

Functional Mixture Regression Control Chart

Christian Capezza¹, Fabio Centofanti¹, Davide Forcina¹, Antonio Lepore¹,
and Biagio Palumbo^{*1}

¹*Department of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio 80,
80125, Naples, Italy*

Abstract

Industrial applications often exhibit multiple in-control patterns due to varying operating conditions, which makes a single functional linear model (FLM) inadequate to capture the complexity of the true relationship between a functional quality characteristic and covariates, which gives rise to the multimode profile monitoring problem. This issue is clearly illustrated in the resistance spot welding (RSW) process in the automotive industry, where different operating conditions lead to multiple in-control states. In these states, factors such as electrode tip wear and dressing may influence the functional quality characteristic differently, resulting in distinct FLMS across subpopulations. To address this problem, this article introduces the functional mixture regression control chart (FMRCC) to monitor functional quality characteristics with multiple in-control patterns and covariate information, modeled using a mixture of FLMS. A monitoring strategy based on the likelihood ratio test is proposed to monitor any deviation from the estimated in-control heterogeneous population. An extensive Monte Carlo simulation study is performed to compare the FMRCC with competing monitoring schemes that have already appeared in the literature, and a case study in the monitoring of an RSW process in the automotive industry, which motivated this research, illustrates its practical applicability.

Keywords: Functional Data Analysis, Profile Monitoring, Statistical Process Control, Functional Mixture Regression, Multiple Functional Linear Models

1 Introduction

In modern statistical process monitoring (SPM) applications, experimental measurements of the quality characteristic of interest are often characterized by complex and high-dimensional formats that can best be represented by functional data or profiles (Ramsay

*Corresponding author. e-mail: biagio.palumbo@unina.it

and Silverman, 2005; Kokoszka and Reimherr, 2017) and stimulated growing interest in profile monitoring (Noorossana et al., 2011; Zou et al., 2012; Chou et al., 2014; Grasso et al., 2014; Paynabar et al., 2016; Wang et al., 2018b; Ren et al., 2019; Capezza et al., 2024b, 2023). Other relevant contributions include those by Jin and Shi (1999), Colosimo and Pacella (2010), Grasso et al. (2016), Grasso et al. (2017), Menafoglio et al. (2018), Maleki et al. (2018), and Jones et al. (2021).

Profile monitoring, and in general SPM, is commonly implemented in two phases. The first (Phase I) aims to identify a clean dataset to be assumed to be representative of the in-control (IC) state of the process, hereinafter referred to as Phase I or reference sample, through retrospective monitoring of an initial dataset drawn from the process; the second (Phase II) is concerned with prospective monitoring of new observations (Qiu, 2013), hereinafter referred to as Phase II observations or Phase II dataset.

Traditional profile monitoring methods assume that the IC process is *unimodal*, i.e., it exhibits a single IC state. However, real industrial processes are often *multimodal*, i.e., they operate under different unobserved IC states, and give rise to the so-called *multimode profile monitoring* problem. To face this problem, Grasso et al. (2017) propose a method based on curve classification to determine the mode of the data on top of a functional control charting scheme. Park and Shrivastava (2014) presented a procedure to monitor time-correlated multimode processes using a mixture of time-series models in a Bayesian framework. Wang et al. (2018a) developed a feature extraction approach to group and monitor near-circular shape profiles through a likelihood ratio test-based EWMA control chart.

However, all works on multimode profile monitoring that appeared so far were not designed to incorporate any information from additional concurrent variables, hereinafter referred to also as predictors or covariates, influencing the functional quality characteristic to be monitored, hereinafter referred to also as the response. When available, incorporating covariate information may drastically improve any SPM scheme's performance.

It is well known indeed that, if a covariate manifests itself with extreme realization, an SPM scheme based on the observations of the quality characteristic alone may wrongly judge the process out of control (OC) or, dually, there are situations where the covariates

are not extreme, and the process may incorrectly appear IC.

In the profile monitoring framework, a basic solution was given by Centofanti et al. (2021), who translated the regression control chart (Mandel, 1969) into a functional regression control chart (FRCC) framework and implemented it under the assumption that a functional response is influenced by multiple functional covariates through a functional linear model (FLM). The residuals of the model were then monitored by simultaneously applying Hotelling’s T^2 and squared prediction error (SPE) control charts, as in Woodall et al. (2004). A software implementation of the FRCC is openly available in the R package `funcharts` (Capezza et al., 2023). However, while the FRCC is presented as a broader framework, the implementation presented in the original work relies on the assumption that all subjects sampled from the process obey the same FLM, and thus cannot apply in the multimode profile monitoring problem, where subjects belong to more than one sub-population, in which the FLMs describing the influence of the covariates on the response are different.

In fact, this is the case for the resistance spot welding (RSW) process in the automotive industry that motivated this research. From a technological perspective, the quality of the RSW process, as also noted by Capezza et al. (2021), can be monitored through observations of the dynamic resistance curve (DRC), which is recognized as the quality footprint of the metallurgical development of a spot weld (Dickinson et al., 1980) and represents the functional quality characteristic of interest, also called the response. Details on DRC behavior can be found in Section 4.

The RSW dataset consists of 1802 IC observations, measured in [$m\Omega$], referring to the same spot weld location on different car bodies and 37 OC DRC observations corresponding to known defective spot welds. For illustrative purposes, the left panel of Figure 1 displays the IC observations colored by the wear of the electrode tips that performed the spot weld, measured through the number of welds before the electrode tip dressing, whereas, on the right panel, DRCs are colored by the number of electrode tip dressings itself. These panels clearly show that both the electrode tip wear and the dressing represent informative scalar covariates, which may explain the heterogeneous DRC shapes depicted in Figure 1 while not necessarily influencing the final quality of the welded joint. In addition to these two

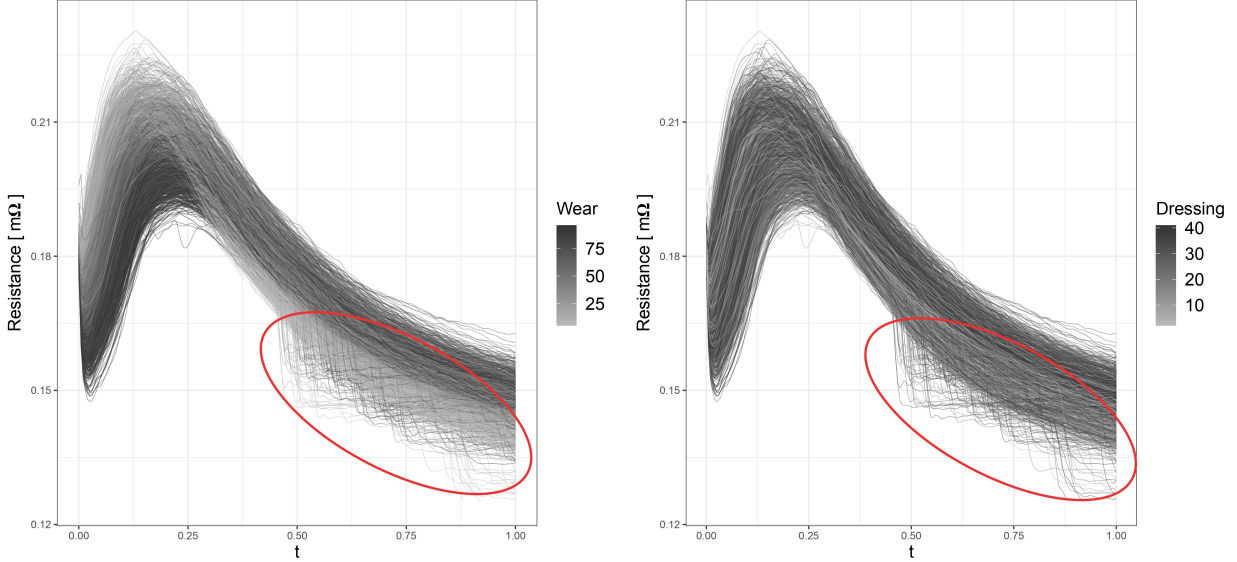


Figure 1: Wear effect (left panel) and dressing effect (right panel) on the DRCs for point 20268.

covariates, many other phenomena can cause DRC heterogeneity across different subpopulations. One known phenomenon is the expulsion of molten material caused by an abnormal welding current flowing through the electrode tips due to an abnormal clamping force that causes a significant drop in the DRCs and may cause the DRC behavior highlighted by the balloons in the left and right panels of Figure 1.

Mixtures of regression models are undoubtedly useful for modeling such heterogeneity, as pointed out by (DeSarbo and Cron, 1988; Jones and McLachlan, 1992) in the case of scalar response versus scalar predictor variables. Extensions to the functional case are presented by Yao et al. (2011); Ciarleglio and Ogden (2016); Zhao et al. (2012) in the scalar-on-function regression setting. In particular, Yao et al. (2011) proposed a mixture of functional regression models which relate a univariate scalar response with a functional predictor variable by adopting the eigenbasis representation of the latter, reducing to a framework similar to the classic mixture of regression models. Wang et al. (2016) studied the mixture of functional regression models in the case of a univariate functional response and multivariate functional covariates through concurrent FLMS. Finally, Devijver (2017) extended this case with one functional predictor variable to non-concurrent FLMS, also referred to as total FLMS (Ramsay and Silverman, 2005), where the value of a univariate functional response is related to the entire domain of a functional predictor variable. It is

important to note that all functional mixture regression (FMR) models are more flexible in modeling unobserved heterogeneity than existing functional clustering methods. The latter methods focus only on clustering the functions themselves, while the former focuses on detecting clusters characterized by different regression structures.

In this article, a functional mixture regression control chart (FMRCC) is proposed to deal with the multimodal profile monitoring problem through an FMR approach with functional response and multivariate functional or scalar predictors instead of modeling different modes only through functional clustering (see, e.g., Capezza et al. (2021); Centofanti et al. (2024)). To deal with the infinite dimensionality of the data, both the response and predictors are projected onto an appropriate basis function system. This allows for the application of existing methods based on the finite mixture regression model and the expectation-maximization (EM) algorithm on the basis coefficients. Finally, an SPM strategy based on the likelihood ratio test (LRT) is proposed to monitor any deviation from the estimated IC heterogeneous population. A Monte Carlo simulation study is performed to assess the SPM performance of the FMRCC scheme. The practical applicability of the FMRCC is demonstrated through a case study in the monitoring of the aforementioned RSW process.

The remainder of the article is organized as follows. Section 2 provides a complete description of the FMRCC. In Section 3, the performance of the FMRCC is compared to that of the other three control charts by means of a simulation study. In Section 4, the case study is presented. Section 5 concludes the article. The Supplementary Materials contains additional details on the data generation process in the simulation study. All calculations and plots were obtained using the programming language R (R Core Team, 2021).

2 The Functional Mixture Regression Control Chart Framework

2.1 Preliminaries

The proposed FMRC is a general framework for multimode profile monitoring where different modes are characterized by different FLMS, with a univariate functional response. Assume that N observations of random functions $\tilde{X}_1, \dots, \tilde{X}_p$ and \tilde{Y} are available with \tilde{X}_j , $j = 1, \dots, p$, representing predictors having values in $L^2(\mathcal{S})$, and \tilde{Y} representing the response having values in $L^2(\mathcal{T})$. $L^2(\mathcal{D})$ denotes the Hilbert space of square-integrable functions defined in the compact set $\mathcal{D} \subseteq \mathbb{R}$, with the inner product $\langle f, g \rangle = \int_{\mathcal{S}} f(s)g(s)ds$. Let $\mathbb{H}^{\tilde{X}} = (L^2(\mathcal{S}))^p$ denote the Hilbert space whose elements are vectors of functions in $L^2(\mathcal{S})$. Then $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_p)^T$ is a random vector of functions, whose realizations are in $\mathbb{H}^{\tilde{X}}$. Assume that $\tilde{\mathbf{X}}$ has a smooth mean function $\boldsymbol{\mu}^{\tilde{X}} = (\mu_1^{\tilde{X}}, \dots, \mu_p^{\tilde{X}})^T$, with $\mu_j^{\tilde{X}} = E(\tilde{X}_j)$ and a covariance function $\mathbf{C}^{\tilde{X}} = \{C_{i,j}^{\tilde{X}}\}_{1 \leq i,j \leq p}$, with $C_{i,j}^{\tilde{X}}(s_1, s_2) = \text{Cov}(\tilde{X}_i(s_1), \tilde{X}_j(s_2))$, for $s_1, s_2 \in \mathcal{S}$. Analogously, let $\mu^{\tilde{Y}} = E(\tilde{Y})$ and $C^{\tilde{Y}}(t_1, t_2) = \text{Cov}(\tilde{Y}(t_1), \tilde{Y}(t_2))$, for $t_1, t_2 \in \mathcal{T}$, be the mean and covariance functions of the response variable, respectively.

In particular, to avoid exhibiting incomparable magnitudes of variation, the transformation approach of Chiou et al. (2014) is utilized hereinafter. To this end, let $\mathbf{X} = (X_1, \dots, X_p)^T = (\mathbf{V}^{\tilde{X}})^{-1}(\tilde{\mathbf{X}} - \boldsymbol{\mu}^{\tilde{X}})$ be the standardized covariate with $\mathbf{V}^{\tilde{X}} = \text{diag}((v_1^{\tilde{X}})^{1/2}, \dots, (v_p^{\tilde{X}})^{1/2})$ and the square root of $v_j^{\tilde{X}}(s) = C_{j,j}^{\tilde{X}}(s, s)$, $\forall s \in \mathcal{S}$. Consequently, the standardized response is defined as $Y = (v^{\tilde{Y}})^{-1/2}(\tilde{Y} - \mu^{\tilde{Y}})$ with $v^{\tilde{Y}}(t) = C^{\tilde{Y}}(t, t)$, $\forall t \in \mathcal{T}$. Then, let \mathbf{C}^X and C^Y denote the covariance functions of the standardized covariates and the response. There exists a multivariate orthonormal eigenbasis $\{\boldsymbol{\psi}_l^X(s)\}_{l=1,2,\dots}$, such that $\langle \boldsymbol{\psi}_l^X, \boldsymbol{\psi}_h^X \rangle_{\mathbb{H}^X} = \delta_{lh}$, with δ_{lh} the Kronecker delta. The corresponding non-negative eigenvalues $\{\lambda_l^X\}_{l=1,2,\dots}$ are nonnegative and are supposed, without loss of generality, to be nonincreasing with l and such that $\mathbf{C}^X(s_1, s_2) = \sum_{l=1}^{\infty} \lambda_l^X \boldsymbol{\psi}_l^X(s_1) \boldsymbol{\psi}_l^X(s_2)^T$, for $s_1, s_2 \in \mathcal{S}$. Similarly, there exists an orthonormal eigenbasis $\{\psi_m^Y(t)\}_{m=1,2,\dots}$ such that $C^Y(t_1, t_2) = \sum_{m=1}^{\infty} \lambda_m^Y \psi_m^Y(t_1) \psi_m^Y(t_2)^T$, for $t_1, t_2 \in \mathcal{T}$ where $\{\lambda_m^Y\}_{m=1,2,\dots}$ are the corresponding nondecreasing eigenvalues such that $\langle \psi_l^Y, \psi_h^Y \rangle_{\mathbb{H}^Y} = \delta_{lh}$.

2.2 Model specification

Suppose N observations are spread among an unknown number K of mutually exclusive clusters, each of which is characterized by a different FLM describing the influence of the covariates on the response, with the probability that each observation belonging to the k -th cluster is π_k , with $\sum_{k=1}^K \pi_k = 1$. Moreover, let us denote by $\mathbf{Z} = (Z_1, \dots, Z_K)^T$, $k = 1, \dots, K$, the unknown component-label vector corresponding to $(\tilde{\mathbf{X}}, \tilde{Y})$, where Z_k equals 1 if the observation is in the k -th cluster and 0 otherwise. Conditioned to $Z_k = 1$, that is, to the k -th cluster, the functional linear model to be estimated is

$$Y(t) = \beta_{0k}(t) + \int_{\mathcal{S}} (\boldsymbol{\beta}_k(s, t))^T \mathbf{X}(s) ds + \varepsilon(t), \quad t \in \mathcal{T}, k = 1, \dots, K, \quad (1)$$

where $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{kp})$ is k -th regression coefficient vector, whose elements are bivariate functions in the space of square-integrable functions defined on the closed interval $\mathcal{S} \times \mathcal{T}$, β_{0k} and ε are k -th functional intercept and the functional error term, respectively, both defined on the compact domain \mathcal{T} . The functional error term ε is supposed independent of \mathbf{X} and having $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = \nu_\varepsilon^2$. Thus, the regression function for the k -th cluster is

$$E(Y(t)|\mathbf{X}) = \beta_{0k}(t) + \int_{\mathcal{S}} (\boldsymbol{\beta}_k(s, t))^T \mathbf{X}(s) ds, \quad t \in \mathcal{T}, k = 1, \dots, K. \quad (2)$$

From the multivariate functional principal component (MFPC) decomposition (Chiou et al., 2014; Happ and Greven, 2018), the predictors and the response can be represented as

$$\mathbf{X}(s) = \sum_{l=1}^{\infty} \xi_l^X \boldsymbol{\psi}_l^X(s), \quad Y(t) = \sum_{m=1}^{\infty} \xi_m^Y \psi_m^Y(t), \quad (3)$$

where $\{\boldsymbol{\psi}_l^X(s)\}$ and $\{\psi_m^Y(t)\}$ are the eigenfunctions defined in section 2.1 and $\xi_l^X = \langle \mathbf{X}, \boldsymbol{\psi}_l^X \rangle_{\mathbb{H}^X}$ and $\xi_m^Y = \langle Y, \psi_m^Y \rangle$ are the functional principal component scores which, unconditionally, are uncorrelated and satisfy $E(\xi_l^X) = 0$, $E(\xi_l^X \xi_h^X) = \lambda_l^X \delta_{lh}$ and $E(\xi_m^Y) = 0$, $E(\xi_m^Y \xi_h^Y) = \lambda_m^Y \delta_{mh}$. Since the eigenfunctions of a square-integrable random function form a complete orthonormal basis, the regression coefficient vector and the functional intercept can be expressed as

$$\boldsymbol{\beta}_k(s, t) = \sum_{l,m=1}^{\infty} b_{lmk} \boldsymbol{\psi}_l^X(s) \boldsymbol{\psi}_m^Y(t), \quad \beta_{0k}(t) = \sum_{m=1}^{\infty} b_{0mk} \boldsymbol{\psi}_m^Y(t), \quad s \in \mathcal{S}, t \in \mathcal{T}, k = 1, \dots, K, \quad (4)$$

respectively, where b_{lmk} and b_{0k} are the coefficients of the basis expansions. Furthermore, considering the following expansion of the functional error term ε ,

$$\varepsilon(t) = \sum_{m=1}^{\infty} \varepsilon_m \boldsymbol{\psi}_m^Y(t), \quad (5)$$

with $\varepsilon_m = \langle \varepsilon, \boldsymbol{\psi}_m^Y \rangle$, and by plugging Equations (3), (5) and (4) in (1), we obtain

$$Y(t) = \sum_{m=1}^{\infty} b_{0mk} \boldsymbol{\psi}_m^Y(t) + \int_{\mathcal{S}} \left(\sum_{l,m=1}^{\infty} b_{lmk} \boldsymbol{\psi}_l^X(s) \boldsymbol{\psi}_m^Y(t) \right)^T \sum_{l=1}^{\infty} \xi_l^X \boldsymbol{\psi}_l^X(s) ds + \sum_{m=1}^{\infty} \varepsilon_m \boldsymbol{\psi}_m^Y(t), \quad t \in \mathcal{T}, k = 1, \dots, K. \quad (6)$$

In the context of multivariate functional regression, the estimation of model parameters in Equation (6) requires regularization of the predictor and response functions, which is achieved by considering the truncated principal component decomposition \mathbf{X}_L and Y_M of \mathbf{X} and Y . That is,

$$\mathbf{X}_L(s) = \sum_{l=1}^L \xi_l^X \boldsymbol{\psi}_l^X(s), \quad Y_M(t) = \sum_{m=1}^M \xi_m^Y \boldsymbol{\psi}_m^Y(t), \quad (7)$$

where L and M denote the minimum number of components necessary to explain a given fraction of the total variation in the data. Other selection criteria, e.g., based on Akaike information criterion (AIC) or Bayesian information criterion (BIC), could also be used. Accordingly, we get the truncated version of Equation (4) as follows

$$\boldsymbol{\beta}_k(s, t) = \sum_{l=1}^L \sum_{m=1}^M b_{lmk} \boldsymbol{\psi}_l^X(s) \boldsymbol{\psi}_m^Y(t), \quad \beta_{0k}(t) = \sum_{m=1}^M b_{0mk} \boldsymbol{\psi}_m^Y(t), \quad s \in \mathcal{S}, t \in \mathcal{T}, k = 1, \dots, K. \quad (8)$$

Then, using $Y_M(t)$ in Equation (7) instead of $Y(t)$ and by the orthonormality of $\{\psi_l^X(s)\}$ and $\{\psi_m^Y(t)\}$, Equation (1) is reduced to the truncated multivariate linear regression model

$$\boldsymbol{\xi}_M^Y = \mathbf{b}_{0k} + (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_L^X + \boldsymbol{\varepsilon}_M, \quad k = 1, \dots, K, \quad (9)$$

where $\boldsymbol{\xi}_M^Y = (\xi_1^Y, \dots, \xi_M^Y)^T$, $\boldsymbol{\xi}_L^X = (\xi_1^X, \dots, \xi_L^X)^T$, $\boldsymbol{\varepsilon}_M$ is the truncated version of the basis coefficients of ε in Equation (5), that is, multivariate Gaussian random errors with zero mean and covariance $\boldsymbol{\Sigma}_k$, $\mathbf{B}_{LMk} = \{b_{lmk}\}_{l=1, \dots, L, m=1, \dots, M}$ and $\mathbf{b}_{0k} = (b_{01k}, \dots, b_{0Mk})$. Additionally, Equation (9) can be reduced to

$$\boldsymbol{\xi}_M^Y = (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_L^X + \boldsymbol{\varepsilon}_M, \quad k = 1, \dots, K, \quad (10)$$

by incorporating the intercept term \mathbf{b}_{0k} in \mathbf{B}_{LMk} and by adding 1 to the head of $\boldsymbol{\xi}_L^X$. Hence, given N independent realizations (\mathbf{X}_i, Y_i) of (\mathbf{X}, Y) , $i = 1, \dots, N$, the probability density function (pdf) of $\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X$ comes from the following mixture:

$$f(\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X) = \sum_{k=1}^K \pi_k \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_{i,L}^X, \boldsymbol{\Sigma}_k), \quad (11)$$

where $\phi(\cdot, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Equation (11) is the classical K -component Gaussian mixture regression model (McLachlan and Peel, 2004). Therefore, the log-likelihood function corresponding to $(\boldsymbol{\xi}_{1,M}^Y, \boldsymbol{\xi}_{1,L}^X), \dots, (\boldsymbol{\xi}_{N,M}^Y, \boldsymbol{\xi}_{N,L}^X)$ is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{i=1}^N \log \sum_{k=1}^K \pi_k \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_{i,L}^X, \boldsymbol{\Sigma}_k), \quad (12)$$

where $\boldsymbol{\Psi} = (\pi_k, \mathbf{B}_{LMk}, \boldsymbol{\Sigma}_k)_{k=1, \dots, K}$ is the parameter set to be estimated. In the presence of a large number of components K , having a different matrix parameter $\boldsymbol{\Sigma}_k$ in each component may lead to too many parameters and then to overfitting. Therefore, we also consider alternative and more parsimonious parameterizations of $\boldsymbol{\Sigma}_k$. Then, the problem of model selection is discussed at the end of the next Section 2.3. In particular, in this article we consider the following parametrizations, i.e., the spherical family with one common scalar parameter λ and diagonal matrix $\boldsymbol{\Sigma}_k = \lambda \mathbf{I}$ for each $k = 1, \dots, K$; the spherical

family with K scalar parameters $\lambda_1, \dots, \lambda_K$, where $\Sigma_k = \lambda_k \mathbf{I}$; the parameterization with common covariance $\Sigma_k = \Sigma$, not necessarily diagonal; the full parametrization where each component $k = 1, \dots, K$ has its own covariance Σ_k .

2.3 The estimation method

An estimator $\hat{\Psi}$ of Ψ in Equation (12) can be calculated by maximizing the log-likelihood in Equation (12). Unfortunately, maximization should be performed numerically through dedicated algorithms, such as the expectation-maximization (EM) algorithm (Dempster et al., 1977; McLachlan and Krishnan, 2007). The EM algorithm first requires the construction of the complete-data log-likelihood by elaborating Equation (12) as follows

$$\log L_c(\Psi) = \sum_{i=1}^N \sum_{k=1}^K Z_{ki} \log[\pi_k \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_{i,L}^X, \Sigma_k)]. \quad (13)$$

Then, starting from an initial solution $\Psi^{(0)}$, the version of the EM algorithm for a mixture regression model (Jones and McLachlan, 1992) such as that introduced in Equation (11) alternates between the expectation and maximization steps, referred to as the E step and M step, respectively, until there is no appreciable change in the logarithmic likelihood values. At iteration q , the E-step computes the conditional expectation of the complete-data log-likelihood function in Equation (13), using the current parameter vector $\Psi^{(q)}$; that is

$$Q(\Psi, \Psi^{(q)}) = \sum_{i=1}^N \sum_{k=1}^K \tau_{ki}^{(q)} \log[\pi_k \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMk})^T \boldsymbol{\xi}_{i,L}^X, \Sigma_k)]. \quad (14)$$

The E-step only requires computing the posterior probabilities of component membership $\tau_{ki}^{(q)}$, $i = 1, \dots, N$, for each of the K components, namely

$$\tau_{ki}^{(q)} = \frac{\pi_k^{(q)} \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMk}^{(q)})^T \boldsymbol{\xi}_{i,L}^X, \Sigma_k^{(q)})}{\sum_{h=1}^K \pi_h^{(q)} \phi(\boldsymbol{\xi}_{i,M}^Y; (\mathbf{B}_{LMh}^{(q)})^T \boldsymbol{\xi}_{i,L}^X, \Sigma_h^{(q)})}. \quad (15)$$

The M-step updates the value of the parameter vector Ψ by maximizing Equation (14) with respect to Ψ . The mixing proportions updates are

$$\pi_k^{(q+1)} = \frac{1}{N} \sum_{i=1}^N \tau_{ki}^{(q)}, \quad (16)$$

while the regression parameters \mathbf{B}_{LMk} and Σ_k are obtained by maximizing the complete-data log-likelihood function in Equation (14) with respect to $(\mathbf{B}_{LMk}, \Sigma_k)$. This corresponds to solving K least-squares problems with the posterior probabilities $(\tau_{ki}^{(q)})_{k=1, \dots, K}$ as weights, which give

$$\mathbf{B}_{LMk}^{(q+1)} = \left[\sum_{i=1}^N \tau_{ki}^{(q)} (\boldsymbol{\xi}_{i,L}^X)^T \boldsymbol{\xi}_{i,L}^X \right]^{-1} \sum_{i=1}^N \tau_{ki}^{(q)} \boldsymbol{\xi}_{i,L}^X (\boldsymbol{\xi}_{i,M}^Y)^T, \quad (17)$$

$$\Sigma_k^{(q+1)} = \frac{1}{\sum_{i=1}^N \tau_{ki}^{(q)}} \sum_{i=1}^N \tau_{ki}^{(q)} \left[\boldsymbol{\xi}_{i,M}^Y - (\mathbf{B}_{LMk}^{(q+1)})^T \boldsymbol{\xi}_{i,L}^X \right] \left[\boldsymbol{\xi}_{i,M}^Y - (\mathbf{B}_{LMk}^{(q+1)})^T \boldsymbol{\xi}_{i,L}^X \right]^T. \quad (18)$$

Dempster et al. (1977) showed that at each iteration of the EM algorithm, the log-likelihood in Equation (12) is non-decreasing. Additionally, the series of parameter estimates converges towards a local maximum of the log-likelihood function (Wu, 1983).

As suggested in McLachlan and Peel (2004) and implemented in popular software packages `EMMIXskew` (Wang et al., 2009) and `mclust` (Scrucca et al., 2016), the K-means algorithm can be utilized to obtain an initial solution $\Psi^{(0)}$. The BIC is used for the parameterization of the covariance matrix Σ_k as well as for the choice of the number of clusters K , when the number of components is not known prior to or evident by data exploration. BIC is one of the most widely adopted methods and has provided good results in various applications of model-based clustering (Fraley and Raftery, 2002).

2.4 The SPM strategy

Let us assume that a set of functional principal component scores $(\boldsymbol{\xi}_{i,L}^X, \boldsymbol{\xi}_{i,M}^Y)$ corresponding to the functional observations $(\tilde{\mathbf{X}}_i, \tilde{Y}_i)$, $i = 1, \dots, N$, is available from Phase I, and that it is drawn from an IC heterogeneous population as defined by Jones and McLachlan (1992) and denoted by Π . Then, let us suppose that we are interested in testing whether the scores $(\boldsymbol{\xi}_{N+1,L}^X, \boldsymbol{\xi}_{N+1,M}^Y)$ of a new functional observation $(\tilde{\mathbf{X}}_{N+1}, \tilde{Y}_{N+1})$ belong to Π or not.

That is, we are interested in testing the following null hypothesis H_0 (IC process) versus the alternative hypothesis H_1 (OC process)

$$H_0 : \boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X \in \boldsymbol{\Pi}, \quad H_1 : \boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X \notin \boldsymbol{\Pi}. \quad (19)$$

The test statistic can be constructed using a likelihood ratio test (LRT), as the ratio between the likelihood functions maximized under H_0 and H_1 , and denoted by $L_0(\boldsymbol{\Psi})$ and $L_1(\boldsymbol{\Psi}, \boldsymbol{\beta})$, respectively. The former can be written as

$$L_0(\boldsymbol{\Psi}) = \left(\prod_{i=1}^N f(\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X; \boldsymbol{\Psi}) \right) f(\boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X; \boldsymbol{\Psi}), \quad (20)$$

where $f(\boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X; \boldsymbol{\Psi})$ is the IC pdf defined as in Equation (11) evaluated at a new observation, whereas the latter is similarly expressed as

$$L_1(\boldsymbol{\Psi}, \boldsymbol{\beta}) = \left(\prod_{i=1}^N f(\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X; \boldsymbol{\Psi}) \right) h(\boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X; \boldsymbol{\beta}), \quad (21)$$

where $h(\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X; \boldsymbol{\beta})$ denotes the pdf with parameter vector $\boldsymbol{\beta}$ from which the new observation is sampled. However, since there is only one observation under H_1 , according to Wang et al. (1997), a simple constant density $h(\boldsymbol{\xi}_{N+1,M}^Y | \boldsymbol{\xi}_{N+1,L}^X) \equiv c$ should be used and dropped from the LRT statistic in these cases. Therefore, following the principle of equal ignorance, the outlier distribution is not explicitly required (Sain et al., 1999). By denoting the likelihood based on the Phase I sample with

$$\tilde{L}_1(\boldsymbol{\Psi}) = \prod_{i=1}^N f(\boldsymbol{\xi}_{i,M}^Y | \boldsymbol{\xi}_{i,L}^X; \boldsymbol{\Psi}), \quad (22)$$

the LRT statistic can be written as

$$\Lambda = \frac{\sup_{\boldsymbol{\Psi} \in \Omega} L_0(\boldsymbol{\Psi})}{\sup_{\boldsymbol{\Psi} \in \Omega} \tilde{L}_1(\boldsymbol{\Psi})}, \quad (23)$$

where Ω is the complete parameter space. In particular, the denominator of Equation (23) is computed only once by plugging in the estimate $\hat{\Psi}$ of the parameter set Ψ calculated, as described in Section 2.3, with the Phase I sample, whereas the numerator $\sup_{\Psi \in \Omega} L_0(\Psi)$ is computed at each new observation, using $\hat{\Psi}$ as an initial value. Wang et al. (1997) demonstrated that, as $n \rightarrow \infty$, Λ approaches $f(\xi_{N+1,M}^Y | \xi_{N+1,L}^X; \hat{\Psi})$. That is, for large n , due to the minimal conditioning effect that a new observation has on the maximizer of Equation (20), there is no significant change in the parameter estimate. This asymptotic approximation, which is used throughout this article, leads to the simplification of Equation (23), which, for large n , can be rewritten as

$$\Lambda = f(\xi_{N+1,M}^Y | \xi_{N+1,L}^X; \hat{\Psi}). \quad (24)$$

For obvious reasons, the monitoring statistic $W = -\log \Lambda$ can be used more practically in place of (24).

Unlike Sain et al. (1999) and Wang et al. (1997), who calculate the distribution of the test statistic under H_0 using a bootstrap procedure, here we assume that the Phase I sample is suitably partitioned into a *training* and *tuning* datasets as also done in Colosimo and Pacella (2010), Capezza et al. (2024b) and Centofanti et al. (2021), to reduce the undesirable effect caused by uncertainty in the estimation of the MFPCA model (Ramaker et al., 2004; Kruger and Xie, 2012). The training set is used to estimate the model parameters, whereas the tuning set is used to estimate the control chart limit for the monitoring statistic calculated based on the estimated model. The upper control limit W_α^{lim} is obtained non-parametrically by means of the $(1 - \alpha)$ -quantile of the empirical distribution of W , where α is the desired type-I error rate.

In Phase II monitoring, to test hypotheses in Equation (19) based on the current functional observation, denoted by $(\tilde{\mathbf{X}}^*, \tilde{Y}^*)$, the functional principal component scores $\hat{\xi}_l^{X^*} = \langle \mathbf{X}^*, \hat{\psi}_l^X \rangle_{l=1, \dots, L}$ and $\hat{\xi}_m^{Y^*} = \langle Y^*, \hat{\psi}_m^Y \rangle_{m=1, \dots, M}$ are obtained from $\mathbf{X}^* = (\hat{\mathbf{V}}^{\tilde{X}})^{-1}(\tilde{\mathbf{X}}^* - \hat{\boldsymbol{\mu}}^{\tilde{X}})$ and $Y^* = (\hat{v}^{\tilde{Y}})^{-1/2}(\tilde{Y}^* - \hat{\mu}^{\tilde{Y}})$. Then, the corresponding realization of the monitoring statistic denoted by \hat{W}^* is calculated as

$$\hat{W}^* = -\log f(\hat{\xi}_M^{Y^*} | \hat{\xi}_L^{X^*}; \hat{\Psi}). \quad (25)$$

If $\hat{W}^* > W_\alpha^{lim}$, we reject H_0 , and an alarm is issued to signal a possible OC state of the process.

2.4.1 The studentized FMRCC

To reduce the influence of covariate mean shifts on the residual mean, especially when the Phase I sample size is small, as in Centofanti et al. (2021), we propose a *studentized* version of the residuals and thus of the monitoring statistic

$$\hat{W}^* = -\log \sum_{k=1}^K \hat{\pi}_k \phi \left(\hat{\boldsymbol{\xi}}_M^{Y^*}; (\hat{\mathbf{B}}_{LMk})^T \hat{\boldsymbol{\xi}}_L^{X^*}, \hat{\boldsymbol{\Sigma}}_k^* \right) \quad (26)$$

to be used in place of Equation (25). In Equation (26), the elements of $\hat{\boldsymbol{\Sigma}}_k^*$, which is the covariance matrix of the prediction error of $(\tilde{\mathbf{X}}^*, \tilde{Y}^*)$, are calculated as

$$\hat{\sigma}_{rhk}^* = \left(\hat{\sigma}_{rhk} + (\hat{\boldsymbol{\xi}}_L^{X^*})^T \text{Cov}\{\hat{\mathbf{B}}_{LM(r)k}, \hat{\mathbf{B}}_{LM(h)k}\} \hat{\boldsymbol{\xi}}_L^{X^*} \right)_{r,h=1,\dots,M}, \quad (27)$$

where $\hat{\sigma}_{rhk}$ is the element corresponding to the r -th row and the h -th column of $\hat{\boldsymbol{\Sigma}}_k$, $\hat{\mathbf{B}}_{LM(r)k}$ and $\hat{\mathbf{B}}_{LM(h)k}$ are the r -th and h -th row of $\hat{\mathbf{B}}_{LMk}$, respectively. Recalling the EM algorithm, $\text{Cov}\{\hat{\mathbf{B}}_{LM(r)k}, \hat{\mathbf{B}}_{LM(h)k}\}$ can be obtained accordingly as

$$\text{Cov}\{\hat{\mathbf{B}}_{LM(r)k}, \hat{\mathbf{B}}_{LM(h)k}\} = \hat{\sigma}_{rhk} \left[\hat{\boldsymbol{\xi}}_L^X \boldsymbol{\Lambda}_k (\hat{\boldsymbol{\xi}}_L^X)^T \right]^{-1} \hat{\boldsymbol{\xi}}_L^X \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k (\hat{\boldsymbol{\xi}}_L^X)^T \left[\hat{\boldsymbol{\xi}}_L^X \boldsymbol{\Lambda}_k (\hat{\boldsymbol{\xi}}_L^X)^T \right]^{-1}, \quad (28)$$

where $\boldsymbol{\Lambda}_k$ is the $n \times n$ diagonal matrix with diagonal elements the posterior probabilities τ_{ki} , $i = 1, \dots, N$. In this article, we will always use the studentized FMRCC when not differently specified.

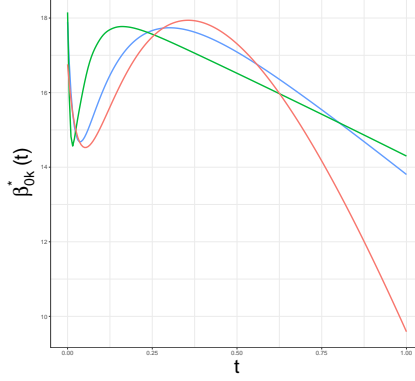


Figure 2: Generated functional intercepts $\beta_{0k}^*(t)$

3 Simulation Study

3.1 Data generation

To evaluate the performance of the proposed FMRCC, the functional response has been generated according to the following FLM

$$Y(t) = (1 - \Delta_2)\beta_{0k}(t) + \int_{\mathcal{S}} \Delta_2(\beta_k(s, t))^T X(s)ds + \varepsilon(t), \quad t \in \mathcal{T}, k = 1, \dots, K, \quad (29)$$

by setting $K = 3$ clusters and $p = 1$ functional covariate. On top of these choices, which do not impact the generality of the study, three different simulation scenarios are generated based on the difference in the FLM structure across clusters. The parameter $\Delta_2 = \{0, 0.5, 1\}$ introduced in Equation (29) controls the generation of three scenarios: $\Delta_2 = 0$ corresponds to clusters that differ only in the intercept $\beta_{0k}(t)$, with no influence of covariates; $\Delta_2 = 1$ corresponds to clusters that differ only in $\beta_k(s, t)$ with no intercept, $\forall k = 1, 2, 3$; $\Delta_2 = 0.5$ generates clusters characterized simultaneously by different functional intercepts and regression coefficient functions.

The functional intercepts $\beta_{0k}(t)$ and the regression coefficient functions $\beta_k(s, t)$ are obtained as follows

$$\beta_{0k}(t) = (1 - \Delta_1)\beta_{01}^*(t) + \Delta_1\beta_{0k}^*(t) \quad t \in \mathcal{T}, k = 2, 3, \quad (30)$$

$$\beta_k(s, t) = (1 - \Delta_1)\beta_1^*(s, t) + \Delta_1\beta_k^*(s, t) \quad t \in \mathcal{T}, k = 2, 3, \quad (31)$$

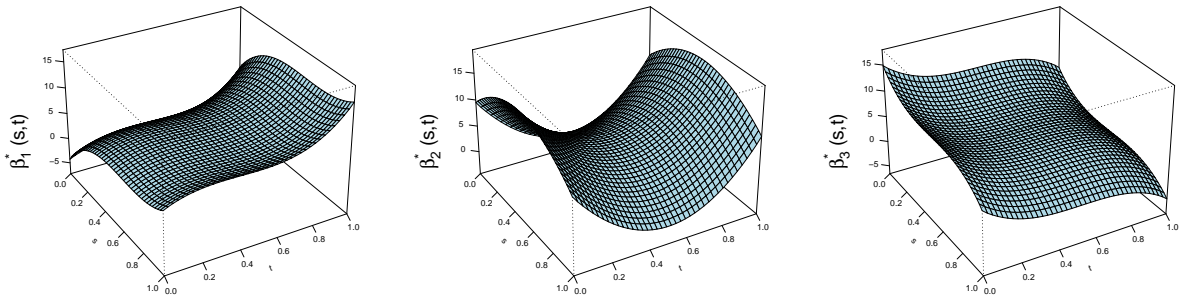


Figure 3: Generated coefficient functions $\beta_k^*(s, t)$

where $\beta_{0k}^*(t)$ and $\beta_k^*(s, t)$, $k = 1, 2, 3$, are the different functional intercepts and regression coefficient functions depicted in Figure 2 and Figure 3, respectively. The parameter $\Delta_1 = \{0, 0.33, 0.66, 1\}$ controls the dissimilarity between the generated clusters: the larger Δ_1 the more distinct the clusters. Accordingly, $\Delta_1 = 0$ corresponds to a single cluster. A more detailed description of the data generation process is given in the Supplementary Materials.

Although data are observed through noisy discrete values, each component of the generated observations is obtained by the spline smoothing approach with a roughness penalty in the second derivative, using 80 cubic B-splines, with the penalty parameter chosen via the generalized cross-validation criterion (GCV), as commonly described by Ramsay and Silverman (2005).

The purpose of the simulation is to assess the performance of FMRCC in identifying any deviation from the IC heterogeneous population, as defined in Section 2.4, in the presence of changes in the mean function of \tilde{Y} conditional on \tilde{X} , $E(\tilde{Y}(t)|\tilde{X})$. From Equation (2) the shifts in $E(\tilde{Y}(t)|\tilde{X})$ can result from the changes in $\beta_{0k}(t)$ and $\beta_k(s, t)$. However, the latter, in addition, can also affect the variability of the functional regression residuals. Because we are interested in the performance of FMRCC in identifying mean function shifts in response, given that the variability of residuals is assumed constant, only shifts caused by changes in $\beta_{0k}(t)$ are considered. Both linear and quadratic types of the shift are considered to represent a change in the slope or curvature of the FLM.

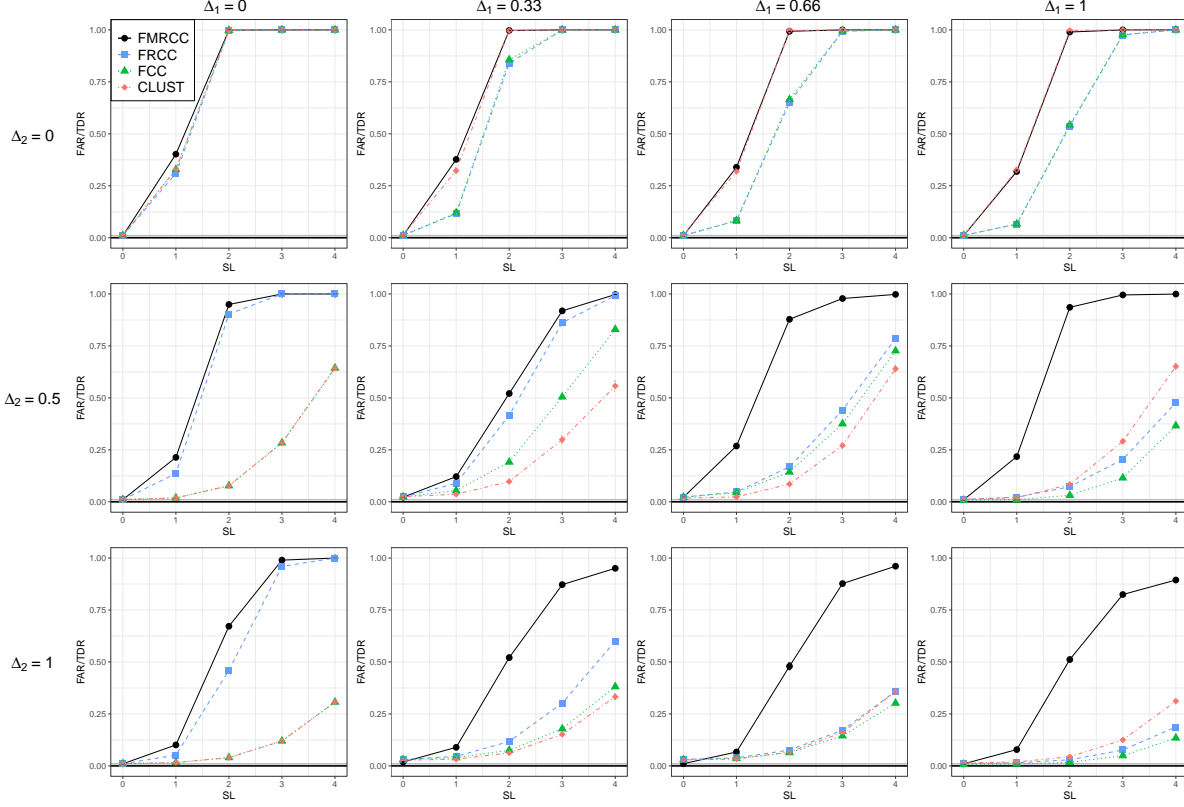


Figure 4: Mean FAR (Severity = 0) or TDR achieved in Phase II by FMRCC, FRCC, FCC, and CLUST for each combination of Δ_1 and Δ_2 for the shift type linear as a function of the severity level.

3.2 Simulation results and discussion

The proposed FMRCC is compared to three different profile monitoring methods: (a) the functional regression control chart (FRCC) of Centofanti et al. (2021); (b) a functional control chart (FCC), which monitors the scores coming from the functional principal component decomposition of the standardized response Y via Hotelling's T^2 and SPE control charts; (c) the FCC applied after a functional clustering step, based on the works of Jacques and Preda (2013) and Grasso et al. (2017), referred to as CLUST. More specifically, the latter consists of a model-based clustering step on the principal component scores of Y and the use of Hotelling's T^2 and SPE control chart on each cluster separately. The BIC and the maximum a posteriori principle (MAP) are adopted to choose the number of clusters and assign the observations to each group, respectively.

Four severity levels of the mean shift (see the Supplementary Materials) are explored. For each combination of Δ_1 , Δ_2 , shift type and severity level, 100 simulation runs are

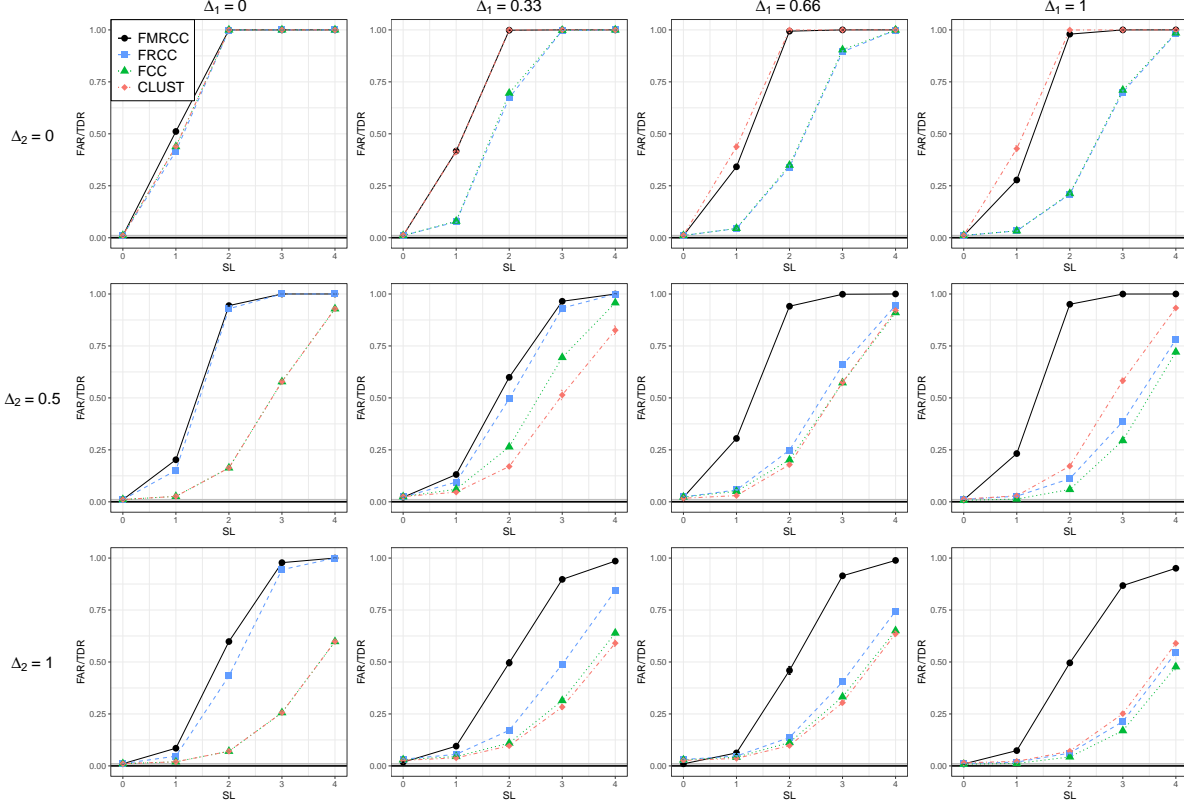


Figure 5: Mean FAR (Severity = 0) or TDR achieved in Phase II by FMRCC, FRCC, FCC, and CLUST for each combination of Δ_1 and Δ_2 for the shift type quadratic as a function of the severity level.

performed. According to Section 2.4, in each run, a training set and a tuning set are independently simulated from IC patterns with 400 and 1000 observations for each cluster, respectively, to mimic a Phase I sample. The number of principal components L and M retained for X and Y is chosen so that the fraction of variance explained (FVE) is at least 95%.

In Phase II, a testing set of 3000 OC observations is randomly generated, and the control chart performance is evaluated by means of the mean true detection rate (TDR) and the mean false alarm rate (FAR), which are estimated as the average proportion, over the simulation runs, of points that fall outside the control limits, whilst the process is, respectively, OC or IC. The mean FAR should be as similar as possible to the overall type-I error probability, whereas the mean TDR should be as large as possible.

The simulation results for the linear shift and the quadratic shift, respectively, are displayed in Figure 4 and Figure 5, which show the mean FAR and TDR as a function of

the severity level for each combination of Δ_1 , Δ_2 , and severity level. For the linear shift, when the clusters differ in the functional intercept $\beta_{0k}(t)$ and there is no functional slope ($\Delta_2 = 0$), the first row of panels in Figure 4 shows that, as the dissimilarity between the clusters increases, FRCC and FCC decrease their performance, while, as expected, CLUST and FMRCC have stable and similar TDRs. However, when no clusters are present (corresponding to $\Delta_1 = 0$) FMRCC performs slightly better than FRCC. When clusters differ in functional intercept and, simultaneously, in regression coefficient functions (corresponding to $\Delta_2 = 0.5$), FMRCC does not detect the correct number of clusters for a value of $\Delta_1 = 0.33$, while it succeeds in detecting the correct number of clusters for a value of $\Delta_1 = 0$, $\Delta_1 = 0.66$ and $\Delta_1 = 1$. As expected, incorporating covariate information is key to the proposed FMRCC performance. In particular, CLUST selects the correct number of clusters for each value of Δ_1 , due to the presence of different functional intercepts $\beta_{0k}(t)$. However, the poor performance of the FRCC, which does not account for the different FLM structures, shows its inability to deal with a heterogeneous population. A more specific advantage of the proposed FMRCC arises when the clusters differ in regression coefficient functions $\beta_k(s, t)$ only ($\Delta_2 = 1$), while keeping the intercept $\beta_{0k}(s, t) = 0$. In this situation, depicted in the third row of Figure 4, FMRCC largely outperforms all competing methods and succeeds in detecting the correct number of clusters for each value of Δ_1 , while CLUST fails since the common functional intercept induces it to prefer only one cluster. FRCC, even worse than the previous case, is inadequate for capturing the different regression structures. For the quadratic shift, the simulation results shown in Figure 5 are analogous to the case shown in Figure 4. Minor differences arise in the competing method's capability of detecting a change in the profile pattern curvature compared to a modification in slope. FMRCC still outperforms all competing methods when the regression coefficient functions differ from 0, i.e., $\Delta_2 \neq 0$, while remaining competitive when $\Delta_2 = 0$.

4 Case Study

The practical applicability of the FMRCC is demonstrated through the case study, already mentioned in the introduction, on the SPM of an RSW process in the automotive industry through the observations of the DRC, which is known to be an informative proxy of the

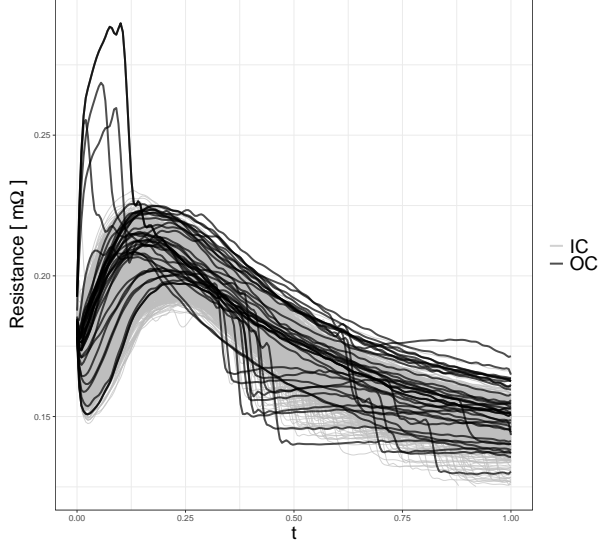


Figure 6: 37 OC DRCs used in Phase II on the IC Phase I DRCs.

final quality of a welded spot. Several factors are known to affect the behavior of DRCs, such as the expulsion of molten material from the weld and the wear of the electrode tips (Capezza et al., 2024a). In particular, the expulsion phenomenon is associated with a notable resistance drop and is usually due to an excessive current and/or inadequate clamping force in the welding spot (Valaee-Tale et al., 2020). However, recent findings from downstream ultrasound inspections, routinely used to certify the actual quality of spot welds on a very low percentage of car bodies, suggest that many spot welds showing signs of molten material expulsion are not necessarily defective. In fact, under certain conditions, expulsion does not always compromise the final quality of the spot weld. Therefore, the corresponding DRCs, despite exhibiting heterogeneous behavior, should be included in the Phase I sample, leading to multimodal IC data. Along the same line, electrode tip wear, which is known to cause changes in electrical, thermal, and mechanical conditions at the electrode tip and metal sheets interfaces (Manladan et al., 2017), is counteracted by periodic electrode tip dressing to avoid inadequate final quality of welded joints. Therefore, its consequences on DRC behavior, already depicted in Figure 1, may still be representative, up to a certain level, of the IC state and lead to another IC mode in the 1802 IC DRCs of the RSW dataset considered in the Phase I sample. The remaining 37 OC DRCs were instead used as Phase II observations. For illustrative purposes, in Figure 6, the OC DRCs (black line) are superimposed on the IC DRCs (grey line). It is worth remarking that, although

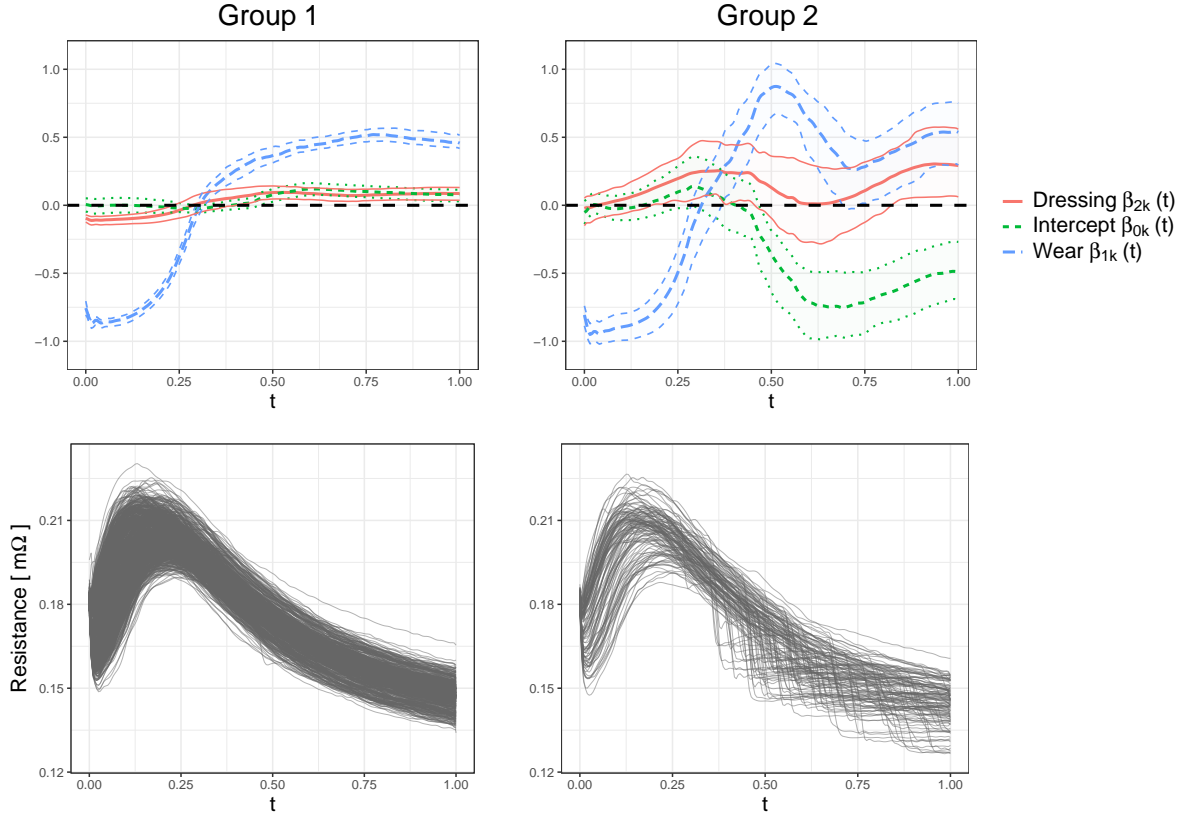


Figure 7: Top panels: Estimated coefficient functions of the two groups detected by the FMRCC along with 95% pointwise confidence intervals (CI). Bottom panels: The corresponding DRCs assigned by the MAP rule to each group.

the FMRCC framework is developed in the more general case of univariate functional response and multivariate functional predictors, in this case study the two predictors are scalar and will not be subject to dimension reduction. As described in Section 2, the Phase I sample is partitioned into training and tuning sets by randomly selecting 901 observations without remittance for the former and the remaining 901 observations for the latter. Five components of the functional response Y are retained to account for at least 95% of the total variance explained. Two mutually exclusive groups with different regression structures are suggested by BIC. The estimates of the regression functions with the 95% pointwise confidence intervals shown in the top panels of Figure 7 are obtained based on 100 bootstrap samples via the non-parametric bootstrap approach of Yao et al. (2011), which includes a label-switching strategy to avoid non-identifiability of component labels. The corresponding DRCs assigned by the MAP rule are shown in the bottom panels of 7. The intercept term of group 2 reveals a drop in DRCs, which clearly corresponds to

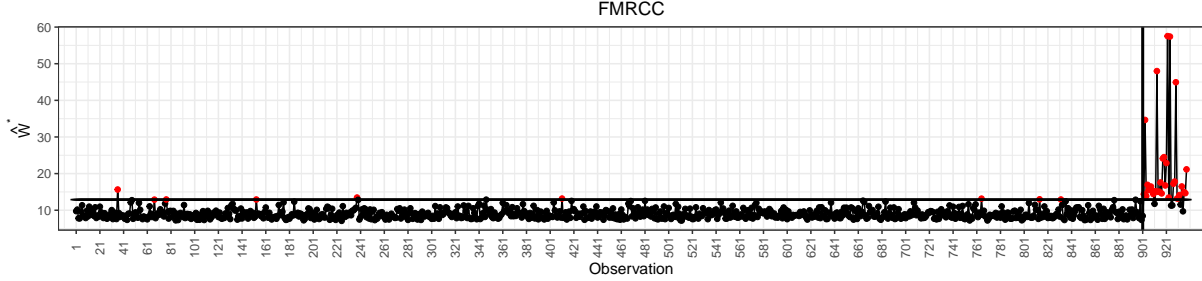


Figure 8: The proposed FMRCC applied to the RSW dataset. The vertical line separates the monitoring statistic calculated for the tuning set on the left and the Phase II observations on the right, while the control limit is shown as a horizontal line.

spot welds affected by the expulsion phenomenon. The wear effect appears to be clearly mapped by a resistance decrease in the first portion of the DRC domain in both clusters, while the dressing effect manifests itself by an overall modest increase of the resistance in the DRC, with a slight decrease in a small initial domain portion.

Figure 8 shows the proposed FMRCC method applied to the RSW dataset. The vertical line separates the monitoring statistic \hat{W}^* calculated for the tuning set, on the left, and the Phase II observations, on the right, while the control limit is represented by the horizontal line. This figure shows that 32 over 37 OC DRCs are correctly detected, resulting in an estimated TDR, denoted by \widehat{TDR} , of 0.864. This value is reported in Table 1 along with that achieved by the competing methods on the RSW dataset, with the same training and tuning sets and Phase II observations. Finally, to quantify the uncertainty of the \widehat{TDR} , Table 1 reports the bootstrap 95% confidence intervals (CI) of the empirical bootstrap distribution (Efron and Tibshirani, 1986) of \widehat{TDR} for the FMRCC and the competing methods presented in Section 3. The CI achieved by the FMRCC is above and non-overlapping those of all competing approaches and confirms its superior performance also in this real scenario, where the capability of simultaneously accounting for the variability explained by the covariates and the heterogeneous structure of the population is decisive.

5 Conclusions

A new framework, referred to as functional mixture regression control chart (FMRCC), is proposed for statistical process monitoring (SPM) of a functional quality characteristic linked to functional and/or scalar covariates by possibly different functional linear models

	\widehat{TDR}	CI
FMRCC	0.864	[0.756,0.960]
FRCC	0.486	[0.310,0.581]
FCC	0.486	[0.323,0.608]
CLUST	0.621	[0.472,0.729]

Table 1: Estimated TDR values, denoted as \widehat{TDR} , on the Phase II observations and the corresponding bootstrap 95% confidence interval (CI) for each monitoring method.

(FLM). The FMRCC is the first profile monitoring framework that is able, by means of a likelihood-ratio test (LRT), to jointly enhance the SPM performance by exploiting additional information on covariates and allowing the subjects to obey different FLMS, thus simultaneously modeling the variability explained by covariates and the hidden heterogeneity of the population structure.

From an extensive Monte Carlo simulation, the FMRCC demonstrated, in the function-on-function regression setting, its superiority in identifying mean shifts in the response over three competing methods suggested by the multimode profile monitoring literature. The practical applicability of the proposed method is illustrated through the case study that motivated this research on the SPM of a resistance spot welding (RSW) process in the automotive industry where heterogeneity of the functional quality characteristic, the dynamic resistance curve (DRC), is modeled through additional information on the wear of the electrode tips and the number of tip dressings the electrode has been subject to. In the case study, the FMRCC confirms that it outperforms all competitors with a more prompt ability to identify the out-of-control state of the RSW process.

Future research can be addressed to extend the proposed framework to nonlinear functional models or to combine it with profile registration techniques and handle profile observations with different domains. In addition, time-weighted monitoring strategies, such as cumulative sum (CUSUM) and exponentially weighted moving average (EWMA), could be implemented to increase the reactivity of the control chart when small and persistent shifts from the in-control population are of interest.

Supplementary Materials

Supplement A contains additional details about the data generation process in the simulation study, omitted from the main manuscript for brevity.

Funding

The research activity of A. Lepore and F. Centofanti were carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). The research activity of B. Palumbo was carried out within the MOST - Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033.17-06-2022, CN00000023). This work reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

- Capezza, C., Capizzi, G., Centofanti, F., Lepore, A., and Palumbo, B. (2024a). An adaptive multivariate functional ewma control chart. *Journal of Quality Technology*.
- Capezza, C., Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B., and Vantini, S. (2023). Funcharts: Control charts for multivariate functional data in r. *Journal of Quality Technology*, 55(5):566–583.
- Capezza, C., Centofanti, F., Lepore, A., and Palumbo, B. (2021). Functional clustering methods for resistance spot welding process data in the automotive industry. *Applied Stochastic Models in Business and Industry*, 37(5):908–925.
- Capezza, C., Centofanti, F., Lepore, A., and Palumbo, B. (2024b). Robust multivariate functional control chart. *Technometrics*.
- Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B., and Vantini, S. (2021). Functional regression control chart. *Technometrics*, 63(3):281–294.
- Centofanti, F., Lepore, A., and Palumbo, B. (2024). Sparse and smooth functional data clustering. *Statistical Papers*, 65(2):795–825.

- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, pages 1571–1596.
- Chou, S.-H., Chang, S. I., and Tsai, T.-R. (2014). On monitoring of multiple non-linear profiles. *International Journal of Production Research*, 52(11):3209–3224.
- Ciarleglio, A. and Ogden, R. T. (2016). Wavelet-based scalar-on-function finite mixture regression models. *Computational Statistics & Data Analysis*, 93:86–96.
- Colosimo, B. M. and Pacella, M. (2010). A comparison study of control charts for statistical monitoring of functional data. *International Journal of Production Research*, 48(6):1575–1601.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. 39(1):1–22.
- DeSarbo, W. S. and Cron, W. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5:249–282.
- Devijver, E. (2017). Model-based regression clustering for high-dimensional data: application to functional data. *Advances in Data Analysis and Classification*, 11:243–279.
- Dickinson, D., Franklin, J., Stanya, A., et al. (1980). Characterization of spot welding behavior by dynamic electrical parameter monitoring. *Welding Journal*, 59(6):170.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75.
- Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631.
- Grasso, M., Colosimo, B. M., and Pacella, M. (2014). Profile monitoring via sensor fusion: the use of pca methods for multi-channel data. *International Journal of Production Research*, 52(20):6110–6135.
- Grasso, M., Colosimo, B. M., and Tsung, F. (2017). A phase i multi-modelling approach for profile monitoring of signal data. *International Journal of Production Research*, 55(15):4354–4377.
- Grasso, M., Menafoglio, A., Colosimo, B. M., and Secchi, P. (2016). Using curve-registration information for profile monitoring. *Journal of Quality Technology*, 48(2):99–127.
- Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659.
- Jacques, J. and Preda, C. (2013). Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing*, 112:164–171.
- Jin, J. and Shi, J. (1999). Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, 41(4):327–339.
- Jones, C. L., Abdel-Salam, A.-S. G., and Mays, D. (2021). Practitioners guide on para-

- metric, nonparametric, and semiparametric profile monitoring. *Quality and Reliability Engineering International*, 37(3):857–881.
- Jones, P. and McLachlan, G. J. (1992). Fitting finite mixture models in a regression context. *Australian Journal of Statistics*, 34(2):233–240.
- Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC.
- Kruger, U. and Xie, L. (2012). *Statistical monitoring of complex multivariate processes: with applications in industrial process control*. John Wiley & Sons.
- Maleki, M. R., Amiri, A., and Castagliola, P. (2018). An overview on recent profile monitoring papers (2008–2018) based on conceptual classification scheme. *Computers & Industrial Engineering*, 126:705–728.
- Mandel, B. (1969). The regression control chart. *Journal of Quality Technology*, 1(1):1–9.
- Manladan, S., Yusof, F., Ramesh, S., Fadzil, M., Luo, Z., and Ao, S. (2017). A review on resistance spot welding of aluminum alloys. *The International Journal of Advanced Manufacturing Technology*, 90(1):605–634.
- McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley.
- McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.
- Menafoglio, A., Grasso, M., Secchi, P., and Colosimo, B. M. (2018). Profile monitoring of probability density functions via simplicial functional pca with application to image data. *Technometrics*, 60(4):497–510.
- Noorossana, R., Saghaei, A., and Amiri, A. (2011). *Statistical analysis of profile monitoring*. John Wiley & Sons.
- Park, C. and Shrivastava, A. K. (2014). Multimode geometric-profile monitoring with correlated image data and its application to nanoparticle self-assembly processes. *Journal of Quality Technology*, 46(3):216–233.
- Paynabar, K., Zou, C., and Qiu, P. (2016). A change-point approach for phase-i analysis in multivariate profile monitoring and diagnosis. *Technometrics*, 58(2):191–204.
- Qiu, P. (2013). *Introduction to statistical process control*. CRC press.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramaker, H.-J., van Sprang, E. N., Westerhuis, J. A., and Smilde, A. K. (2004). The effect of the size of the training set and number of principal components on the false alarm rate in statistical process monitoring. *Chemometrics and intelligent laboratory systems*, 73(2):181–187.
- Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer.

- Ren, H., Chen, N., and Wang, Z. (2019). Phase-II monitoring in multichannel profile observations. *Journal of Quality Technology*, 51(4):338–352.
- Sain, S. R., Gray, H., Woodward, W. A., and Fisk, M. D. (1999). Outlier detection from a mixture distribution when training data are unlabeled. *Bulletin of the Seismological Society of America*, 89(1):294–304.
- Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *The R journal*, 8(1):289.
- Valaee-Tale, M., Sheikhi, M., Mazaheri, Y., Ghaini, F. M., and Usefifar, G. R. (2020). Criterion for predicting expulsion in resistance spot welding of steel sheets. *Journal of Materials Processing Technology*, 275:116329.
- Wang, K., Li, J., and Tsung, F. (2018a). Registration-free monitoring of multimode near-circular shape profiles. *Quality and Reliability Engineering International*, 34(4):529–542.
- Wang, K., Ng, S.-K., and McLachlan, G. J. (2009). Multivariate skew t mixture models: applications to fluorescence-activated cell sorting data. In *2009 Digital Image Computing: Techniques and Applications*, pages 526–531. IEEE.
- Wang, S., Huang, M., Wu, X., and Yao, W. (2016). Mixture of functional linear models and its application to CO₂-GDP functional data. *Computational Statistics & Data Analysis*, 97:1–15.
- Wang, S., Woodward, W. A., Gray, H., Wiechecki, S., and Sain, S. R. (1997). A new test for outlier detection from a multivariate mixture distribution. *Journal of Computational and Graphical Statistics*, 6(3):285–299.
- Wang, Y., Mei, Y., and Paynabar, K. (2018b). Thresholded multivariate principal component analysis for phase I multichannel profile monitoring. *Technometrics*, 60(3):360–372.
- Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, 36(3):309–320.
- Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103.
- Yao, F., Fu, Y., and Lee, T. C. (2011). Functional mixture regression. *Biostatistics*, 12(2):341–353.
- Zhao, Y., Ogden, R. T., and Reiss, P. T. (2012). Wavelet-based lasso in functional linear regression. *Journal of computational and graphical statistics*, 21(3):600–617.
- Zou, C., Ning, X., and Tsung, F. (2012). Lasso-based multivariate linear profile monitoring. *Annals of operations research*, 192(1):3–19.

Supplementary Materials to “Functional Mixture Regression Control Chart”

Christian Capezza¹, Fabio Centofanti¹, Davide Forcina¹, Antonio Lepore¹,
and Biagio Palumbo*¹

¹*Department of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125, Naples, Italy*

A Details on Data Generation in the Simulation Study

The data are generated according to the FLM

$$\tilde{Y}(t) = (1 - \Delta_2)\beta_{0k}(t) + \int_{\mathcal{S}} \Delta_2(\beta_k(s, t))^T \tilde{X}(s) ds + \varepsilon(t), \quad t \in \mathcal{T}, k = 1, \dots, K, \quad (\text{A.1})$$

where the compact domains \mathcal{S} and \mathcal{T} are set, without loss of generality, equal to $[0, 1]$ and the number of covariates p and clusters K are set equal to 1 and 3, respectively. The functional covariates $\tilde{X}(s)$ are generated by the eigenfunction set $\{\psi_l\}_{l=1,2,\dots}$, as in Centofanti et al. (2021), by considering the correlation function G through the following steps.

1. Set $G(s, t) = G(z)$, where $z = |s - t|$, as the *powered exponential* correlation function (Stein, 1999). The general form of the correlation function and parameter used are listed in Table 1.
2. Calculate the eigenvalues λ_l and the corresponding eigenfunctions ψ_l , $l = 1, 2, \dots, L^*$, of G , where L^* is set equal to 50.

*Corresponding author. e-mail: biagio.palumbo@unina.it

Table 1. Powered exponential correlation function and parameter for covariates generation in the simulation study.

	ρ	ν
$P(z) = \exp - \left(\frac{ z }{\rho}\right)^\nu$	1	0.5

Table 2. Parameters a, b, c, d, e used for the functional coefficient functions generation.

	a	b	c	d	e
$\beta_1^*(s, t)$	0.3	0.3	0.3	0.3	5
$\beta_2^*(s, t)$	0.2	0.15	0.9	0.9	-5
$\beta_3^*(s, t)$	0.9	0.9	-0.3	-0.3	5

3. Let \tilde{X} as

$$\tilde{X} = \sum_{l=1}^{L^*} \xi_l^{\tilde{X}} \psi_l^{\tilde{X}}, \quad (\text{A.2})$$

with $\xi_{L^*}^{\tilde{X}} = \left(\xi_1^{\tilde{X}}, \dots, \xi_{L^*}^{\tilde{X}}\right)^T$ generated by means of a multivariate normal distribution with mean 0 and covariance $\text{Cov}\left(\xi_{L^*}^{\tilde{X}}\right) = \text{diag}(\lambda_1, \dots, \lambda_{L^*})$.

Covariate generation is made by means of the R package `funcharts` of Capezza et al. (2023).

Inspired by the coefficient functions generation process proposed in Centofanti et al. (2022), the regression coefficient functions $\beta_k^*(s, t)$, depicted in Figure 3, are obtained through the following reference model for $k = 1, 2, 3$

$$\beta_k^*(s, t) = \left(\frac{t-0.5}{c}\right)^3 + \left(\frac{s-0.5}{d}\right)^3 + \left(\frac{t-0.5}{b}\right)^2 - \left(\frac{s-0.5}{a}\right)^2 + e, \quad s, t \in [0, 1], \quad (\text{A.3})$$

where the parameters c, d, b, a, e are real numbers depicted in Table 2.

Furthermore, the functional intercepts $\beta_{0k}^*(t)$, depicted in Figure 2, are obtained through the following reference model for $k = 1, 2, 3$

$$\begin{aligned} \beta_{0k}^*(t) = & f + 0.3117 \exp[-371.4u_k(t)] + 0.5284\{1 - \exp[gu_k(t)]\} \\ & - 423.3\{1 + \tanh[-hu_k(t) + 0.1715]\} \quad t \in \mathcal{T}, \quad (\text{A.4}) \end{aligned}$$

where the parameters f, g and h are real numbers depicted in Table 3 and $u_k(t)$ is defined

Table 3. Parameters f , g and h used for the functional intercepts generation.

	f	g	h
$\beta_{01}^*(t)$	0.2074	0.8217	26.15
$\beta_{02}^*(t)$	0.187	0.2	27
$\beta_{03}^*(t)$	0.3	4	24

as follows

$$u_k(t) = \frac{t - \min(t)}{\max(t) - \min(t)}(m - 0.0045) + 0.0045 \quad t \in \mathcal{T}, \quad (\text{A.5})$$

with $m = 0.15, 0.4, 0.08$ for $k = 1, 2, 3$ respectively. Note that the mean functions $\beta_{0k}^*(t)$ are generated to resemble a typical DRC through the phenomenological model for the RSW process presented in Schwab et al. (2012).

The parameter $\Delta_1 = \{0, 0.33, 0.66, 1\}$ represents a measure of dissimilarity between the generated clusters according to the equations

$$\beta_{0k}(t) = (1 - \Delta_1)\beta_{01}^*(t) + \Delta_1\beta_{0k}^*(t) \quad t \in \mathcal{T}, k = 2, 3 \quad (\text{A.6})$$

$$\beta_k(s, t) = (1 - \Delta_1)\beta_1^*(s, t) + \Delta_1\beta_k^*(s, t) \quad t \in \mathcal{T}, k = 2, 3. \quad (\text{A.7})$$

A higher value of Δ_1 corresponds to more distinct clusters, whereas a value of $\Delta_1 = 0$ corresponds to the absence of clusters.

The parameter $\Delta_2 = \{0, 0.5, 1\}$ weighs the functional intercept term and the regression coefficient function term according to Equation (A.1). In particular, a value of $\Delta_2 = 0$ corresponds to clusters that differ only in $\beta_{0k}(t)$, while $\beta_k(s, t) = 0 \quad \forall k = 1, 2, 3$. On the other hand, a value of $\Delta_2 = 1$ corresponds to clusters that differ only in $\beta_k(s, t)$, while $\beta_{0k}(t) = 0 \quad \forall k = 1, 2, 3$. Finally, a value of $\Delta_2 = 0.5$ indicates the presence of clusters characterized simultaneously by different functional intercepts and regression coefficient functions.

Moreover, to have noisy realizations of the response, a functional error term $k\varepsilon(t) = \sum_{i=1}^{20} e_i\psi_i^e(t)$ is generated, where e_i are independent realizations of a standard normal random variable and $\psi_i^e(t)$ are cubic B-splines with evenly spaced knot sequence. The constant k is chosen to have a signal-to-noise ratio (SNR) defined as Städler et al. (2010) equal to 10. Finally, it is assumed that the generated data were observed discretely at 500 equally

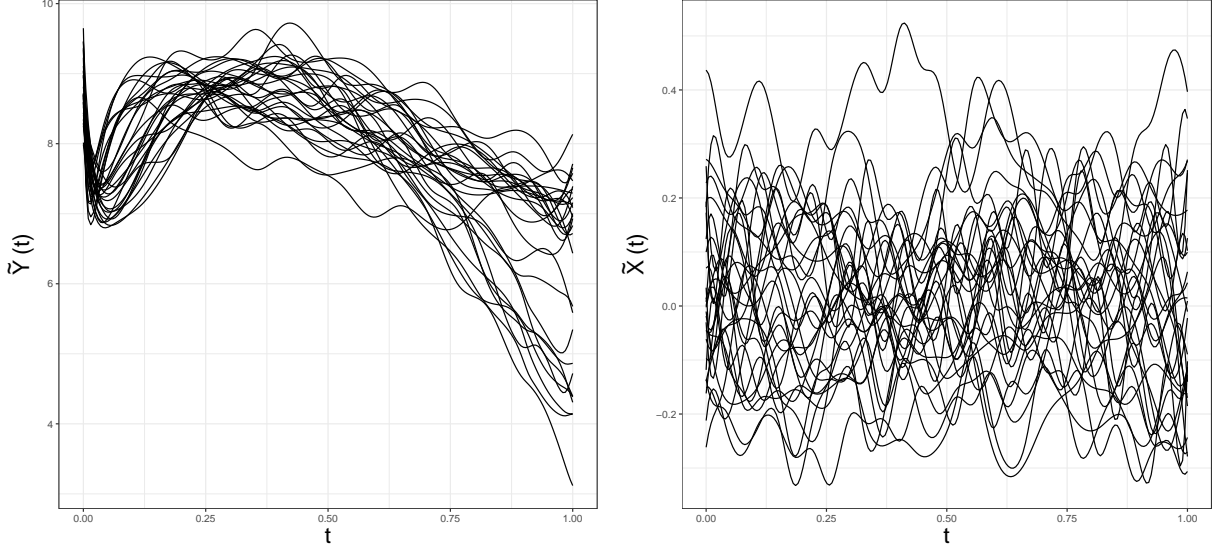


Figure 1. Sample of 30 IC randomly generated \tilde{Y} and \tilde{X} functions corresponding to $\Delta_1 = 1$ and $\Delta_2 = 0.5$

Table 4. Shift types according to Equation A.8

Shift	δ_1	δ_1
Linear	0	$s \cdot q_l$
Quadratic	$s \cdot q_q$	0

spaced time points throughout the domain $[0, 1]$.

For illustrative purposes, a sample of 30 IC randomly generated \tilde{Y} and \tilde{X} functions corresponding to $\Delta_1 = 1$ and $\Delta_2 = 0.5$ are shown in Figure 1.

Without loss of generality, the Phase II out-of-control functional profiles are generated from the first cluster using the model in Equation (A.1) with an additional shift term $\delta(t)$ defined as follows

$$\delta(t) = \delta_1 t^2 + \delta_2 t, \quad t \in [0, 1]. \quad (\text{A.8})$$

The real numbers δ_1 and δ_2 are set equal to positive severities as reported in Table 4, where two different shift types are considered. The terms q_l and q_q are set equal to 1.2 and 1.6, respectively, and s represents five severity levels equal to 0, 0.375, 0.75, 1.25, 1.5.

The linear shift is representative of a change in the slope of the profile pattern, whereas the quadratic shift represents curvature modification.

References

- Capezza, C., F. Centofanti, A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini (2023). Fun-charts: Control charts for multivariate functional data in r. *Journal of Quality Technology* 55(5), 566–583.
- Centofanti, F., M. Fontana, A. Lepore, and S. Vantini (2022). Smooth lasso estimator for the function-on-function linear regression model. *Computational Statistics & Data Analysis* 176, 107556.
- Centofanti, F., A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini (2021). Functional regression control chart. *Technometrics* 63(3), 281–294.
- Schwab, I., M. Senn, and N. Link (2012). Improving expert knowledge in dynamic process monitoring by symbolic regression. In *2012 Sixth International Conference on Genetic and Evolutionary Computing*, pp. 132–135. IEEE.
- Städler, N., P. Bühlmann, and S. Van De Geer (2010). ℓ_1 -penalization for mixture regression models. *Test* 19, 209–256.
- Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.