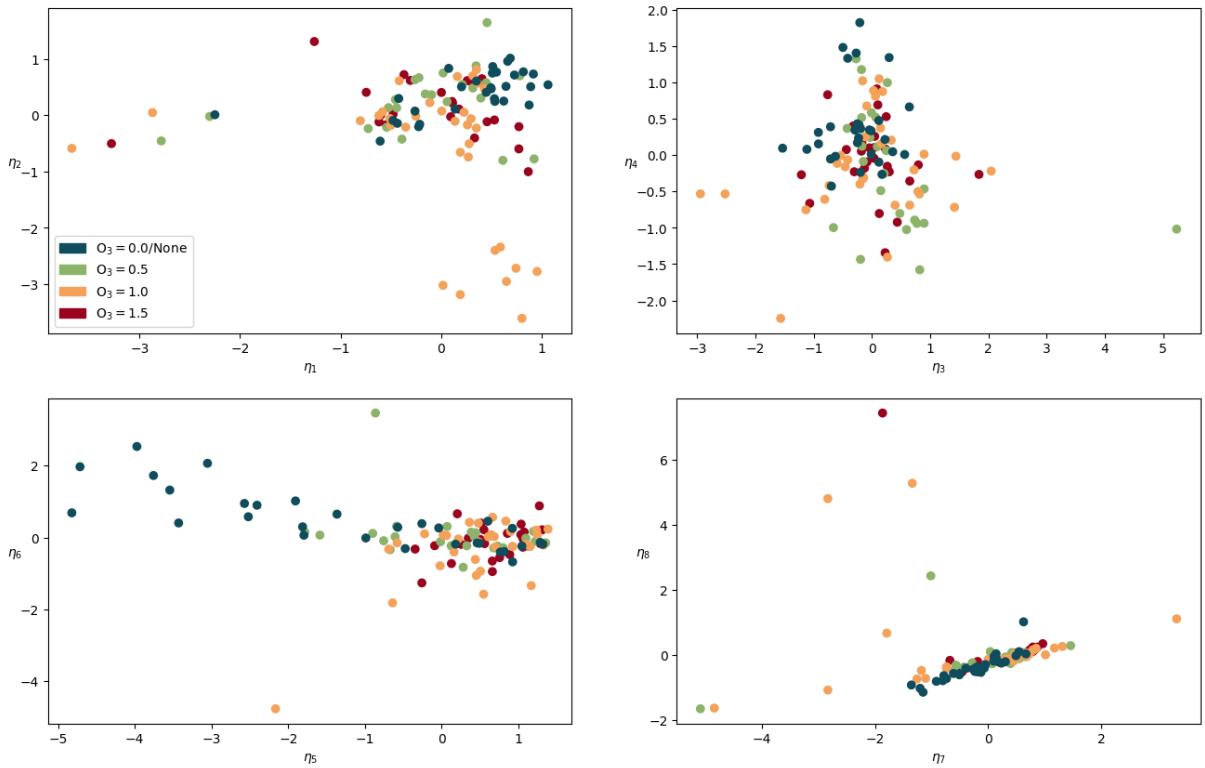


Figure 5.9: Mean of the projection on the Latent Factor space for the elements of the dataset, color-coded by level of ozonation. The 5th Latent Factor appears closely related to the level of ozonation.

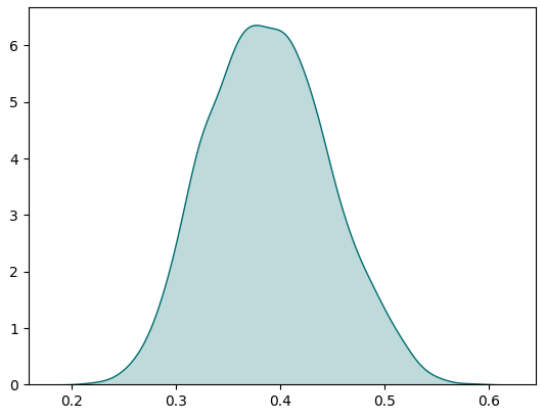


Figure 5.10: Posterior density of the coefficient of determination $R^2_{1\to 2}|\boldsymbol{Y}$, as described in Equation (3.5).

6 | Conclusions and future developments

This dissertation has introduced and rigorously analyzed the Bivariate Bayesian Latent Factor (BBLF) model, showcasing its theoretical robustness and practical utility in the domain of multi-target inference from functional data. Through detailed mathematical formulation, comprehensive model specification, and robust validation against both simulated and empirical datasets, this work firmly establishes the BBLF model as an effective framework for elucidating latent structures within bivariate functional observations. Integrating the Varimax-RSP algorithm has been pivotal, significantly enriching the interpretability of these latent structures.

6.1. Conclusions

The contributions of this research are multifaceted. Primarily, it has authenticated the BBLF model's proficiency in accurately capturing latent factors within bivariate functional data, providing a novel perspective for analyzing complex environmental phenomena, especially in NOM characterization through spectroscopic techniques. Moreover, the model's application across various datasets highlights its versatility and potential to drive sophisticated analytical tools within and beyond environmental science domains.

This thesis's theoretical developments and practical applications significantly advance the field of Bayesian latent factor modeling, with a particular emphasis on functional data analysis. The creation of a Python package and the extensive exploration of the model's posterior and posterior predictive distributions underscore the BBLF model's suitability and promise for broad adoption in applied functional data analysis settings.

6.2. Future Developments

Building on the groundwork established by this research, several avenues for future exploration and enhancement are identified:

1. **Excitation-Emission Matrix Standardization:** Developing a standardized protocol for the model's application in absorbance and fluorescence spectroscopy will facilitate its integration by domain experts, aligning it with the analytical processes of existing statistical tools but with reduced complexities in result validation.

2. **Bayesian PARAFAC:** Employing similar mathematical methodologies as those used in this thesis, a model that emulates the bilinear latent component structure assumed in PARAFAC can be formulated. A comparative analysis between this model and the BBLF could serve to validate the bilinear assumption's efficacy.

3. **Clusterization in the Latent Factors space:** In Appendix D, we found that in many NOM Characterization situations, samples often form natural groups or clusters. This suggests that we could improve our model by using a Gaussian Mixture model for the latent factors, in place of the current normal distribution. It also opens up the possibility of grouping the samples based on the fitted latent factors, which could help us develop better tools for analyzing the data.

4. **Mathematical Investigation into Basis Coefficients Space Projection:** The simplification achieved in Section 3.1.2, predicated on the homoschedastic white noise assumption for measurement noise modeling, prompts further inquiry. Specifically, the potential for analogous simplifications under more complex covariance structures merits detailed exploration.

In summary, the BBLF model represents a significant stride in the statistical analysis of bivariate functional data. The paths outlined for future research promise to refine and expand upon the model's capabilities and explore new horizons in Bayesian latent factor modeling and functional data analysis at large.

Bibliography

[1] P. Baldi. *Stochastic Calculus: An Introduction Through Theory and Exercises*. Springer, 01 2017. ISBN 978-3-319-62225-5. doi: 10.1007/978-3-319-62226-2.

[2] B. Cantoni, L. Riva, C. Romani, I. Epifani, F. Trovò, C. Punta, and M. Antonelli. Safecrew: a holistic project to reduce nom and dbps in drinking water and minimize the human health risk under climate change scenarios. In *Gordon Research Conference*, 2023.

[3] B. Cantoni, J. Ianes, B. Bertolo, S. Ziccardi, F. Maffini, and M. Antonelli. Adsorption on activated carbon combined with ozonation for the removal of contaminants of emerging concern in drinking water. *Journal of Environmental Management*, 350:119537, 2024. ISSN 0301-4797. doi: https://doi.org/10.1016/j.jenvman.2023.119537. URL `https://www.sciencedirect.com/science/article/pii/S0301479723023253`.

[4] W. Chen and H.-Q. Yu. Advances in the characterization and monitoring of natural organic matter using spectroscopic approaches. *Water Research*, 190:116759, 2021. ISSN 0043-1354. doi: https://doi.org/10.1016/j.watres.2020.116759. URL `https://www.sciencedirect.com/science/article/pii/S0043135420312926`.

[5] I. Delpla, A. V. Jung, E. Baures, M. Clement, and O. Thomas. Impacts of climate change on surface water quality in relation to drinking water production. *Environmental international*, 35(8):1225–1233, Nov 2009. doi: 10.1016/j.envint.2009.07.001. Epub ahead of print.

[6] Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain. fields: Tools for spatial data, 2021. URL `https://github.com/dnychka/fieldsRPackage`. R package version 15.2.

[7] D. Dubber, J. Knappe, and L. W. Gill. Characterisation of organic matter and its transformation processes in on-site wastewater effluent percolating through soil using fluorescence spectroscopic methods and parallel factor analysis (parafac). *Water*, 13

(19), 2021. ISSN 2073-4441. doi: 10.3390/w13192627. URL https://www.mdpi.com/2073-4441/13/19/2627.

[8] B. S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, January 2011. ISBN 9780470749913. doi: 10.1002/9780470977811. First published: 7 January 2011.

[9] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3 edition, 2013. doi: 10.1201/b16018. URL https://doi.org/10.1201/b16018.

[10] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[11] C. C. Jang JH, Manatunga AK and L. Q. A. Bayesian multiple imputation approach to bivariate functional data with missing components. *Statistics in Medicine*, 40 (4-22):4772–4793, 2021. doi: 10.1002/sim.9093.

[12] G. N. Lance and W. T. Williams. Computer Programs for Hierarchical Polythetic Classification ("Similarity Analyses"). *The Computer Journal*, 9(1):60–64, 05 1966. ISSN 0010-4620. doi: 10.1093/comjnl/9.1.60. URL https://doi.org/10.1093/comjnl/9.1.60.

[13] J. A. LEENHEER and J.-P. CROUP. Characterizing aquatic dissolved organic matter. *Environmental science & technology*, 37(1):18A–26A, 2003.

[14] A. Matilainen, E. Gjessing, T. Lahtinen, L. Hed, A. Bhatnagar, and M. Sillanpää. An overview of the methods used in the characterisation of natural organic matter (nom) in relation to drinking water treatment. *Chemosphere*, 83:1431–1442, 2011.

[15] S. Montagna, S. Tokdar, B. Neelon, and D. Dunson. Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68:1064–1073, 12 2012. doi: 10.2307/41806025.

[16] K. R. Murphy, C. A. Stedmon, D. Graeber, and R. Bro. Fluorescence spectroscopy and multi-way techniques. parafac. *Anal. Methods*, 5:6557–6566, 2013. doi: 10.1039/C3AY41160E.

[17] K. R. Murphy, C. A. Stedmon, D. Graeber, and R. Bro. Fluorescence spectroscopy and multi-way techniques. parafac. *Anal. Methods*, 5:6557–6566, 2013. doi: 10.1039/C3AY41160E. URL http://dx.doi.org/10.1039/C3AY41160E.

[18] P. Papastamoulis and I. Ntzoufras. On the identifiability of bayesian factor analytic models. *Statistics and Computing*, 32(2):23, 2022.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[20] C. Ramos-Carreño, J. L. Torrecilla, M. Carbajo-Berrocal, P. Marcos, and A. Suárez. scikit-fda: A python package for functional data analysis, 2023.

[21] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005. ISBN 9780262256834. doi: 10.7551/mitpress/3206.001.0001. URL `https://doi.org/10.7551/mitpress/3206.001.0001`.

[22] S. D. Richardson, M. J. Plewa, E. D. Wagner, R. Schoeny, and D. M. DeMarini. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutation Research/Reviews in Mutation Research*, 636(1):178–242, 2007. ISSN 1383-5742. doi: https://doi.org/10.1016/j.mrrev.2007.09.001. URL `https://www.sciencedirect.com/science/article/pii/S138357420700035X`. The Sources and Potential Hazards of Mutagens in Complex Environmental Matrices - Part II.

[23] H. L. Shang and R. J. Hyndman. fds: Functional data sets, 2018. URL `https://cran.r-project.org/package=fds`. R package version 1.8.

[24] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*, 2024. URL `https://mc-stan.org`. 2.34.

[25] B. Ursino, I. Epifani, and F. Trovò. bayesian-multitarget-latent-factors, 2024. URL `https://github.com/Addicted2BayesianEpistemology/bayesian_multitarget_latent_factors`. GitHub Repository.

[26] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2015.

[27] A. Vehtari, A. Gelman, and J. Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432, 2017. doi: 10.1007/s11222-016-9696-4. URL `https://doi.org/10.1007/s11222-016-9696-4`.

[28] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold,

R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17: 261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[29] Y. Wang. *Smoothing splines: Methods and applications*. 06 2011.

[30] World Health Organization (WHO) TEAM. Guidelines for Drinking-Water Quality, 4th Edition, Incorporating the 1st Addendum, Year of publication, e.g., 2022. URL `https://www.who.int/publications/i/item/9789241549950`. Radiation and health (RAD), Water, Sanitation, Hygiene and Health (WSH) Overview.

[31] U. J. Wünsch, R. Bro, C. A. Stedmon, P. Wenig, and K. R. Murphy. Emerging patterns in the global distribution of dissolved organic matter fluorescence. *Anal. Methods*, 11:888–893, 2019. doi: 10.1039/C8AY02422G. URL `http://dx.doi.org/10.1039/C8AY02422G`.

A | Prior Statistical Moments of the Latent Factors

Proposition A.1 (Expected Loading Factors).

$$\mathbb{E}\left[\Lambda_{j\,\ell}^{(\iota)}\right] = 0 \qquad \forall j, \forall \ell, \forall \iota$$

Proof.

$$\mathbb{E}\left[\Lambda_{j\,\ell}^{(\iota)}\right] = \mathbb{E}\left[\mathbb{E}\left[\Lambda_{j\,\ell}^{(\iota)} \middle| \tau_{\lambda\,\ell}^{(\iota)}\right]\right] = \mathbb{E}\left[0\right] = 0$$

□

Proposition A.2 (Expected Reciprocal of the Precision $\tau_{\lambda}^{(\iota)}$ of the Loading Factors). *Let $\alpha_1 > 1$ and $\alpha_2 > 1$, then:*

$$\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

Proof.

$$\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \mathbb{E}\left[\prod_{\gamma=1}^{\ell}\left(\delta_{\gamma}^{(\iota)}\right)^{-1}\right] = \prod_{\gamma=1}^{\ell}\mathbb{E}\left[\left(\delta_{\gamma}^{(\iota)}\right)^{-1}\right]$$

If $\ell = 1$,

$$\mathbb{E}\left[\left(\tau_{\lambda\,1}^{(\iota)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{0} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{1-1}$$

This is derived using the well-known moments of the Inverse-Gamma distribution.
If $\ell \geq 2$:

$$\mathbb{E}\left[\left(\delta_{1}^{(\iota)}\right)^{-1}\right]\prod_{\gamma=2}^{\ell}\mathbb{E}\left[\left(\delta_{\gamma}^{(\iota)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1}\prod_{\gamma=1}^{\ell}\frac{1}{\alpha_2 - 1} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

□

Remark A.2.1. Let now $\alpha_2 > 1$. Following the same line of reasoning as previously, we

find:

$$\mathbb{E}\left[\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1}\right] = \left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

Proposition A.3 (Variance of the Loading Factors)**.** *If $\alpha_1 > 1$, $\alpha_2 > 1$ and $\nu > 2$, then*

$$\mathbb{V}\text{ar}\left[\Lambda_{j\,\ell}^{(\iota)}\right] = \frac{\nu}{(\nu-2)(\alpha_1-1)(\alpha_2-1)^{\ell-1}}$$

Proof.

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\Lambda_{j\,\ell}^{(\iota)}\right] &= \mathbb{V}\text{ar}\left[\mathbb{E}\left[\Lambda_{j\,\ell}^{(\iota)}\middle|\,\tau_{\lambda\,\ell}^{(\iota)}\right]\right] + \mathbb{E}\left[\mathbb{V}\text{ar}\left[\Lambda_{j\,\ell}^{(\iota)}\middle|\,\tau_{\lambda\,\ell}^{(\iota)}\right]\right] \\
&= \mathbb{E}\left[\mathbb{V}\text{ar}\left[\Lambda_{j\,\ell}^{(\iota)}\middle|\,\tau_{\lambda\,\ell}^{(\iota)}\right]\right] \\
&= \mathbb{E}\left[\frac{\nu}{\nu-2}\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] \\
&= \frac{\nu}{\nu-2}\,\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right]
\end{aligned}$$

□

Proposition A.4 (Variance of the Reciprocal of the Precision $\tau_{\lambda}^{(\iota)}$ of the Loading Factors)**.** *If $\alpha_1 > 2$ and $\alpha_2 > 2$, then:*

$$\mathbb{V}\text{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)\,(\alpha_1-2)\,[(\alpha_2-1)(\alpha_2-2)]^{\ell-1}} \cdot \left[1 - \frac{\alpha_1-2}{\alpha_1-1}\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right]$$

To prove Proposition A.4 we need the following Lemma.

Lemma A.4.1. *Let $0 < A < 1$ and $0 < B < 1$. A recurrence equation of the type:*

$$\begin{aligned}
F_\ell &= \frac{B^2}{1-B}F_{\ell-1} + \frac{A^2}{1-B}B^{2\ell-1} \\
F_1 &= \frac{A^3}{1-A}
\end{aligned}$$

has solution:

$$F_\ell = \frac{A^2}{1-A}\left(\frac{B^2}{1-B}\right)^{\ell-1}\left\{A + (1-A)\left[1-(1-B)^{\ell-1}\right]\right\}$$

Proposition A.4. If $\ell = 1$ then:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,1}^{(\iota)}\right)^{-1}\right] = \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,1}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)^2(\alpha_1-2)}$$

If $\ell \geq 2$ then:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right] + \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right]$$

where

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right] &= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right] \\
&= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\right]\right] \\
&= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right] \\
&= \left[\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right]^2\mathbb{V}\mathrm{ar}\left[\delta_{\lambda\,\ell}^{(\iota)}\right] \\
&= \left[\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right]^2\frac{1}{(\alpha_2-1)^2(\alpha_2-2)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right] &= \mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-2}\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(\iota)}\right]\right] \\
&= \mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\right]\mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-2}\right] \\
&\overset{(A.4.1)}{=} \mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\right] + \frac{1}{(\alpha_2-1)(\alpha_2-2)}
\end{aligned}$$

There we used:

$$\begin{aligned}
\mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-2}\right] &= \mathbb{E}^2\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] + \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \\
&= \left(\frac{1}{\alpha_2-1}\right)^2 + \frac{1}{(\alpha_2-1)^2(\alpha_2-2)} = \frac{1}{(\alpha_2-1)(\alpha_2-2)} \qquad \text{(A.4.1)}
\end{aligned}$$

This proves the following recurrence equation:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_2-1)(\alpha_2-2)}\,\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(\iota)}\right)^{-1}\right] + \frac{1}{\alpha_2-2}\left(\frac{\alpha_2-1}{\alpha_1-1}\right)^2\left(\frac{1}{\alpha_2-1}\right)^{2\ell}$$
$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,1}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)^2(\alpha_1-2)}$$

Finally, we recognize the relationship between the recurrence equation of LemmaA.4.1 and the one for $\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right]$, due to the substitution:

$$A := \frac{1}{\alpha_1-1}, \qquad B := \frac{1}{\alpha_2-1}$$

which induces:

$$\frac{1}{\alpha_1-2} = \frac{A}{1-A}, \qquad \frac{1}{\alpha_2-2} = \frac{B}{1-B}$$

Direct substitution in the solution of Lemma A.4.1 yields:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)\,(\alpha_1-2)}\left[\frac{1}{(\alpha_2-1)\,(\alpha_2-2)}\right]^{\ell-1}\cdot\left[1-\frac{\alpha_1-2}{\alpha_1-1}\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right]$$

□

Remark A.4.1. By observing that

$$\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1} = \prod_{\gamma=2}^{\ell}\delta_{\gamma}^{(\iota)}$$

one can infer the formula for $\mathbb{V}\mathrm{ar}\left[\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1}\right]$ using the same result of Proposition A.4, with the substitution $\alpha_1 = \alpha_2$ and adjusting for the value of ℓ. It follows that

$$\mathbb{V}\mathrm{ar}\left[\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1}\right] = \left[\frac{1}{(\alpha_2-1)\,(\alpha_2-2)}\right]^{\ell-1}\cdot\left[1-\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right] \qquad \forall\alpha_2>2.$$

Remark A.4.2.

$$\alpha_1, \alpha_2 > 2 \quad \Longrightarrow \quad \frac{\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right]}{\mathbb{E}^2\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right]} = \frac{\alpha_1 - 1}{\alpha_1 - 2}\left(\frac{\alpha_2 - 1}{\alpha_2 - 2}\right)^{\ell-1} - 1$$

$$\alpha_2 > 2 \quad \Longrightarrow \quad \frac{\mathbb{V}\mathrm{ar}\left[\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1}\right]}{\mathbb{E}^2\left[\left(\frac{\tau_{\lambda\,\ell}^{(\iota)}}{\tau_{\lambda\,1}^{(\iota)}}\right)^{-1}\right]} = \left(\frac{\alpha_2 - 1}{\alpha_2 - 2}\right)^{\ell-1} - 1$$

Proposition A.5.

$$\mathbb{C}\mathrm{ov}\left[\Lambda_{j_1\,\ell_1}^{(i_1)},\, \Lambda_{j_2\,\ell_2}^{(i_2)}\right] = 0 \qquad \forall (i_1, j_1, \ell_1) \neq (i_2, j_2, \ell_2).$$

Proof. This is a consequence of 1) the independence of the entries of the latent factors conditional on $\boldsymbol{\tau}_\lambda$ and 2) the latent factors are centered conditional on $\boldsymbol{\tau}_\lambda$. □

Remark A.5.1. The covariance is null due to the centered nature of the latent factors, but an internal structure is still present and is captured by the fourth mixed moments $\mathbb{E}\left[\left(\Lambda_{j_1\,\ell_1}^{(i_1)}\Lambda_{j_2\,\ell_2}^{(i_2)}\right)^2\right]$. This allows to forget the sign of the deviation from the mean, and focus only on the co-deviation in absolute terms:

If $(i_1, \ell_1) = (i_2, \ell_2)$, but $j_1 \neq j_2$ then:

$$\begin{aligned}
\left(\frac{\nu - 2}{\nu}\right)^2 \mathbb{E}\left[\left(\Lambda_{j_1\,\ell}^{(\iota)}\Lambda_{j_2\,\ell}^{(\iota)}\right)^2\right] &= \mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-2}\right] \\
&= \mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] + \mathbb{E}^2\left[\left(\tau_{\lambda\,\ell}^{(\iota)}\right)^{-1}\right] \\
&= \frac{1}{\alpha_1 - 1}\frac{1}{\alpha_1 - 2}\left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)}\right]^{\ell-1}.
\end{aligned}$$

If $i_1 = i_2$ but $\ell_1 \neq \ell_2$, define $\ell_m := \min(\ell_1, \ell_2)$, $\ell_M := \max(\ell_1, \ell_2)$:

$$\begin{aligned}
\left(\frac{\nu - 2}{\nu}\right)^2 \mathbb{E}\left[\left(\Lambda_{j_1\,\ell_1}^{(\iota)}\Lambda_{j_2\,\ell_2}^{(\iota)}\right)^2\right] &= \mathbb{E}\left[\left(\tau_{\lambda\,\ell_1}^{(\iota)}\right)^{-1}\left(\tau_{\lambda\,\ell_2}^{(\iota)}\right)^{-1}\right] \\
&= \left\{\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell_m}^{(\iota)}\right)^{-1}\right] + \mathbb{E}^2\left[\left(\tau_{\lambda\,\ell_m}^{(\iota)}\right)^{-1}\right]\right\}\left(\frac{1}{\alpha_2 - 1}\right)^{|\ell_1 - \ell_2|} \\
&= \frac{1}{\alpha_1 - 1}\frac{1}{\alpha_1 - 2}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell_M - 1}\left(\frac{1}{\alpha_2 - 2}\right)^{\ell_m - 1}
\end{aligned}$$

Finally, if $i_1 \neq i_2$:

$$\begin{aligned}
\left(\frac{\nu-2}{\nu}\right)^2 \mathbb{E}\left[\left(\Lambda^{(1)}_{j_1\,\ell_1}\Lambda^{(2)}_{j_2\,\ell_2}\right)^2\right] &= \mathbb{E}\left[\left(\tau^{(1)}_{\lambda\,\ell_1}\right)^{-1}\left(\tau^{(2)}_{\lambda\,\ell_2}\right)^{-1}\right] \\
&= \mathbb{E}\left[\left(\tau^{(1)}_{\lambda\,\ell_1}\right)^{-1}\right]\mathbb{E}\left[\left(\tau^{(2)}_{\lambda\,\ell_2}\right)^{-1}\right] \\
&= \left(\frac{1}{\alpha_1-1}\right)^2\left(\frac{1}{\alpha_2-1}\right)^{\ell_1+\ell_2-2} .
\end{aligned}$$

B | Prior Predictive Moments

Let us begin with a result already discussed in Section 3.1.1 of the main text:

Proposition B.1 (Prior Predictive Distribution).

$$\boldsymbol{Y}_s | \Theta \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}X_{\cdot,s} \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}X_{\cdot,s} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \qquad s = 1, \ldots, N$$

where

$$\boldsymbol{\Sigma}_{ii} = {\psi^{(\iota)}}^2 \boldsymbol{I} + \boldsymbol{B}^{(\iota)} \left[\Delta_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(\iota)} {\boldsymbol{\Lambda}^{(\iota)}}^T \right] {\boldsymbol{B}^{(\iota)}}^T$$
$$\boldsymbol{\Sigma}_{ij} = \boldsymbol{B}^{(\iota)} \boldsymbol{\Lambda}^{(\iota)} {\boldsymbol{\Lambda}^{(j)}}^T {\boldsymbol{B}^{(j)}}^T$$

Proposition B.2 (Expectation of the Prior Predictive).

$$\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)}\right] = \boldsymbol{0}, \qquad \forall s = 1, \ldots, N, \forall \iota = 1, 2$$

Proof.

$$\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)}\right] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)} \middle| \Theta\right]\right] = \mathbb{E}\left[\boldsymbol{B}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right] = \sum_{j=1}^{p_\iota}\sum_{\ell=1}^{k}\sum_{i=1}^{r} {\boldsymbol{B}_{\cdot,j}}^{(\iota)} \mathbb{E}\left[\Lambda_{j\ell}^{(\iota)}\right] \mathbb{E}\left[\beta_{\ell i}\right] X_{i,s} = \boldsymbol{0}$$

□

Proposition B.3 (Variance of the Prior Predictive). *Assuming $\nu, \chi > 2$ and $\alpha_\sigma, \alpha_\psi, \alpha_1, \alpha_2 > 1$. If $\alpha_2 \neq 2$, then*

$$\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\right] = \frac{\beta_\psi}{\alpha_\psi - 1} + \left(\sum_{j=1}^{p_\iota} {B_{h,j}^{(\iota)}}^2\right) \cdot \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \right.$$
$$\left. + \frac{\nu}{\nu - 2} \frac{\alpha_2 - 1}{\alpha_1 - 1} \frac{1}{\alpha_2 - 2} \left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right] \left[1 + \frac{\chi}{\chi - 2}\left(\sum_{i=1}^{r} X_{i,s}^2\right)\right] \right\}$$

If $\alpha_2 = 2$, then

$$\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\right] = \frac{\beta_\psi}{\alpha_\psi - 1} + \left(\sum_{j=1}^{p_\iota} {B_{h,j}^{(\iota)}}^2\right) \cdot \left\{\frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu - 2}\frac{k}{\alpha_1 - 1}\left[1 + \frac{\chi}{\chi - 2}\left(\sum_{i=1}^{r} X_{i,s}^2\right)\right]\right\}$$

Proof.

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\right] &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\middle|\Theta\right]\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\middle|\Theta\right]\right] \\
\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\middle|\Theta\right]\right] &= \mathbb{E}\left[{\psi^{(\iota)}}^2 + \boldsymbol{B}_{h,\cdot}^{(\iota)}\left[\Delta_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(\iota)}{\boldsymbol{\Lambda}^{(\iota)}}^T\right]{\boldsymbol{B}_{h,\cdot}^{(\iota)}}^T\right] \\
&= \frac{\beta_\psi}{\alpha_\psi - 1} + \boldsymbol{B}_{h,\cdot}^{(\iota)}\left\{\mathbb{E}\left[\Delta_\theta^{(\iota)}\right] + \mathbb{E}\left[\boldsymbol{\Lambda}^{(\iota)}{\boldsymbol{\Lambda}^{(\iota)}}^T\right]\right\}{\boldsymbol{B}_{h,\cdot}^{(\iota)}}^T \\
\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\middle|\Theta\right]\right] &= \mathbb{V}\mathrm{ar}\left[\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right] = \mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right)^2\right]
\end{aligned}$$

□

Lemma B.3.1.

$$\mathbb{E}\left[\Delta_\theta^{(\iota)}\right] = \frac{\beta_\sigma}{\alpha_\sigma - 1}\boldsymbol{I}$$

Lemma B.3.2.

$$\mathbb{E}\left[\boldsymbol{\Lambda}^{(\iota)}{\boldsymbol{\Lambda}^{(\iota)}}^T\right] = \frac{\nu}{\nu - 2}\frac{\alpha_2 - 1}{\alpha_1 - 1}\frac{1}{\alpha_2 - 2}\left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right]\boldsymbol{I}$$

Proof. It is sufficient observe that

$$\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}{\boldsymbol{\Lambda}^{(\iota)}}^T\right)_{j_1 j_2}\right] = \sum_{\ell=1}^{k}\mathbb{E}\left[\Lambda_{j_1\ell}^{(\iota)}\Lambda_{j_2\ell}^{(\iota)}\right] = \begin{cases} \sum_{\ell=1}^{k}\frac{\nu}{\nu-2}\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-1} & \text{if } j := j_1 = j_2, \\ 0 & \text{if } j_1 \neq j_2. \end{cases}$$

and that

$$\sum_{\ell=1}^{k}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1} = \frac{1 - \left(\frac{1}{\alpha_2-1}\right)^k}{1 - \frac{1}{\alpha_2-1}} = \frac{\alpha_2 - 1}{\alpha_2 - 2}\left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right]$$

(with the understanding that in the case $\alpha_2 = 2$ we get k). □

Lemma B.3.3.

$$\mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right)^2\right] = \frac{\chi}{\chi-2}\frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\sum_{j=1}^{p_\iota}\sum_{i=1}^{r}{B_{h,j}^{(\iota)}}^2 X_{i,s}^2$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right)^2\right] &= \mathbb{E}\left[\sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota} B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_1,s}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_2,s}\right] \\
&= \sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota} B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}\,\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_1,s}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_2,s}\right] \\
&= \sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r} B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}X_{i_1,s}X_{i_2,s}\,\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_1,i_1}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_2,i_2}\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_1,i_1}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_2,i_2}\right] &= \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\beta_{\ell_1,i_1}\beta_{\ell_2,i_2}\right] \\
&= \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\right]\mathbb{E}\left[\beta_{\ell_1,i_1}\beta_{\ell_2,i_2}\right] \\
&= \frac{\chi}{\chi-2}\sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\right]\delta^K\left(\ell_1,\ell_2\right)\delta^K\left(i_1,i_2\right) \\
&= \frac{\chi}{\chi-2}\sum_{\ell=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell}^{(\iota)}\Lambda_{j_2,\ell}^{(\iota)}\right]\delta^K\left(i_1,i_2\right) \\
&= \frac{\chi}{\chi-2}\frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\delta^K\left(i_1,i_2\right)\delta^K\left(j_1,j_2\right)\ .
\end{aligned}$$

□

Proposition B.4 (Limiting behavior of the Variance of the Prior Predictive)**.** *Let* $\nu,\chi,\alpha_2>2$ *and* $\alpha_\sigma,\alpha_\psi,\alpha_1>1$*, then*

$$\lim_{k\to+\infty}\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)}\right] = \frac{\beta_\psi}{\alpha_\psi-1}+\left(\sum_{j=1}^{p_\iota}{B_{h,j}^{(\iota)}}^2\right)\cdot\left\{\frac{\beta_\sigma}{\alpha_\sigma-1}+\frac{\nu}{\nu-2}\frac{1}{\alpha_1-1}\frac{\alpha_2-1}{\alpha_2-2}\left[1+\frac{\chi}{\chi-2}\left(\sum_{i=1}^{r}X_{i,s}^2\right)\right]\right\}$$

Proof. The thesis is a direct consequence of the fact that

$$\lim_{k\to+\infty}\left(\frac{1}{\alpha_2-1}\right)^k = 0 \qquad \forall\alpha_2>2$$

□

Proposition B.5 (Total Variance of the Prior Predictive).

Let $\nu, \chi, \alpha_2 > 2$ and $\alpha_\sigma, \alpha_\psi, \alpha_1 > 1$:

$$\Sigma_T^k(\boldsymbol{Y}_s) := \left(\sum_{h_1=1}^{L_1} \mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h_1,s}\right]\right) + \left(\sum_{h_2=1}^{L_2} \mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h_2,s}\right]\right) = (L_1+L_2)\frac{\beta_\psi}{\alpha_\psi - 1} + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right) \times$$
$$\times \left\{\frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu-2}\frac{\alpha_2 - 1}{(\alpha_1-1)(\alpha_2-2)}\left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right]\left(1 + \frac{\chi}{\chi - 2}\|\boldsymbol{X}_{\cdot,s}\|_2^2\right)\right\}$$

$$\Sigma_T^\infty(\boldsymbol{Y}_s) = (L_1+L_2)\frac{\beta_\psi}{\alpha_\psi - 1} + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)\left\{\frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu-2}\frac{\alpha_2 - 1}{(\alpha_1-1)(\alpha_2-2)}\left(1 + \frac{\chi}{\chi - 2}\|\boldsymbol{X}_{\cdot,s}\|_2^2\right)\right\}$$

where

$$\boldsymbol{B} := \begin{bmatrix} \boldsymbol{B}^{(1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}^{(2)} \end{bmatrix} \quad .$$

Proposition B.6 (Covariance of the Prior Predictive).
Assuming $\nu, \chi > 2$ and $\alpha_\sigma, \alpha_\psi, \alpha_1, \alpha_2 > 1$:

If $\alpha_2 \neq 2$ and $h_1 \neq h_2$, then

$$\mathbb{C}\mathrm{ov}\left[Y_{h_1,s}^{(\iota)}, Y_{h_2,s}^{(\iota)}\right] = \left(\sum_{j=1}^{p_\iota} B_{h_1,j}^{(\iota)} B_{h_2,j}^{(\iota)}\right) \times$$
$$\times \left\{\frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu-2}\frac{\alpha_2 - 1}{\alpha_1 - 1}\frac{1}{\alpha_2 - 2}\left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right]\left[1 + \frac{\chi}{\chi - 2}\left(\sum_{i=1}^r X_{i,s}^2\right)\right]\right\}$$

If $\alpha_2 = 2$ and $h_1 \neq h_2$, then

$$\mathbb{C}\mathrm{ov}\left[Y_{h_1,s}^{(\iota)}, Y_{h_2,s}^{(\iota)}\right] = \left(\sum_{j=1}^{p_\iota} B_{h_1,j}^{(\iota)} B_{h_2,j}^{(\iota)}\right) \cdot \left\{\frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu-2}\frac{\alpha_2 - 1}{\alpha_1 - 1}\frac{k}{\alpha_2 - 2}\left[1 + \frac{\chi}{\chi - 2}\left(\sum_{i=1}^r X_{i,s}^2\right)\right]\right\}.$$

Moreover,

$$\mathbb{C}\mathrm{ov}\left[Y_{h_1,s}^{(1)}, Y_{h_2,s}^{(2)}\right] = 0$$

Proof. The proof of this result is obtained by following the exact same steps as in the proof of Proposition B.3, with the only difference being the replacement of the total variance

formula with the total covariance formula:

$$\mathbb{C}\,\mathrm{ov}\left[Y_{h_1,s}^{(\iota_1)}, Y_{h_2,s}^{(\iota_2)}\right] = \mathbb{E}\left[\mathbb{C}\,\mathrm{ov}\left[Y_{h_1,s}^{(\iota_1)}, Y_{h_2,s}^{(\iota_2)}\middle|\,\Theta\right]\right] + \mathbb{C}\,\mathrm{ov}\left[\mathbb{E}\left[Y_{h_1,s}^{(\iota_1)}\middle|\,\Theta\right], \mathbb{E}\left[Y_{h_2,s}^{(\iota_2)}\middle|\,\Theta\right]\right]$$

□

Remark B.6.1. The covariance between different functional targets amounts to zero because of Proposition A.5. The Remark A.5.1 still holds, and we expect to capture more details if we looked at the fourth mixed moments.

Proposition B.7 (Between samples Covariance of the Prior Predictive)**.**
Assuming $\nu, \chi > 2$ and $\alpha_\sigma, \alpha_\psi, \alpha_1, \alpha_2 > 1$, $\alpha_2 \neq 2$, then

$$\mathbb{C}\,\mathrm{ov}\left[\boldsymbol{Y}_{h_1,s_1}^{(\iota_1)}, \boldsymbol{Y}_{h_2,s_2}^{(\iota_2)}\right] = \\ = \begin{cases} \frac{\chi}{\chi-2}\frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\sum_{j=1}^{p_\iota}\sum_{i=1}^{r} B_{h_1,j}^{(\iota)} B_{h_2,j}^{(\iota)} X_{i,s_1} X_{i,s_2} & \text{if } \iota := \iota_1 = \iota_2 \\ 0 & \text{if } \iota_1 \neq \iota_2 \end{cases}$$

Proof.

$$\begin{aligned} \mathbb{C}\mathrm{ov}\left[\boldsymbol{Y}_{h_1,s_1}^{(\iota_1)}, \boldsymbol{Y}_{h_2,s_2}^{(\iota_2)}\right] &= \mathbb{E}\left[\mathbb{C}\mathrm{ov}\left[\boldsymbol{Y}_{h_1,s_1}^{(\iota_1)}, \boldsymbol{Y}_{h_2,s_2}^{(\iota_2)}\middle|\,\Theta\right]\right] + \mathbb{C}\mathrm{ov}\left[\mathbb{E}\left[\boldsymbol{Y}_{h_1,s_1}^{(\iota_1)}\middle|\,\Theta\right], \mathbb{E}\left[\boldsymbol{Y}_{h_2,s_2}^{(\iota_2)}\middle|\,\Theta\right]\right] \\ &= \mathbb{C}\mathrm{ov}\left[\left(\boldsymbol{B}_{h_1,\cdot}^{(\iota_1)}\boldsymbol{\Lambda}^{(\iota_1)}\boldsymbol{\beta}X_{\cdot,s_1}\right), \left(\boldsymbol{B}_{h_2,\cdot}^{(\iota_2)}\boldsymbol{\Lambda}^{(\iota_2)}\boldsymbol{\beta}X_{\cdot,s_2}\right)\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{B}_{h_1,\cdot}^{(\iota_1)}\boldsymbol{\Lambda}^{(\iota_1)}\boldsymbol{\beta}X_{\cdot,s_1}\right)\left(\boldsymbol{B}_{h_2,\cdot}^{(\iota_2)}\boldsymbol{\Lambda}^{(\iota_2)}\boldsymbol{\beta}X_{\cdot,s_2}\right)\right] \\ &= \sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r} B_{h_1,j_1}^{(\iota_1)} B_{h_2,j_2}^{(\iota_2)} X_{i_1,s_1} X_{i_2,s_2}\,\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota_1)}\boldsymbol{\beta}\right)_{j_1,i_1}\left(\boldsymbol{\Lambda}^{(\iota_2)}\boldsymbol{\beta}\right)_{j_2,i_2}\right] \end{aligned}$$

□

C | Calibration of Post-Varimax-RSP Latent Components

We simulated $S = 500$ datasets by sampling from the prior distribution of the BBLF model. The choice for the basis function was that of 10 cubic b-splines for $\boldsymbol{Y}^{(1)}$, and Fourier basis truncated to the third harmonic, plus a linear trend for $\boldsymbol{Y}^{(2)}$. We set $k = 3$ latent factors. 5 covariates per data point were drawn independently from a normal distribution with a standard deviation of 0.3. We sampled $N = 100$ data points $\boldsymbol{Y}$. Observations for $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ were made on $L_1 = L_2 = 40$ evenly spaced points over the intervals $(-1, 1)$ and $(0, 5)$, respectively. We set $\alpha_1 = 2.1$, $\alpha_2 = 2.5$, $\nu = \chi = 4$.

The posterior Monte Carlo Markov Chain samples are taken using Hamiltonian Monte Carlo No U-turn sampler [24], with 4 chains, a burn-in of 500 samples, maximum tree depth of 12. 1000 samples were drawn per chain.

Varimax-RSP was applied to the latent component's posterior distribution, and the *true* latent components were aligned to the resulting Monte Carlo samples as described in Algorithm 3.2.

For each of the $S = 500$ samples from the prior distribution, for each latent component $\ell = 1, \ldots, k$, we compute the number of locations of measurement such that the resulting 90% Highest Density Interval of the Varimaxed Posterior latent components samples contains the *true* latent component at that location $(\boldsymbol{B}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)})_{h,\ell}$.

The results are stored in $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ respectively, which are vectors of length S. If the algorithm producing the samples produces well-calibrated samples, we expect that

$$\frac{1}{S}\sum_{s=1}^{S}\frac{Z_s^{(1)}}{kL_1} \overset{S\to+\infty}{\longrightarrow} 0.9, \qquad \frac{1}{S}\sum_{s=1}^{S}\frac{Z_s^{(2)}}{kL_2} \overset{S\to+\infty}{\longrightarrow} 0.9$$

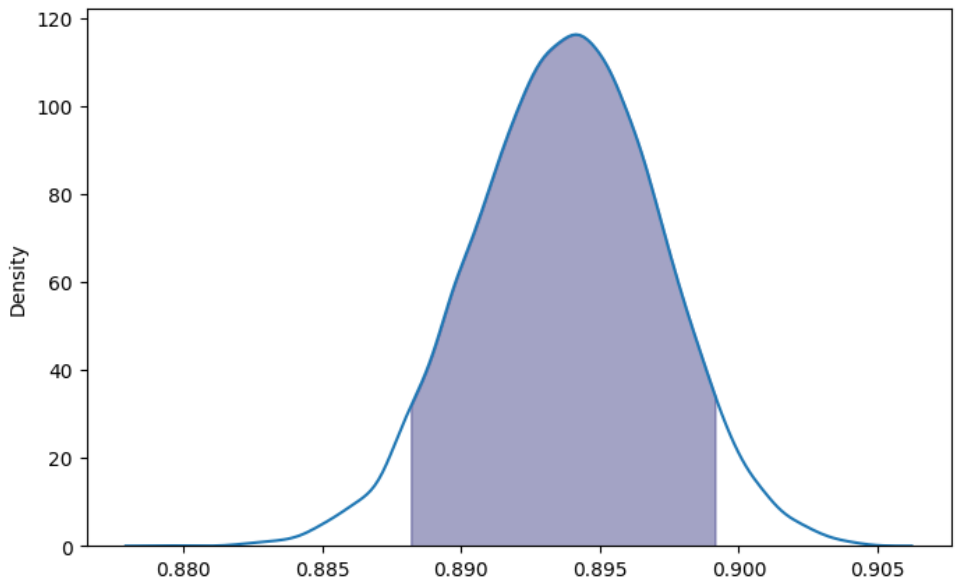


Figure C.1: The blue line depicts the Kernel Density Estimation of the parameter μ, the shaded region of the figure delimits the 90% Highest Density Interval.

The results are:

$$\frac{1}{500}\sum_{s=1}^{500}\frac{Z_s^{(1)}}{120} \simeq 0.896, \qquad \frac{1}{500}\sum_{s=1}^{500}\frac{Z_s^{(2)}}{120} \simeq 0.893$$

To analyze the results we applied the following robust Bayesian model:

$$\begin{aligned}
Z_s^{(1)}|p_s^{(1)} &\sim \text{Binomial}(120, p_s^{(1)}) \\
Z_s^{(2)}|p_s^{(2)} &\sim \text{Binomial}(120, p_s^{(2)}) \\
p_s^{(1)}|\mu, K, K_s^{(1)} &\sim \text{Beta}(K \cdot K_s^{(1)} \cdot \mu, K \cdot K_s^{(1)} \cdot (1-\mu)) \\
p_s^{(2)}|\mu, K, K_s^{(2)} &\sim \text{Beta}(K \cdot K_s^{(2)} \cdot \mu, K \cdot K_s^{(2)} \cdot (1-\mu)) \\
K_s^{(1)} &\sim \text{Ga}(15, 15) \\
K_s^{(2)} &\sim \text{Ga}(15, 15) \\
K &\sim \text{Ga}(2, 2) \\
\mu &\sim \text{Beta}(4, 2)
\end{aligned}$$

The resulting properties of the posterior distribution of the quantity of interest μ, given

$Z^{(1)}$ and $Z^{(2)}$, are as follows:

$$\begin{aligned}
\mathbb{E}[\mu|Z^{(1)},\, Z^{(2)}] &\simeq 89.38\% \\
\mathrm{HDI}_L^{90\%}[\mu|Z^{(1)},\, Z^{(2)}] &\simeq 88.8\% \\
\mathrm{HDI}_U^{90\%}[\mu|Z^{(1)},\, Z^{(2)}] &\simeq 89.9\% \\
\mathbb{P}[\mu > 0.9|Z^{(1)},\, Z^{(2)}] &\simeq 2.9\%
\end{aligned}$$

indicating a divergence from calibration by a small effect size, which should not be of importance in most practical scenarios.

D | Clustering and Correlation Analysis between Absorbance and Fluorescence spectra

A parallel approach was considered at the beginning of the thesis work, which was based on hierarchical clustering techniques.

Due to the unsupervised nature of this problem (i.e., lacking prior knowledge about the actual content of NOM in each water sample), the idea was to apply unsupervised clustering techniques separately to absorbance spectra and excitation-emission matrices describing fluorescence spectra. If the two clustering approaches provided similar results, it would suggest the possibility of deducing fluorescence spectra based on absorbance spectra in certain instances.

Subsequently, we discuss the data pre-processing steps performed, the dissimilarity measures used, the hierarchical clustering techniques applied, and the post-processing strategies employed to derive meaningful insights from these spectroscopic measurements.

D.1. Data Pre-processing

The pre-processing steps have been designed to address noise, reconstruct missing data, and standardize sampling values across datasets developed by different teams with several measurement tools. We recall that UV-Vis and LC-OCD measurements are represented as vectors, while fluorescence data is represented as a matrix. However, the following methodologies are applicable to both types of data.

For one-dimensional data such as UV-Vis absorbance spectra and LC-OCD profiles, we smooth them by means of cubic smoothing splines [29]. Cubic smoothing splines provide a flexible and effective means to smooth the data while preserving important features and trends. They help mitigate random fluctuations and outliers, resulting in cleaner and more reliable data for downstream analysis.

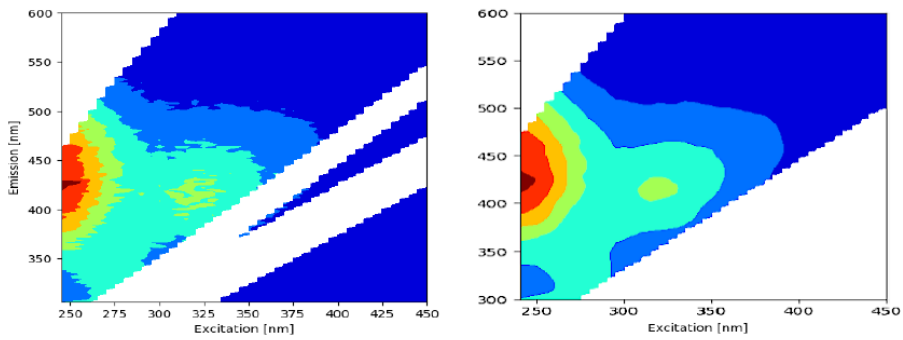


Figure D.1: Fluorescence spectrum that has been smoothed and limited to the relevant region for fluorescence spectroscopy for NOM characterization. Raw data (left) and smoothed version (right).

Instead, in the case of two-dimensional data, like fluorescence spectra, we use the thin-plate smoothing splines, that make multidimensional data well-suited for smoothing 2D spectra. Thin-plate smoothing splines maintain the integrity of spatial relationships within the data while reducing noise and enhancing the clarity of underlying patterns. An example of this process is presented in Figure D.1.

The Generalized Cross-Validation (GCV) technique was used to fine-tune the smoothing parameters required for both spline methods. The GVC method enables us to find the optimal equilibrium between reducing the erratic fluctuations in the output function while preserving the fundamental patterns exhibited by the data. The documentation on the libraries used for this analysis is available in the R software documentation [6].

Since the output of spline smoothing consists of mathematical functions, there is the flexibility to sample them at any frequency and time. This ensures both uniform sampling across all available datasets, and the computation of spectra at specific points, highlighted as significant in the NOM characterization literature [14, 31].

The smoothed data are then normalized, dividing them by the area/volume under the curve/surface. Working with such normalized data allows us to focus the analysis solely on the shapes and disregard about the absolute intensity of the data.

D.2. Hierarchical Clustering

Now we discuss the clustering technique that was employed.

D.2.1. Dissimilarity Measure

A hierarchical clustering procedure is based on the choice of a dissimilarity measure between samples. For our spectroscopic data analysis, we opt for the Canberra distance, which is a standard choice in the framework of hierarchical clustering when dealing with positive quantities [12]. For two points x, y its formal definition is as follows:

$$D_C(x, y) = \frac{|x - y|}{|x| + |y|}$$

As a consequence, we define the dissimilarity between two absorbance spectra as:

$$D(A_1, A_2) = \int \frac{|A_1(\omega) - A_2(\omega)|}{|A_1(\omega)| + |A_2(\omega)|} d\omega$$

We take account of the discretization induced by the sampling of the spectra:

$$D(A_1, A_2) = \sum_i \frac{|A_1(\omega_i) - A_2(\omega_i)|}{|A_1(\omega_i)| + |A_2(\omega_i)|}$$

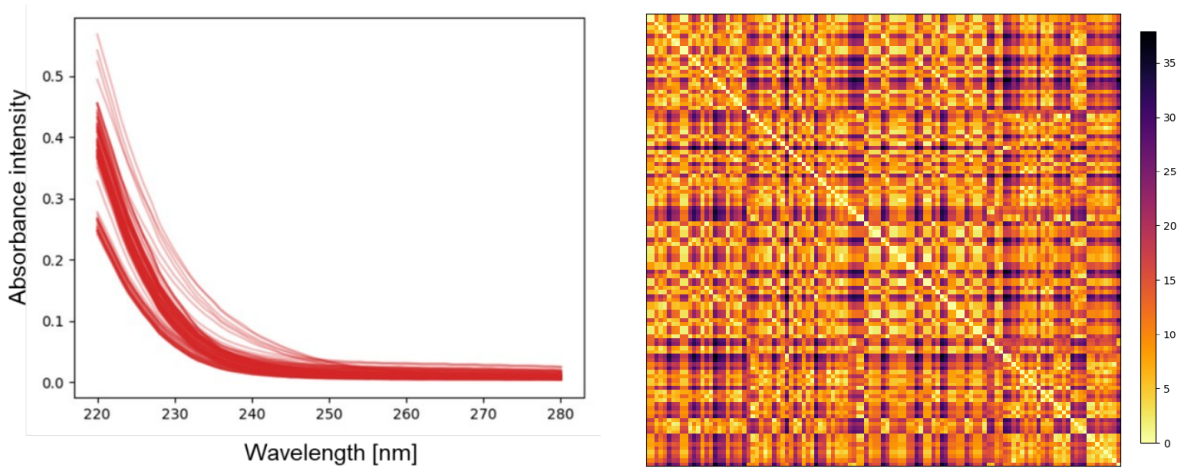


Figure D.2: Conversion of a dataset of absorbance spectra (left) to a matrix of pairwise dissimilarities (right).

An example of a heatmap of the dissimilarities between pairs of absorbance spectra is shown in Figure D.2: on the left, we have some different spectra (wavelength vs. intensity), and on the right, each pixel of the map represents a pairwise dissimilarity (the closer the two absorbance spectra are, the lighter the pixel).

D.2.2. Clustering

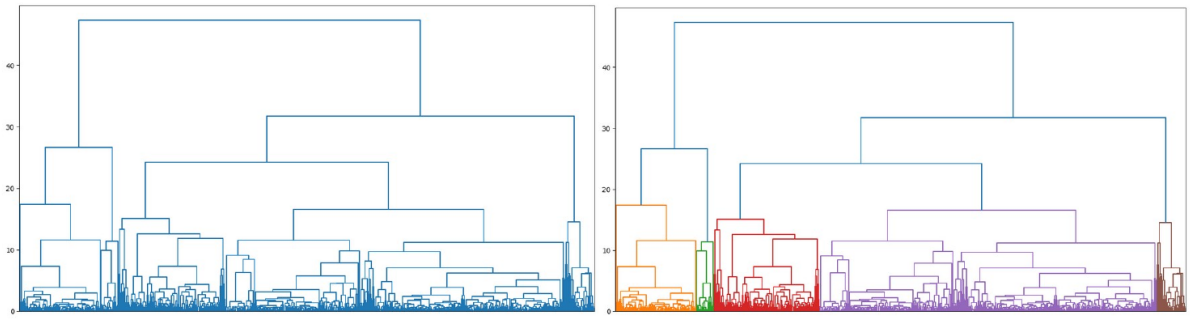


Figure D.3: Dendrogram (left) and corresponding clustering generated from a cutting (right).

After calculating the distances between different samples' measurements, groups of samples with similar NOM characteristics (measured by each analytical method) were identified. Hierarchical clustering is a powerful data analysis technique developed to address unsupervised learning tasks [8]. This technique groups data points into a hierarchical tree-like structure known as a dendrogram. It proves particularly valuable when exploring complex datasets with the aim of identifying inherent structures, patterns, or relationships.

An example of a dendrogram is provided in Figure D.3 (left). At the bottom of the tree, we find the leaves, each representing a sample. As we move up from the bottom, branching represents the clustering of nearest samples. The distance between two distinct clusters is determined through a process called "linking". In our case, we have chosen an average linking method, which estimates the distance between two clusters by averaging all pairwise distances between the samples in those clusters. This approach allows us to group different clusters together using the same rule at each level. Following this process, we eventually cluster all samples into a single cluster given by the highest part of the tree known as the root.

Finally, the analysis of the results of this procedure involves studying the differences between each cluster while varying the analytical tool used for clustering and the depth at which we cut the tree. An example of cutting the tree to generate 5 clusters, each depicted in distinct colours, is shown on the right in Figure D.3.

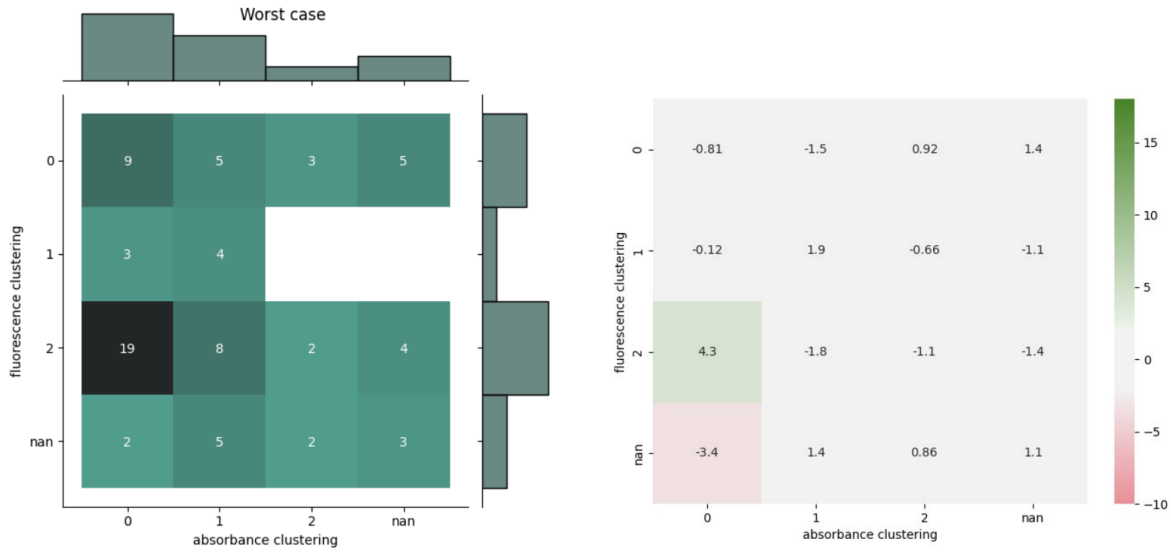


Figure D.4: Example of worst-case scenario. Contingency table (left) and transformed values (right).

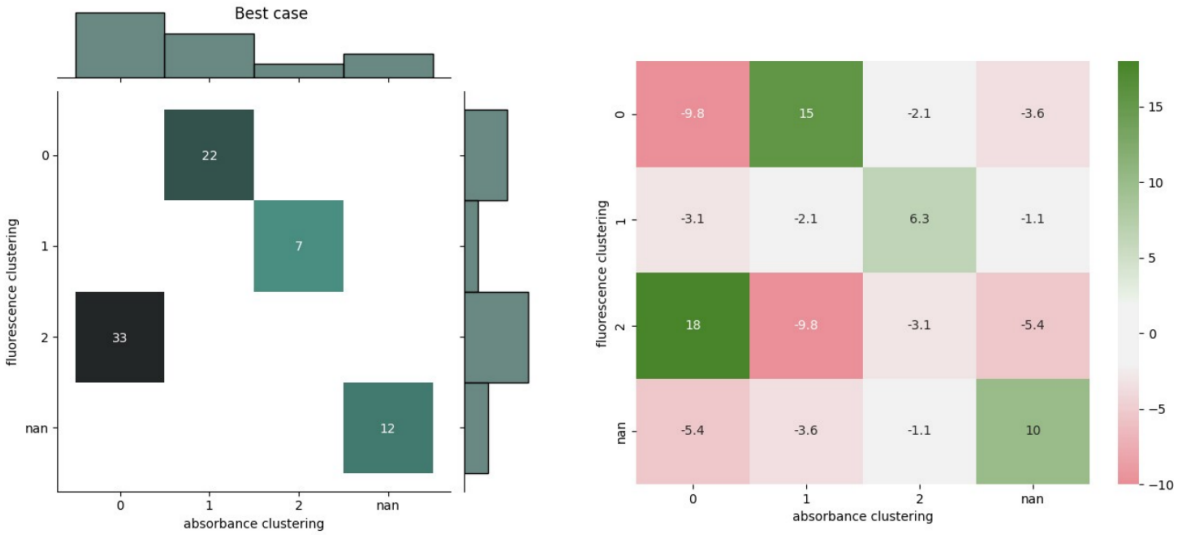


Figure D.5: Example of best-case scenario. Contingency table (left) and transformed values (right).

D.2.3. Post-processing

After clustering the samples according to each measurement method, a post-processing is needed to correlate results obtained with different methods. In fact, the post-processing tools are essential for translating the results of our methodology into actionable insights for future research and practical applications. A contingency table was constructed, as it is

commonly used for discovering and visualizing relationship between two attributes. This technique organizes data into rows and columns, where each cell in the table represents the intersection of corresponding categories for each attribute. In our context, we used these tables (see two examples in Figure D.4 and Figure D.5) to identify relationships between pairs of clustering: on the columns we find the different clusters obtained by absorbance spectra, and in the rows the clusters obtained by the fluorescence EEM. In the contingency table, in each cell it is reported the number of samples falling into the intersection between two clusters built with each measurement method. Furthermore, we introduce a transformation that highlights the clusters with higher correlations. This transformation is colour- coded to represent positive correlations in green and negative correlations in red. The transformation is mathematically defined as the difference between the contingency matrix and the tensor product of the marginal empirical distributions, which are visualized as histograms within the contingency matrix.

We present two extreme examples, both to illustrate the construction of the matrices described above and the potential outcomes of this analysis.

The worst-case scenario represents complete independence among clusters, where knowledge of the clustering for one variable (i.e. absorbance) gives no information about the other (i.e. fluorescence); in fact, for each cluster of absorbance there is equal probability to fall in each one of the fluorescence clusters and the correlations (shown in the right graph) are low. An example of this kind of contingency table (left) and the transformed values (right) are presented in Figure D.4.

Conversely, the best-case scenario, exemplified in Figure D.5, is when for each cluster of one feature (i.e. absorbance), we have a 100%-correspondence with one cluster for the other (i.e. fluorescence). Therefore, we can identify the cluster of a measurement by only providing the corresponding cluster for the other measurement. In this case, only measuring the absorbance spectrum of one sample, we would know in which absorbance cluster it falls and, consequently we would know also in which fluorescence cluster it would fall and which fluorescence EEM characteristics it has, without the need for fluorescence measurements.

D.2.4. Results

We show an example based on a historically available dataset where NOM has been characterized by both UV absorbance spectra and fluorescence EEM. This dataset consists of 113 water samples taken from a DWTP with river source water at three different times between June and July 2021. The dataset was assembled by recording absorbance and

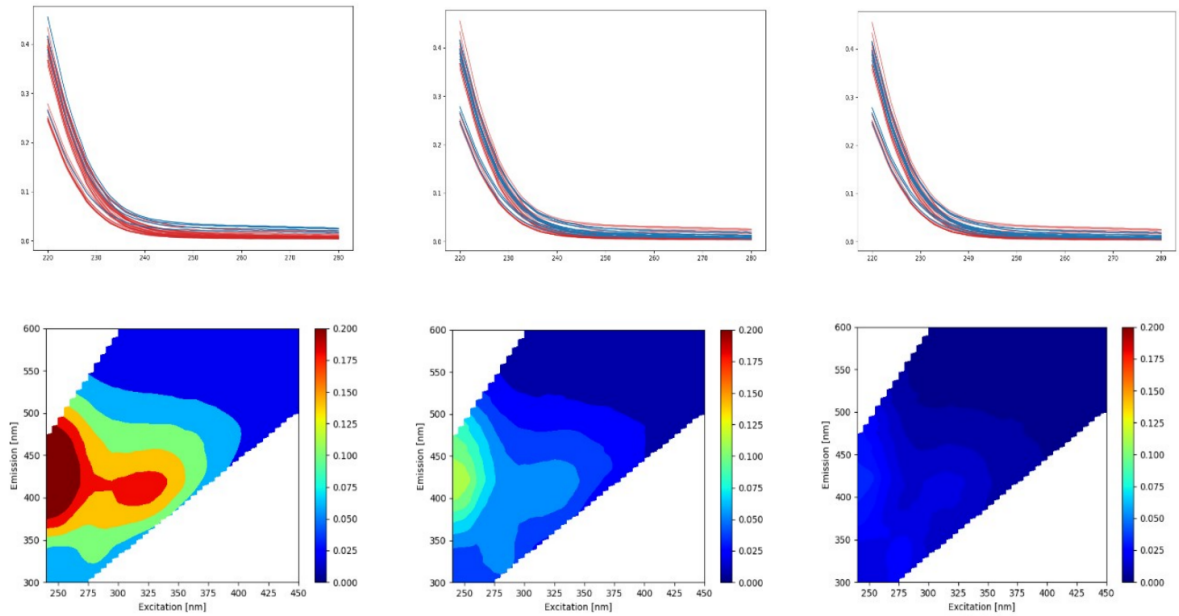


Figure D.6: Top absorbance spectra, bottom the average fluorescence spectra, from left to right: before the ozonation, after the ozonation, after the activated carbon treatment.

fluorescence spectra for samples subjected to various treatment combinations, encompassing an ozonation treatment and 48-hour contact with activated carbon [3]. Absorbance spectra and the average fluorescence are presented in Figure D.6.

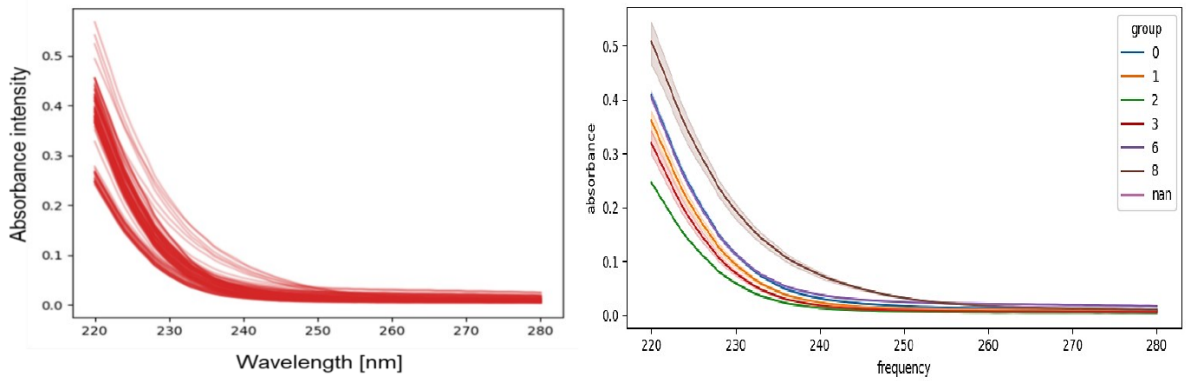


Figure D.7: Clusters produced by cutting the absorbance spectra hierarchical tree at 10 clusters (clusters with fewer than 5 samples are grouped together under the 'nan' cluster) Left: original dataset, Right: average absorbance spectrum per cluster and corresponding pointwise 90% confidence interval.

Figure D.7 presents the absorbance spectra (left) and the corresponding centres of the clusters (right) obtained by applying hierarchical clustering. We can already discern a complex structure in the data, revealing at least two distinct clusters for the absorbance

spectra. Conversely, we can directly observe the impact of advancing the treatment through the diminishing intensity of the spectra, it is important to note that the observed clustering is not directly linked to the treatment effect.

The contingency table for the absorbance vs fluorescence clustering is presented in Figure D.8 (left) as well as the corresponding transformed values (right). We compared the clusters we obtain through absorbance spectra and those of fluorescence spectra. We have 6 clusters created from the absorbance spectra, one per columns, and along the rows the 8 clusters corresponding to the fluorescence spectra. Cells with darker colours indicate a higher number of shared samples, while blank spaces represent clusters with no shared samples.

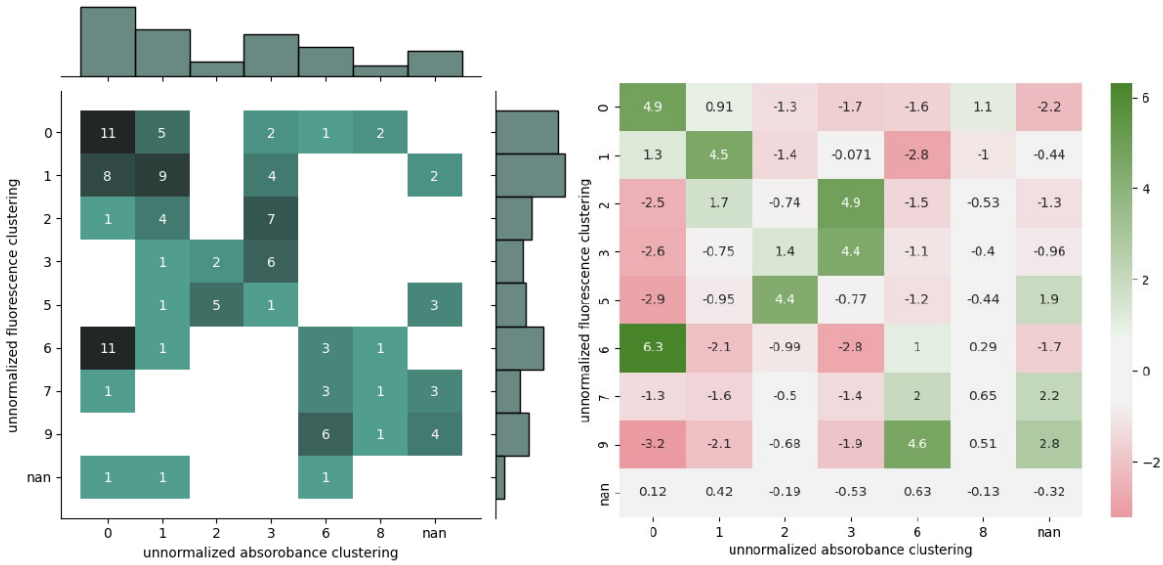


Figure D.8: Clustering with respect to absorbance spectra vs fluorescence spectra

Clusters 0, 1, 3, and 6 are the most numerous for the unnormalized absorbance clustering. Focusing on them we can already extrapolate some information, for instance, the white regions in the matrix in the left part of Figure D.8, tell us that being in cluster 3 for the absorbance implies that such a sample cannot belong to clusters 6-9 for the fluorescence. This allows having only a limited variety of fluorescence for this data. Similarly, if the sample belongs to cluster 0 for the absorbance, it will most likely also belong to cluster 0 or 6 for the fluorescence. The green regions in the right matrix of Figure D.8 show the associative pattern between the clusters. For example, being in cluster 6 for the absorbance sharply increases the likelihood of being in cluster 9 for the fluorescence.

Further analysis can be conducted by looking at the average fluorescence spectra of the clusters a sample might belong to. For instance, cluster 3 for the absorbance is associ-

ated with clusters 2 or 3 for the fluorescence In Figure D.9, we report the normalized fluorescence spectra for these two clusters (cluster 2 on the left and 3 on the right), this allows us to focus on the shape difference between the two spectra. In particular, the two clusters have similar shapes except that Cluster 3 has a peak in the area of the typical peak T (corresponding to protein-like NOM).

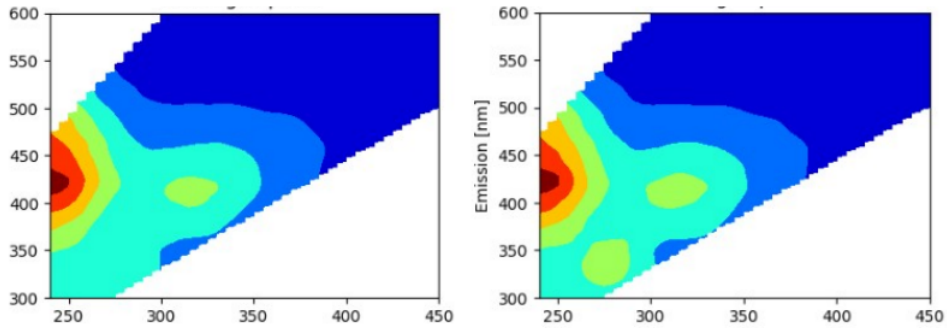


Figure D.9: Normalized fluorescence spectra for clusters 2 (left) and 3 (right) of the fluorescence, the shape difference shows that moving from cluster 2 to cluster 3 corresponds to a redistribution of fluorescence intensity from region III, towards regions IV a V of the regional integration approach commonly used in fluorescence spectroscopy analysis.

E | Stan Code

```
data {
  // dimensions
  int<lower=1> N;
  int<lower=1> L1;
  int<lower=1> L2;
  int<lower=1> p1;
  int<lower=1> p2;
  int<lower=1> k;
  int<lower=1> r;

  // data
  matrix[r,N] X;

  // transformed data trick
  matrix[p1,N] proj_y1;     // B_1^+ y1
  matrix[p2,N] proj_y2;     // B_2^+ y2
  vector[N] ress_y1;        // sqrt(Y'M'MY)
  vector[N] ress_y2;        // sqrt(Y'M'MY)
  matrix[p1,p1] qr_decompose_R1;
  matrix[p2,p2] qr_decompose_R2;

  // basis functions matrices
  matrix[L1,p1] B1;
  matrix[L2,p2] B2;
  matrix[p1,p1] BTBinv1;    // B1'B1
  matrix[p2,p2] BTBinv2;    // B2'B2

  // hyperparameters
  real<lower=0> v;           // dofs latent factors (in the thesis nu)
  real<lower=0> nu;          // dofs beta (int the thesis chi)
  real<lower=0> alpha_psi;
```

```
  real<lower=0> beta_psi;
  real<lower=0> psi_noise_scale_multiplier_1;
  real<lower=0> psi_noise_scale_multiplier_2;
  real<lower=0> alpha_sigma;
  real<lower=0> beta_sigma;
  real<lower=0> theta_noise_scale_multiplier_1;
  real<lower=0> theta_noise_scale_multiplier_2;
  real<lower=0> alpha_1;
  real<lower=0> alpha_2;
}

transformed data {
  matrix[p1+p2,N] y_all;
  y_all[1:p1,:] = proj_y1;
  y_all[(p1+1):(p1+p2),:] = proj_y2;

  real alpha_psi_1_conj = alpha_psi + 0.5*N*(L1-p1);
  real beta_psi_1_conj = beta_psi*psi_noise_scale_multiplier_1 +
                           0.5*sum(square(ress_y1));
  real alpha_psi_2_conj = alpha_psi + 0.5*N*(L2-p2);
  real beta_psi_2_conj = beta_psi*psi_noise_scale_multiplier_2 +
                           0.5*sum(square(ress_y2));
}

parameters {
  matrix[p1,N] theta1;
  matrix[p2,N] theta2;
  matrix[p1,k] Lambda1;
  matrix[p2,k] Lambda2;
  matrix[k,N] eta;
  matrix[k,r] beta;
  vector<lower=0>[p1] tau_theta1;
  vector<lower=0>[p2] tau_theta2;
  vector<lower=0>[k] delta1;
  vector<lower=0>[k] delta2;
  real<lower=0> psi1;
  real<lower=0> psi2;
}
```

```
transformed parameters {
  vector[k] tau_lambda1;
  tau_lambda1[1] = delta1[1];
  for (h in 2:k) {
    tau_lambda1[h] = prod(delta1[1:h]);
  }
  vector[k] tau_lambda2;
  tau_lambda2[1] = delta2[1];
  for (h in 2:k) {
    tau_lambda2[h] = prod(delta2[1:h]);
  }
}

model {

  // lklhood
  target += normal_lupdf( to_vector(
                    qr_decompose_R1*(proj_y1 - theta1) )./(psi1) | 0, 1 );
  target += normal_lupdf( to_vector(
                    qr_decompose_R2*(proj_y2 - theta2) )./(psi2) | 0, 1 );
  to_vector(diag_pre_multiply(sqrt(tau_theta1), theta1 - Lambda1*eta))
            ~ normal(0,theta_noise_scale_multiplier_1);
  to_vector(diag_pre_multiply(sqrt(tau_theta2), theta2 - Lambda2*eta))
            ~ normal(0,theta_noise_scale_multiplier_2);
  to_vector(eta - beta*X) ~ normal(0,1);

  // priors for matrices
  to_vector(diag_post_multiply(Lambda1, sqrt(tau_lambda1)))
            ~ student_t(v, 0, 1);
  to_vector(diag_post_multiply(Lambda2, sqrt(tau_lambda2)))
            ~ student_t(v, 0, 1);
  to_vector(beta) ~ student_t(nu, 0,1);

  // priors
  square(psi1) ~ inv_gamma(alpha_psi_1_conj, beta_psi_1_conj);
  square(psi2) ~ inv_gamma(alpha_psi_2_conj, beta_psi_2_conj);
  tau_theta1 ~ gamma(alpha_sigma, beta_sigma);
  tau_theta2 ~ gamma(alpha_sigma, beta_sigma);
  delta1[1] ~ gamma(alpha_1, 1);
```

```
  delta1[2:k] ~ gamma(alpha_2, 1);
  delta2[1] ~ gamma(alpha_1, 1);
  delta2[2:k] ~ gamma(alpha_2, 1);

  // Jacobian Corrections
  target += log(psi1);
  target += log(psi2);
  target += N*0.5*sum(log(tau_theta1));
  target += p1*0.5*sum(log(tau_lambda1));
  target += N*0.5*sum(log(tau_theta2));
  target += p2*0.5*sum(log(tau_lambda2));
}

generated quantities {
  array[N] real log_lik_y;
  {
    matrix[p1,r] regr_coeffs1 = Lambda1*beta;
    matrix[p1,N] estimated_y1 = regr_coeffs1*X;
    matrix[p2,r] regr_coeffs2 = Lambda2*beta;
    matrix[p2,N] estimated_y2 = regr_coeffs2*X;

    matrix[p1,p1] Sigma_11 = add_diag( Lambda1*(Lambda1') ,
     square(theta_noise_scale_multiplier_1)./tau_theta1 ) + square(psi1)*BTBinv1;
    matrix[p2,p2] Sigma_22 = add_diag( Lambda2*(Lambda2') ,
     square(theta_noise_scale_multiplier_2)./tau_theta2 ) + square(psi2)*BTBinv2;
    matrix[p1,p2] Sigma_12 = Lambda1*(Lambda2');
    matrix[p2,p1] Sigma_21 = Lambda2*(Lambda1');
    matrix[p1+p2,p1+p2] Sigma;
    Sigma[1:p1,1:p1] = Sigma_11;
    Sigma[1:p1,(p1+1):(p1+p2)] = Sigma_12;
    Sigma[(p1+1):(p1+p2),1:p1] = Sigma_21;
    Sigma[(p1+1):(p1+p2),(p1+1):(p1+p2)] = Sigma_22;

    matrix[p1+p2,p1+p2] Choleksy_Decomposed_Sigma = cholesky_decompose(Sigma);

    for (i in 1:N) {
      vector[p1+p2] estimated_y_aux;
      estimated_y_aux[1:p1] = estimated_y1[:,i];
      estimated_y_aux[(p1+1):(p1+p2)] = estimated_y2[:,i];
```

```
      log_lik_y[i] = multi_normal_cholesky_lpdf(
        y_all[:,i] | estimated_y_aux, Choleksy_Decomposed_Sigma
      );
      log_lik_y[i] += -0.5*square(ress_y1[i])/square(psi1);
      log_lik_y[i] += -0.5*(L1-p1)*log(2*pi());
      log_lik_y[i] += -(L1-p1)*log(psi1);
      log_lik_y[i] += -0.5*square(ress_y2[i])/square(psi2);
      log_lik_y[i] += -0.5*(L2-p2)*log(2*pi());
      log_lik_y[i] += -(L2-p2)*log(psi2);
    }
  }
}
```

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisors, Ilenia Epifani, Francesco Trovò, Beatrice Cantoni and Manuela Antonelli, for their unwavering support, guidance, and expertise throughout the duration of this research. Their insights and feedback were invaluable in shaping this thesis into its final form.

I would also like to thank Mario Beraha for the great help he gave me shaping this thesis work.

My heartfelt appreciation goes out to my family, especially mum, dad, and my dear brother, for their love, encouragement, and the countless ways they supported me throughout this journey. Without their patience and understanding, this accomplishment would not have been possible.

I am profoundly grateful to my best friend, Cinzia, for the myriad moments of joy and sorrow we've experienced side by side through the years.

My sincere thanks goes to the professors who have significantly influenced my academic journey. Special recognition goes to Prof.ssa Alessandra Guglielmi for unveiling the fascinating world of Bayesian Statistics to me. I am deeply indebted to Prof. Maurizio Grasselli, whose expertise laid the foundations for my understanding of mathematical principles. I owe a debt of gratitude to Prof. Mattia Frasca and Prof.ssa Valentina Gambuzza, who ignited my passion for Complex Systems and continuously supported and encouraged me throughout my studies. Lastly, I am immensely grateful to Prof.ssa Vittoria Marzullo, who taught me that the essence of teaching lies in a profound love for one's subject.

Lastly, I would like to thank all the friends I made at Politecnico, especially those who accompanied in developing the various course projects. Like J.K.Rowling wrote "There are some things you can't share without ending up liking each other, and knocking out a twelve-foot mountain troll is one of them."