

# Functional regression control chart for monitoring ship CO<sub>2</sub> emissions

Christian Capezza<sup>1</sup>  | Fabio Centofanti<sup>1</sup>  | Antonio Lepore<sup>1</sup>   
 Alessandra Menafoglio<sup>2</sup> | Biagio Palumbo<sup>1</sup>  | Simone Vantini<sup>2</sup> 

<sup>1</sup> Department of Industrial Engineering,  
 University of Naples Federico II, Naples,  
 Italy

<sup>2</sup> MOX - Modelling and Scientific  
 Computing, Department of Mathematics,  
 Politecnico di Milano, Milan, Italy

## Correspondence

Fabio Centofanti, Department of Industrial Engineering, University of Naples Federico II, Naples, Italy.  
 Email: [fabio.centofanti@unina.it](mailto:fabio.centofanti@unina.it)

## Abstract

On modern ships, the quick development in data acquisition technologies is producing data-rich environments where variable measurements are continuously streamed and stored during navigation and thus can be naturally modelled as functional data or profiles. Then, both the CO<sub>2</sub> emissions (i.e. the quality characteristic of interest) and the variable profiles that have an impact on them (i.e. the covariates) are called to be explored in the light of the new worldwide and European regulations on the monitoring, reporting and verification of harmful emissions. In this paper, we show an application of the functional regression control chart (FRCC) with the ultimate goal of answering, at the end of each ship voyage, the question: *given the value of the covariates, is the observed CO<sub>2</sub> emission profile as expected?* To this aim, the FRCC focuses on the monitoring of residuals obtained from a multivariate functional linear regression of the CO<sub>2</sub> emission profiles on the functional covariates. The applicability of the FRCC is demonstrated through a real-case study of a Ro-Pax ship operating in the Mediterranean Sea. The proposed FRCC is also compared with other alternatives available in the literature and its advantages are discussed over some practical examples.

## KEY WORDS

functional regression, industry 4.0, multivariate functional linear regression, profile monitoring, statistical process monitoring

## 1 | INTRODUCTION

Nowadays, the quick development in the data acquisition (DAQ) technologies is producing data-rich industrial environments where massive amounts of data are available. In particular, a large portion of the ship observational data are complex measurement signals that may be envisaged as reflecting smooth variations of quantities generated by continuous functions defined on an infinite compact domain, that is, as *functional data*. Functional data analysis (FDA) is a thriving area of statistics. For a comprehensive overview, the reader could refer to Ramsay and Silverman,<sup>1</sup> Horvath and Kokoszka<sup>2</sup> and Kokoszka and Reimherr.<sup>3</sup> In this functional data setting, profile monitoring<sup>4</sup> is the name of a new branch of statistical

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Quality and Reliability Engineering International* published by John Wiley & Sons Ltd

process control (SPC) that provides a suite of methods to continuously give a solution to the urging issue of evaluating the stability over time of functional quality characteristics. Recent contributions are Colosimo and Pacella,<sup>5</sup> Grasso et al<sup>6</sup> and Menafoglio et al.<sup>7</sup> As in the classical SPC, where data are scalars, profile control charts have the task of monitoring the functional quality characteristic and of triggering a signal when assignable sources of variations (i.e. special causes) act on it. When this happens, the process is said to be out of control (OC). Otherwise, when only normal sources of variation (i.e. common causes) apply, the process is said to be in control (IC).

Only recently, Centofanti et al<sup>8</sup> introduced the functional regression control chart (FRCC) framework to monitor a functional quality characteristic when this is influenced by one or more functional covariates. In particular, the aim of the FRCC is to improve the monitoring of the quality characteristic by including the information coming from the covariates. In this scenario, if one of these covariates manifests itself with an extreme realization, the quality characteristic may wrongly be judged to be OC, even though it shows perfectly reasonable values given the covariates. Otherwise, there may be situations where the covariates are not extreme and the quality characteristic may wrongly appear IC because the information in the covariates is not used to increase the power of the monitoring.

The FRCC framework is the functional extension of the basic Mandel's idea,<sup>9</sup> where the quality characteristic is monitored after being adjusted for the effect of covariates. That is, the control variable is the residual obtained from a regression of the quality characteristic on the covariates, and the focus is spotted on the residual variability not explained by the knowledge of the observed value of the covariates. In a more direct phrasing, the FRCC answers the question: *given the value of the covariates, is the quality characteristic as expected?* Alternatively, the question could be phrased as: *does the assumed model fit the reality of the quality characteristic?* If the answer is no, then special causes may have occurred that are beyond the information brought by the covariates through the chosen regression model. In particular, in the FRCC framework, the quality characteristic and the covariates are linked through a multiple functional linear regression model (MFLR), where both the response and the explanatory variables can be described by functional data. Recent examples of MFLR model can be found in Palumbo et al,<sup>10</sup> Centofanti et al<sup>11, 12</sup> and Chiou et al.<sup>13</sup> The idea of monitoring model residuals arises also in SPC with autocorrelated data, for example, time series.<sup>14–16</sup> In this setting, residuals from an autoregressive model are used to recover independence of the observations, because conventional control charts are known not to work well if the quality characteristic exhibits even low levels of correlation over time.<sup>17</sup> In the regression control chart idea, residuals are used to adjust the quality characteristic for the information in the covariates.

In recent years, profile monitoring has emerged as an effective technique also in the field of maritime transport where the issue of CO<sub>2</sub> emission monitoring is becoming of paramount importance.<sup>8, 18, 19</sup> Indeed, in view of climate change and global warming crises, the maritime transport industry is currently facing new challenges related to CO<sub>2</sub> emissions. Indeed, the Marine Environment Protection Committee of the International Maritime Organization<sup>20–22</sup> has urged shipping companies to set up a framework for the monitoring, reporting and verification (MRV) of CO<sub>2</sub> emissions based on fuel consumption. In the face of these regulations, shipping companies are updating DAQ systems on their fleets, enabling large volumes of observational data to be automatically streamed and transferred to a remote server, bypassing human intervention. A large proportion of these data can be modelled as functional data, thus representing a new challenge for FDA and related SPC methods in this area. The DAQ system installed on modern ships facilitates in fact the collection of functional data related to CO<sub>2</sub> emissions as well as to other functional covariates affecting them, which include the speed of the ship, engine variables such as the propeller pitch variables, environmental variables that describe the wind and sea conditions or the cumulative navigation time (we refer to Section 2.2 for more details on the variables considered in this work). However, most of the approaches that have already appeared in maritime literature<sup>23–25</sup> do not take advantage of the potential help to managerial decision making represented by the modelling of the entire voyage profiles acquired and usually compress information in one or more scalar features extracted from them. In this setting, a recurrent request posed by maritime engineers is related to the CO<sub>2</sub> emissions corresponding to the given values of other recorded covariates. That is, they want to assess if CO<sub>2</sub> emissions are coherent with the values of the covariates, in order to identify unexpected behaviours and take corrective measures. Engineers are less concerned in identifying CO<sub>2</sub> emission profiles that are extreme with respect to their marginal distribution, if this can be explained by some extreme value in the covariates. They are rather interested in CO<sub>2</sub> emission profiles that are not consistent with covariates affecting them because, for example, they can reveal anomalous ship performance.

The FRCC can be used to meet this engineering need. In this paper, we propose to use the FRCC to monitor ship CO<sub>2</sub> emissions throughout each voyage, in order to identify special causes, at given values recorded by functional covariates. Specifically, we consider a particular implementation of the FRCC framework, where the functional quality characteristic, hereinafter referred to also as response, and the functional covariates are related through the multivariate functional linear regression (MFLR) model. In this paper, the MFLR model is estimated based on the multivariate functional principal components analysis (FPCA).<sup>13, 26</sup> Then, studentized residuals are monitored through the simultaneous application of

**TABLE 1** Technical features of the considered Ro-Pax ship

Features	Value
Gross tonnage	32,728
Length	203.9 m
Beam	25 m
Draft	6.8 m
Maximum power for propulsion	46,080 kW
Maximum speed	28.9 knots

the Hotelling's  $T^2$  and the squared prediction error (SPE) control charts.<sup>4–6, 27</sup> In this paper, the FRCC framework is used both *retrospectively*, as an aid to the practitioner to determine the IC state of the process under study and to identify an IC reference sample (Phase I), and, *prospectively* to monitor any departure from the IC state at future voyages (Phase II). The applicability of the FRCC is demonstrated through a real-case study of a Ro-Pax ship operating in the Mediterranean Sea, courtesy of the shipping company Grimaldi group.

This work does not directly aim at a real-time feedback control, that is, in the online monitoring and immediate actions during an ongoing voyage. The FRCC framework is indeed used to signal OC voyages once they are completed, as  $T^2$  and SPE monitoring statistics are going to be calculated on the entire profiles. Instead, the proposed FRCC focuses on tracking automatically (and possibly the whole fleet) the future OC signals, or patterns and trends that may identify, for example, engine malfunctioning, the need for hull cleaning or for any other energy efficiency initiative. Even though no action could be taken during an ongoing voyage, we believe this paper motivates the use of a functional data approach as the FRCC can be indeed used by shipping companies to evaluate, at the end of a voyage, if the observed CO<sub>2</sub> emission profile is anomalous, given the value of the observed covariates. This information can help to prevent and disguise possible problems in future voyages, or to schedule maintenance operations.

The paper is structured as follows. In Section 2, the structure of the data and technological details of the ship equipment are provided for the real-case study at hand. In Section 3, the main materials and methods behind the particular implementation of the FRCC framework are summarized. In Section 4, by means of the real-case study mentioned before, the FRCC is practically applied to monitor CO<sub>2</sub> emissions and compared with alternative methods already appeared in the SPC literature. In Section 5, we draw conclusions. All computations and plots have been obtained using the software environment R.<sup>28</sup>

## 2 | TECHNOLOGICAL BACKGROUND AND DATA STRUCTURE

For confidentiality reasons, in what follows we omit the name of the ship considered in the real-case study. However, to give an idea of the ship type, in Section 2.1, we provide its main technical features. Whereas, in Section 2.2, we describe in detail all the variables and the data used for the analysis.

### 2.1 | Technical features

The main technical features of the ship are illustrated in Table 1. The considered ship is characterized by two engine sets, each consisting of two main diesel engines for propulsion Wärtsilä, Type 16ZAV40S, four-stroke, with a maximum continuous rating of 11,520 kW at 510 revolutions per minute (rpm) and by two variable pitch propellers and a shaft generator for electric power supply. The main engine power is used both for propulsion and electrical generation through the shaft generators, which are themselves keyed on a gearbox. The gearbox has two fast inlet shafts powered by the engine shaft, a slow outlet shaft for the propeller and a faster one to which the shaft alternator is connected. The gear ratio between the engine shaft and the propeller shaft is equal to 3.24, whereas the gear ratio between the engine shaft and the shaft alternator is equal to 0.32. The main diesel engine can be powered by three types of fuel with different percentage of sulphur (S) content in order to be compliant with the regulation in force on the geographical area to be sailed: heavy fuel oil, very low sulphur fuel oil ( $\leq 0.5\% S$ ), ultra-low sulphur fuel oil or marine gas oil ( $\leq 0.1\% S$ ). The electrical power supply of the ship consists of three diesel generators (1840 kVA, 690 V), two shaft generators (2100 kVA, 690 V) and one emergency diesel generator (480 kVA). The main engines can supply power in two different modes, at fixed rpm (constant mode) or at

variable rpm (combined mode). In the constant mode, shaft generators can be used to supply electric power, even though the maximum speed cannot be reached because speed variations are only possible by changing the pitch of the propellers. Whereas, in the combined mode, the ship speed can be regulated by increasing both pitch propeller and engine rpm, but, vice versa, the possibility to engage the shaft generator is lost.

## 2.2 | Data description

Data come from a DAQ system installed on the ship that are transmitted to the cloud with different frequencies varying from 2 to 5 min. The data refer to a specific route, that is, each profile in the data set corresponds to a voyage of the ship and all voyages have the same departure and arrival ports. A period of 11 consecutive months following a dry-dock operation on the ship is considered. Voyages in the first nine months (i.e. from the beginning of February 2020 to the end of October 2020) are used in Phase I, that is, to identify a reference data set, estimate the model and control chart limits. Voyages in the last 2 months (i.e. from the end of October 2020 to the end of December 2020) are used in Phase II to show the performance of the FRCC on monitoring new voyages. We start with a data set of 190 voyages for the Phase I (Section 4.1); then we remove outliers as described in Section 4.1.2 and we end up with a reference data set of 169 voyages; finally, we consider 22 voyages for the Phase II (Section 4.2).

Note that the data refer to the navigation phase. More specifically, the navigation phase begins with the finished with engine order (when the ship leaves the departure port) and ends with the stand by engine order (when the ship enters the arrival port). Moreover, we need to identify an adequate functional domain for each voyage. Even if time is naturally suitable as a functional domain, total travel time can vary from voyage to voyage. Therefore, we prefer to use the fraction of distance traveled over the voyage as the common domain (0,1) of the data.

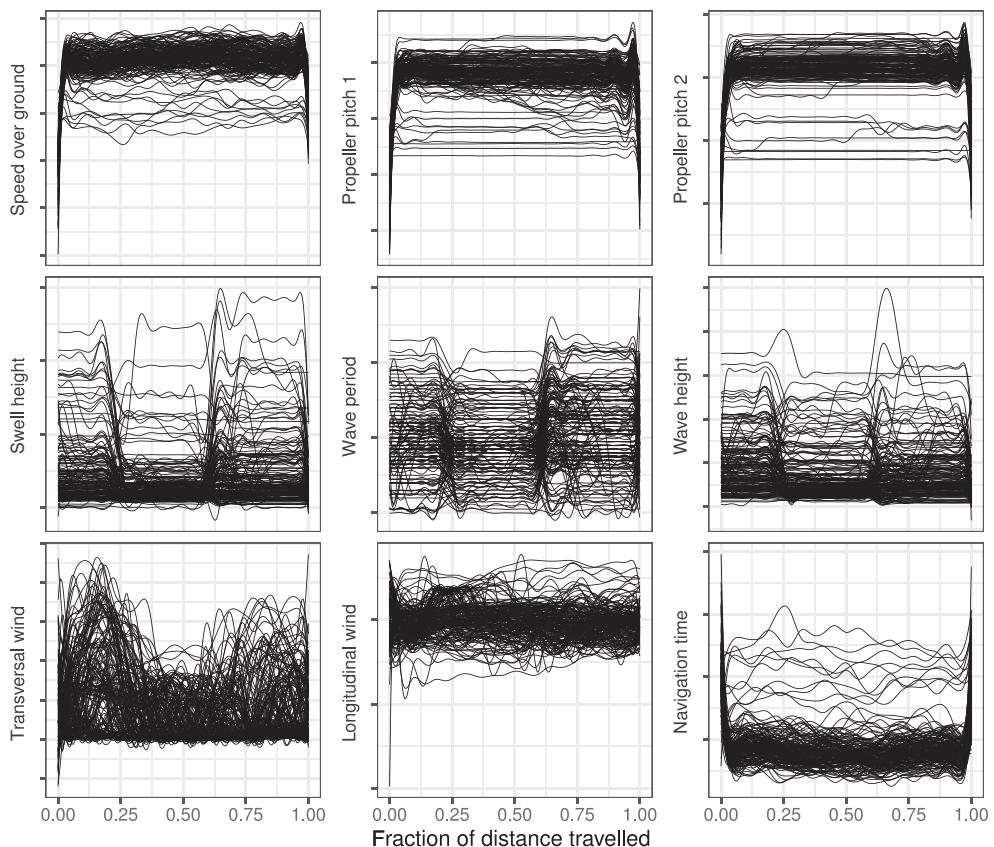
All the signals acquired by the DAQ system are summarized into several variables that we describe here to select later the functional covariates and response considered in the MFLR model. The ship is tracked by its global positioning system (GPS), which provides longitude and latitude coordinates. The course over ground (COG) is the actual direction of progress of a vessel, between two points, with respect to the surface of the earth, measured in degrees. The sailed distance over ground (SDOG) is the distance travelled by the vessel between two points, measured in nautic miles (NM), calculated from the GPS sensor through the Haversine formula. The speed over ground (SOG) is measured in knots (kn) and is the ratio between SDOG and the sailing time, measured in hours. The propeller pitch (P) is measured in degrees and represents the angle between the intersection of the chord line of the blade section and a plane normal to the propeller axis. An anemometer sensor provides data about the true speed ( $V$ ), measured in knots, and direction ( $\psi$ ), measured in degrees, of the wind. The latter is obtained as the difference between the true wind angle in earth system and the COG. Additional information on the wind variables can be found on Bocchetti et al.<sup>25</sup> From the two anemometer variables, the longitudinal component of the wind is calculated as  $V \cos \psi$ , while the transversal component is calculated as  $|V \sin \psi|$ . Note that a positive (respectively, negative) longitudinal component of the wind means that the wind blows from stern (respectively, bow). Moreover, a data fusion process also allows the integration of marine data into the data set, that is, weather forecasts about the sea state furnished by private held weather service provider. The sea state is characterized by the provider through the typical parameters, namely, height and period, used to model waves that, in turn, are roughly divided into two components: wind-driven waves, or simply waves (generated by the immediate local wind) and swell (generated by distant weather systems and usually having larger period). In particular, the height, measured in meters, is defined as the vertical distance from wave crest to wave trough; whereas, the period, measured in seconds, represents the time between successive crests of a train of waves passing a fixed point in a ship, at a fixed angle of encounter.<sup>29</sup> Regarding the CO<sub>2</sub> measurement, MRV regulation proposes direct and indirect methods. The direct method determines the amount of CO<sub>2</sub> emitted measuring the flow of these emissions passing in exhaust gas funnels. Instead, the indirect method calculates the CO<sub>2</sub> emissions based on the fuel consumption. The direct method is based on the determination of CO<sub>2</sub> emitted that flow in exhaust gas stacks based on the measurement of the CO<sub>2</sub> in the exhaust gas and the measurement of the volume of the exhaust gas flow per unit of time. This method is very sensitive to the calibration and the uncertainty related to the measurement devices. Whereas, the class of indirect method calculates CO<sub>2</sub> emissions as a product of the whole amount of fuel consumption of the main and auxiliary engines, boilers, gas turbines and inert gas generators times the so-called *emission factor*, which is calculated as the average emission rate of a greenhouse gas (GHG) relative to the activity data of a source stream, assuming complete oxidation for combustion and complete conversion for all other chemical reactions. In this paper, we use the indirect method and we focus on the main engines only.

In what follows, we list the functional variables chosen for the analysis in this work that are obtained from the signals acquired by the DAQ system. The functional response is the signal corresponding to the CO<sub>2</sub> emissions per hour along the

**TABLE 2** Functional covariates used in the MFLR

	Variable name	Unit of measurement
1	Speed over ground (SOG)	kn
2	Propeller pitch 1	degrees
3	Propeller pitch 2	degrees
4	Swell height	m
5	Wave period	s
6	Wave height	m
7	Longitudinal wind component	kn
8	Transverse wind component	kn
9	Navigation time	h/(NM[%])

Note: For confidentiality reasons, y-axis labels are omitted.

**FIGURE 1** Functional covariates in the reference data set

entire voyage. In order to select functional covariates among the available signals, a very long preliminary investigation was carried out to identify the covariates that could better explain the CO<sub>2</sub> emissions. However, in practice, many signals that could have played the role of covariates were not able to be measured accurately. The intersection between the set of candidate and truly measurable covariates was finally identified after an intensive exchange of information and experience with marine engineers, shipping managers and operators. The following nine functional covariates have been identified, which are, thus, assumed as a characterization of the ship operational conditions: SOG, left propeller pitch, right propeller pitch, transversal and longitudinal component of the wind, wave height, wave period, swell height and derivative of the cumulative navigation time. Table 2 reports the complete list of the functional covariates considered in this work.

Moreover, Figures 1 and 2 show the profiles of covariates and response, respectively, included in the reference data set used for model building and control chart limits estimation. In Section 4, we describe with more details how smooth profiles are obtained from the discrete observations acquired by the DAQ system over time, with 2–5 min frequency.

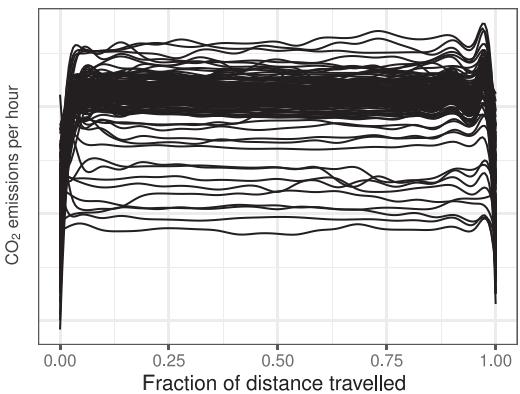


FIGURE 2 Functional response in the reference data set. For confidentiality reasons, y-axis labels are omitted

### 3 | METHODOLOGY

The FRCC is a general framework for profile monitoring that can be divided into three main steps.<sup>8</sup> First, (i) define an MFLR model which links the functional response variable  $\tilde{Y}$ , defined on the compact domain  $\mathcal{T}$  and a vector  $\tilde{\mathbf{X}}$  of random functional covariates  $\tilde{X}_1, \dots, \tilde{X}_p$ , defined on the compact domain  $S$ . Secondly, (ii) define the estimation method of the chosen model and thirdly, (iii) define the monitoring strategy of the functional residual, which is defined as the difference between the fitted value and  $\tilde{Y}$ . In what follows, we assume that  $\tilde{X}_1, \dots, \tilde{X}_p$  and  $\tilde{Y}$  have smooth realizations in  $L^2(S)$  and  $L^2(\mathcal{T})$ , that is, the Hilbert spaces of square integrable functions defined on  $S, \mathcal{T} \subset \mathbb{R}$ .

A specific implementation of the FRCC can be obtained by assuming for the step (i) the following MFLR model:

$$Y(t) = \int_S (\boldsymbol{\beta}(s, t))^T \mathbf{X}(s) ds + \varepsilon(t) \quad t \in \mathcal{T}, \quad (1)$$

where  $Y$  and  $\mathbf{X}$  are the standardized versions of  $\tilde{Y}$  and  $\tilde{\mathbf{X}}$ , obtained through the transformation approach of Chiou et al.<sup>13</sup> The regression coefficient  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ , is a vector where  $\beta_i$ 's are square integrable bivariate functions defined on the closed interval  $S \times \mathcal{T}$ , and the random error function  $\varepsilon$  has zero mean and variance function  $v_\varepsilon^2$ , and is independent of  $\mathbf{X}$ .

For the step (ii), we use an estimation method based on the multivariate Karhunen–Loëve's Theorem.<sup>30</sup> In particular, we assume that the standardized covariate and response variables can be represented as follows:

$$\mathbf{X} = \sum_{i=1}^{\infty} \xi_i^X \psi_i^X \quad Y = \sum_{i=1}^{\infty} \xi_i^Y \psi_i^Y, \quad (2)$$

where  $\psi_i^X = (\psi_{i1}^X, \dots, \psi_{ip}^X)^T$  and  $\psi_i^Y$  are *principal components* (PCs). PCs are defined as the eigenfunctions of the covariance operator of the standardized covariates and response variable corresponding to the eigenvalues  $\lambda_i^X$  and  $\lambda_i^Y$  in descending order, respectively. The scores  $\xi_i^X = \sum_{j=1}^p \int_S X_j(s) \psi_{ij}^X(s) ds$  and  $\xi_i^Y = \int_{\mathcal{T}} Y(t) \psi_i^Y(t) dt$  are such that  $E(\xi_i^X) = 0$ ,  $E(\xi_i^X \xi_j^X) = \lambda_i^X \delta_{ij}$  and  $E(\xi_i^Y) = 0$ ,  $E(\xi_i^Y \xi_j^Y) = \lambda_i^Y \delta_{ij}$ , with  $\delta_{ij}$  the Kronecker delta. As demonstrated in Chiou et al.,<sup>31</sup> the regression coefficient can be then expressed as follows:

$$\boldsymbol{\beta}(s, t) = \sum_{i,j=1}^{\infty} \frac{E(\xi_i^X \xi_j^Y)}{\lambda_i^X} \psi_i^X(s) \psi_j^Y(t) \quad s \in S, t \in \mathcal{T}. \quad (3)$$

An estimator of  $\beta(s, t)$  can be readily obtained by considering the truncated version of Equation (3), that is

$$\boldsymbol{\beta}_{LM}(s, t) = \sum_{i=1}^L \sum_{j=1}^M b_{ij} \psi_i^X(s) \psi_j^Y(t) \quad s \in S, t \in \mathcal{T}, \quad (4)$$

with  $b_{ij} = E(\xi_i^X \xi_j^Y)/\lambda_i^X$ , and  $L, M < \infty$ . Plugging Equation (4) into Equation (1), due to the orthonormality of the PCs  $\psi_i^X$  and  $\psi_i^Y$ , we obtain

$$\xi_M^Y = (\mathbf{B}_{LM})^T \xi_L^X + \epsilon_M, \quad (5)$$

where  $\xi_M^Y = (\xi_1^Y, \dots, \xi_M^Y)^T$ ,  $\xi_L^X = (\xi_1^X, \dots, \xi_L^X)^T$ ,  $\epsilon_M = (\epsilon_1, \dots, \epsilon_M)^T$  and  $\mathbf{B}_{LM} = \{b_{ij}\}_{i=1, \dots, L, j=1, \dots, M}$ , with  $\epsilon_i = \int_{\mathcal{T}} \varepsilon(t) \psi_i^Y(t) dt$ . Therefore, the problem of estimating  $\beta$  reduces to the estimation of the matrix  $\mathbf{B}_{LM}$ , which can be obtained through least squares given  $n$  independent realizations  $(\tilde{\mathbf{X}}_i, \tilde{Y}_i)$  of  $(\tilde{\mathbf{X}}, \tilde{Y})$ . Then, given the least-squares estimator  $\hat{\mathbf{B}}_{LM}$  of  $\mathbf{B}_{LM}$ , the estimator  $\hat{\beta}_{LM}$  of  $\beta$  can be calculated as

$$\hat{\beta}_{LM}(s, t) = (\hat{\psi}^Y(t))^T (\hat{\mathbf{B}}_{LM})^T \hat{\Psi}^X(s) \quad s \in S, t \in \mathcal{T}, \quad (6)$$

where  $\hat{\psi}^Y = (\hat{\psi}_1^Y, \dots, \hat{\psi}_M^Y)^T$ ,  $\hat{\Psi}^X = (\hat{\psi}_1^X, \dots, \hat{\psi}_L^X)^T$ , with  $\hat{\psi}_i^Y$  and  $\hat{\psi}_i^X = (\hat{\psi}_{i1}^X, \dots, \hat{\psi}_{ip}^X)^T$  estimators of  $\psi_i^Y$  and  $\psi_i^X$ , respectively. Finally, an estimator  $\hat{Y}_{LM}$  of  $Y$  is

$$\hat{Y}_{LM} = \sum_{i=1}^L \sum_{j=1}^M \hat{b}_{ij} \xi_i^X \hat{\psi}_j^Y, \quad (7)$$

where  $\hat{b}_{ij}$  are the entries of  $\hat{\mathbf{B}}_{LM}$  and  $\xi_i^X = \sum_{j=1}^p \int_S X_j(s) \hat{\psi}_{ij}^X(s) ds$ .

For the step (iii), we can define the raw functional residual as

$$e(t) = Y(t) - \hat{Y}_{LM}(t) \quad t \in \mathcal{T}. \quad (8)$$

However, by following the remarks in Centofanti et al,<sup>8</sup> we shall better consider a scaled version of it, named *studentized functional residual*, and defined as

$$e_{stu}(t) = \frac{Y(t) - \hat{Y}_{LM}(t)}{\text{Cov}_Y(Y - \hat{Y}_{LM})^{1/2}(t)} \quad t \in \mathcal{T}. \quad (9)$$

The residual variance function is estimated as  $\widehat{\text{Cov}}_Y(Y - \hat{Y}_{LM})(t) = \hat{v}_\varepsilon^2(t) + \hat{\omega}_{LM}(t, t)$ , for  $t \in \mathcal{T}$ , where  $\hat{v}_\varepsilon^2$  is an estimator of  $v_\varepsilon^2$  and  $\hat{\omega}_{LM}$  is defined as

$$\hat{\omega}_{LM}(s, t) = \text{Cov}(\hat{Y}_{LM})(s, t) = \left( \hat{\xi}_L^X \right)^T (\hat{\Xi}_X^T \hat{\Xi}_X)^{-1} \hat{\xi}_L^X \hat{\psi}_M^Y(s)^T \hat{\Sigma}_{\epsilon_M} \hat{\psi}_M^Y(t) \quad s \in S, t \in \mathcal{T}, \quad (10)$$

where  $\hat{\xi}_L^X$  is the estimator of the score vector  $\xi_L^X$  of  $\mathbf{X}$ ,  $\hat{\Xi}_X^T \hat{\Xi}_X$  is the estimator of  $n \text{Cov}(\xi_L^X, \xi_L^X)$ ,  $\hat{\psi}_M^Y$  is the estimator of the vector of the first  $M$  eigenfunctions of  $Y$ , and  $\hat{\Sigma}_{\epsilon_M}$  is the estimator of  $\text{Cov}(\epsilon_M)$ . As stated in Ref. [8], the use of the studentized functional residual, in place of the raw residual, is needed to reduce the effect of covariate mean shifts on the performance of the FRCC to identify OC condition of the quality characteristic, which can be large especially when the coefficient function is poorly estimated. In this case, it can be demonstrated that the interpretation of the FRCC becomes cumbersome, indeed a point falling outside the FRCC control limits could be wrongly assigned to a mean shift in the quality characteristic even though it shows a perfectly reasonable behaviour given the value of the covariates. This is problematic because the aim of the FRCC is to monitor the quality characteristic assigned the value of the covariates, and, thus, if a mean shift in the covariates causes an OC, the chart is saying that the value of quality characteristic disagrees with those of the covariates that, of course, it is not true. In this respect, the studentized functional residual is less influenced by covariate mean shifts than the raw residual. Indeed, the aim of  $\text{Cov}_Y(Y - \hat{Y}_{LM})^{1/2}$  is to weight the raw residual on the basis of its uncertainty, such that for an extreme realization of  $\mathbf{X}$ , the residual is heavily scaled. Thus, the probability of the quality characteristic to be judged IC, when no special causes apply to the process, is larger than that corresponding to the raw residual. Note that, for a large sample sizes, consistently with the data set complexity, the studentized functional residual leads to the same results of the raw residual, because in this case it is independent of the values achieved by the covariates. Therefore, the use of studentized residual allows controlling the false alarm rate in presence of covariate mean

shifts, which comes at a cost of reducing the power of the FRCC in identifying OCs in the quality characteristic, especially for extreme values of the covariates. However, this behaviour is inevitable because comes from the greater uncertainty in the model estimation originated by a limited number of extreme realizations in the reference sample.

We use a monitoring strategy based on the Hotelling's  $T^2$  and the  $SPE$  control charts<sup>4, 6 27, 32</sup> applied to  $e_{stu}$  in Equation (9). In particular, the studentized functional residual  $e_{stu}$  is approximated as

$$e_{stu,K} = \sum_{i=1}^K \xi_i^e \psi_i^e, \quad (11)$$

where the scores  $\xi_i^e = \int_{\mathcal{T}} e(t) \psi_i^e(t) dt$  and the PCs  $\psi_i^e$  are the eigenfunctions corresponding to the eigenvalues  $\lambda_i^e$  in descending order of the covariance function of  $e_{stu}$ . The Hotelling's statistic  $T^2$  is obtained as follows:

$$T_e^2 = \xi^e \Sigma_{\xi^e}^{-1} \xi^e, \quad (12)$$

where  $\Sigma_{\xi^e} = \text{diag}(\lambda_1^e, \dots, \lambda_K^e)$  is the variance–covariance matrix of  $\xi^e = (\xi_1^e, \dots, \xi_K^e)^T$ . Note that  $T_e^2$  is the squared distance of the projection of  $e_{stu}$  from the origin of the space spanned by the PCs standardized for the score variances. Analogously, changes along directions orthogonal to the latter space are monitored by the statistic

$$SPE_e = \int_{\mathcal{T}} (e_{stu}(t) - e_{stu,K}(t))^2 dt. \quad (13)$$

The control charts are designed in Phase I by means of a set of  $n$  functional studentized residuals  $e_{stu,i}, i = 1, \dots, n$ , obtained by  $n$  independent realizations  $(\tilde{\mathbf{X}}_i, \tilde{Y}_i)$  acquired under IC conditions. Phase I includes also the estimation of the MFLR model unknown parameters, the PCs  $\psi_i^e$  and the matrix  $\Sigma_{\xi^e}$  (calculated by means of the sample covariance) as well as the estimation of the control limits for both the Hotelling's  $T^2$  and the  $SPE$  control charts. The latter can be obtained by means of the  $(1 - \alpha)$ -quantiles of the empirical distribution of the two statistics, where  $\alpha$  is chosen to control the overall type I error probability. In the monitoring phase (Phase II), the functional studentized residuals of new data are calculated and an alarm signal is issued if at least one of the corresponding  $T_e^2$  and  $SPE_e$  statistics violates the control limits.

## 4 | RESULTS AND DISCUSSION

In this section, we show the results of the application of the FRCC to the data set described in Section 2.2. In particular, the retrospective and prospective phases, that is, Phase I and Phase II, are described in Section 4.1 and Section 4.2, respectively. Moreover, in Section 4.3, a comparison with simpler monitoring approaches is shown.

We have used the R package `funcharts` to build the FRCC and perform all the analysis shown in this paper. The package is available on CRAN at <https://cran.r-project.org/web/packages/funcharts/index.html>. Moreover, in the [supplementary material](#) we provide an R script and the data to reproduce the results in this paper. Note that the data have been scaled for confidentiality reason.

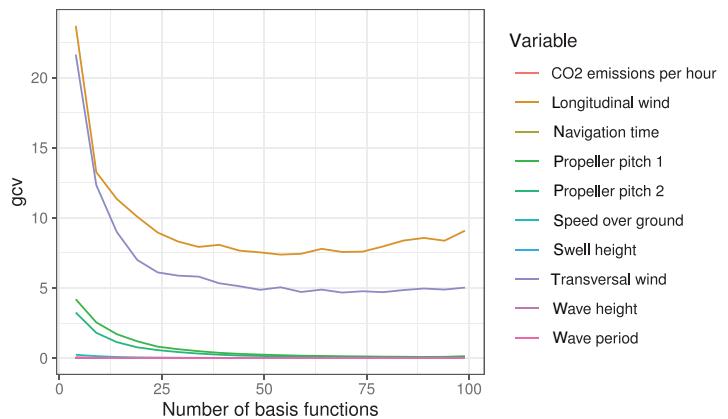
### 4.1 | Phase I

Phase I comprises the recovery of smooth functional data from the discrete observations for each voyage (Section 4.1.1), the identification of the reference data set of IC voyages (Section 4.1.2) and, the estimation of the MFLR model as described in Section 3 (Section 4.1.3).

#### 4.1.1 | Data smoothing

The first step of the analysis is to get smooth functional data from the discrete observations for each voyage of the ship. We use B-spline basis expansion and penalized least squares to estimate the corresponding basis coefficients. A common

**FIGURE 3** Median, across all voyages in the Phase I data set, of the generalized cross-validation (GCV) error as a function of the number of basis functions, for each functional variable



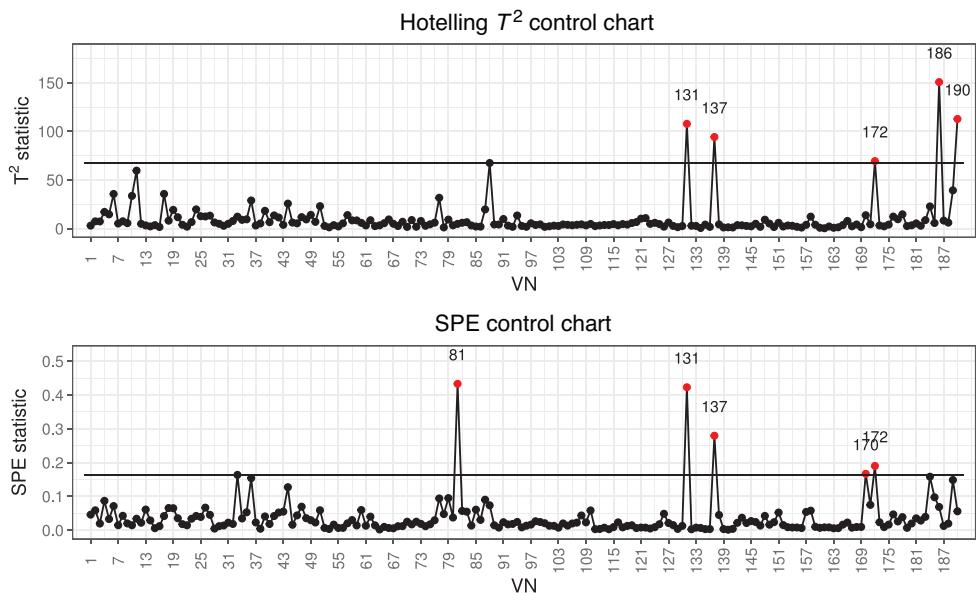
approach is to set a quite large number of basis functions and then select the optimal smoothing parameter by minimizing the generalized cross-validation (GCV) error.<sup>1</sup> However, the number of available discrete points is above 200 for each voyage and, by following this approach, the GCV criterion leads to choosing the smoothing parameter equal to zero in practice for all functional variables. This is a typical problem of overfitting, as also pointed out by Reiss and Ogden,<sup>33</sup> which show that at finite sample sizes GCV is likely to develop multiple minima and undersmooth. Therefore, we encourage parsimony and achieve regularization by choosing a small, efficient number of basis functions, with the smoothing parameter fixed to a small positive value (i.e.  $10^{-10}$ ) to ensure identifiability. In Figure 3, we plot the GCV error against the number of B-spline basis functions. While increasing the number of basis functions reduces the GCV error, we select 25 basis functions for all functional variables as the elbow point of these curves.

#### 4.1.2 | Reference data set

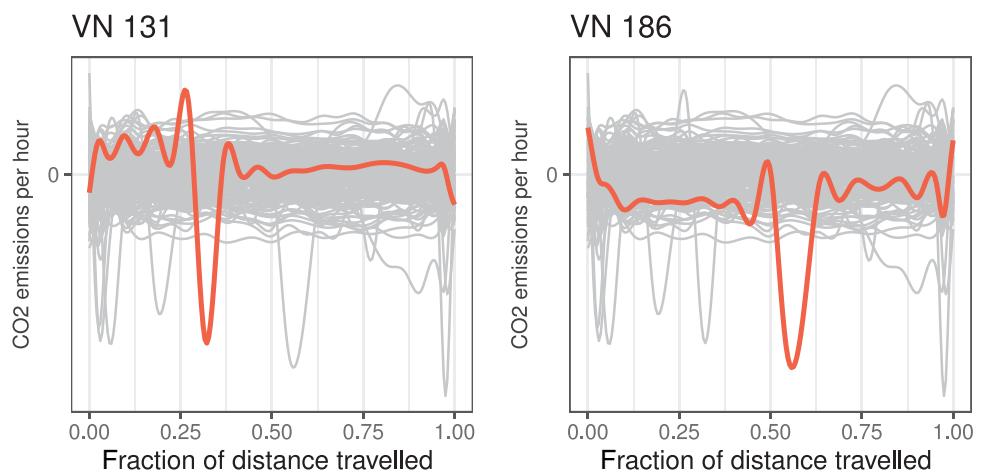
Once functional data are obtained, in order to monitor a set of voyages in a considered period, it is necessary to identify a reference data set of previous IC voyages that can be used for model building and estimation of the Hotelling  $T^2$  and SPE control chart limits. Starting from an historical data set, any voyage that is not representative of the IC conditions has to be removed from it. Specific Phase I techniques are designed for the problem of eliminating anomalous voyages from the historical data set and generally lead to the construction of control charts with different limits from the ones calculated in Phase II. However, the problem of selecting the best method to perform Phase I is beyond the scope of this paper, which is instead mainly focused on Phase II monitoring. In this paper, we use the same FRCC framework also in Phase I and based on experts' opinion, to establish which voyages are actually considered anomalous and have to be excluded. That is, we first use the historical data set to build the FRCC and estimate the control chart limits, and we plot the FRCC for the same voyages; then, voyages signalled as anomalous are carefully investigated by maritime engineers to understand if special causes occurred; in these cases, voyages are removed from the data set; the process is repeated until the final reference data set is obtained, which contains only voyages considered as IC. Starting from the initial historical data set of 190 voyages, at the end of this iterative process of identifying outliers, detecting anomalies, removing anomalous voyages and refitting the model, we get a reference data set of 169 voyages. Figure 4 shows the FRCC applied to the initial data set. The x-axis label indicates the voyage number (VN) and is here intended to be a progressive counting label to denote subsequent voyages in the data set. Moreover, Figure 5 shows some OC studentized residual profiles correctly signalled as anomalous.

#### 4.1.3 | Model building

The FRCC relies on the choice of  $L$  and  $M$  in Equation (4), as well as  $K$  in Equation (11). Figure 6 shows the cumulative fraction of variance explained by the functional PCs in the multivariate functional covariates, the functional response and the functional studentized residuals, respectively. Based on the results in Figure 6, we opt for a more parsimonious choice than that suggested in Centofanti et al<sup>8</sup> and set the thresholds for the explained variability in the data as 80%, 95% and



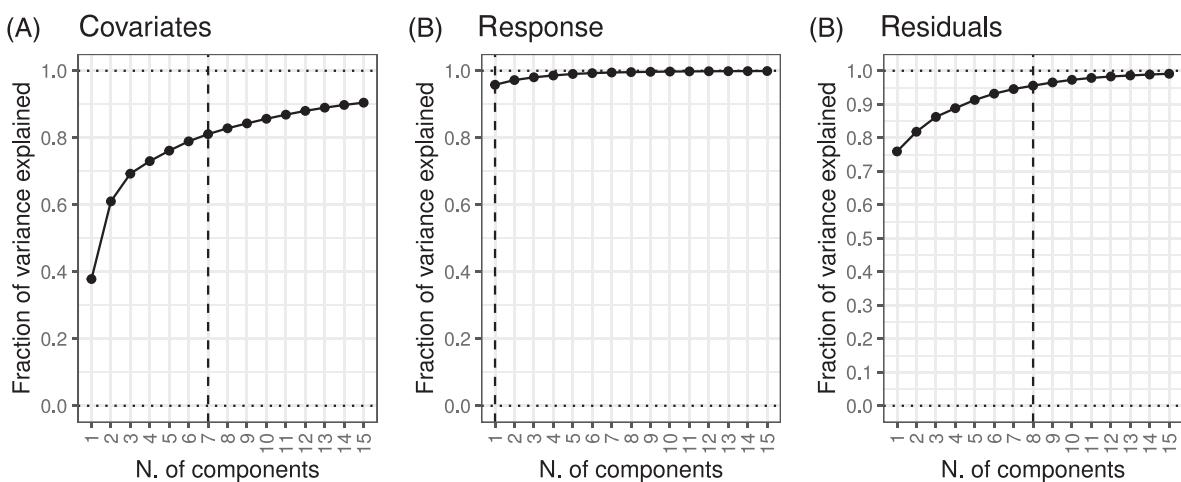
**FIGURE 4** FRCC used for Phase I monitoring on the initial data set of 190 voyages, to remove outliers and define the reference data set



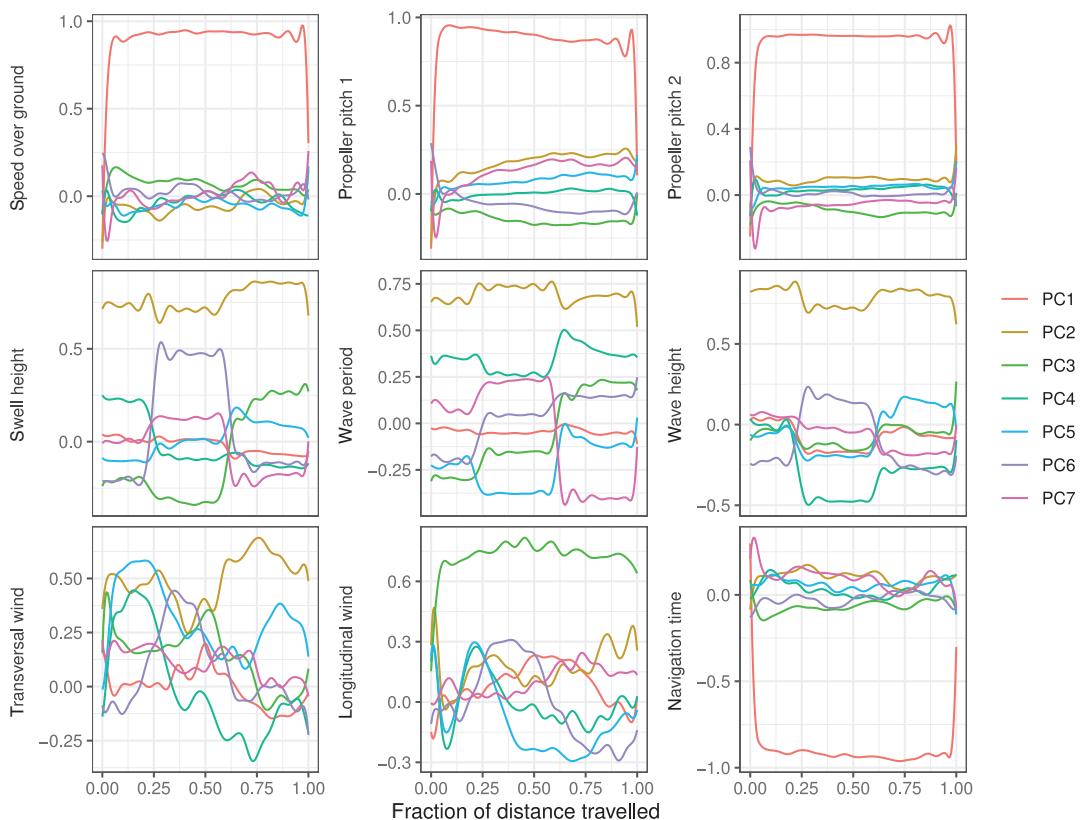
**FIGURE 5** Some functional studentized residuals identified as OC in Phase I, plotted in red against all the other ones in the original Phase I data set, plotted in grey

95%, respectively, that is, we select  $L = 7$ ,  $M = 1$  and  $K = 8$ . The corresponding actual fractions of variance explained are 81%, 96% and 96%.

To allow for a possible interpretation of the selected functional PCs, in Figure 7 we plot the eigenfunctions of the covariance operator of the standardized multivariate functional covariates. Since they all have unit norm, we multiply them by the square root of the corresponding eigenvalues so that profiles with larger norm are PCs that explain a larger fraction of the total variance in the data. It can be readily envisaged that the first PC depends almost entirely on the two propeller pitch variables, the SOG and the navigation time. The latter is negatively correlated with the other variables, and for all these variables, weights are almost constant over the entire functional domain. The second PC strongly depends on the sea state descriptors (swell height, wave height and wave period) and only on the transversal component of the wind. These functional variables all have positive weight, with some parts of the domain showing slightly larger weight than others. The third PC seems to mainly depend on the longitudinal component of the wind alone. In other words, the first PC describes how fast the ship is moving, while the second and third PCs capture two distinct aspects of environmental conditions. Figure 8A shows the first PC (i.e. eigenfunction) of the covariance operator of the standardized functional response, that alone explains most of the variability in the data. This highlights that, after standardization, all



**FIGURE 6** Eigenvalues of the covariance operator of the functional covariates (A), the functional response (B) and the functional studentized residuals (C), estimated on the reference data set. Vertical dashed lines indicate the selected number of components



**FIGURE 7** Eigenfunctions of the covariance operator of the functional covariates, estimated on the reference data set

fuel consumption per hour profiles are mostly constant over the entire domain, apart from the beginning and the end of the voyage. Figure 8B shows the eigenfunctions of the covariance operator of the studentized residuals. The first PC depends on the average value over the entire voyage. The second PC accounts for the difference between the first and the second half of the voyage. Whereas, some of the following PCs seem to assign a larger weight to the boundaries of the functional domain, but, however, their interpretation becomes less straightforward.

Figure 9 shows the estimated functional coefficients obtained as in Equation (6). Since the functional response is approximated with a single functional PC, which, in practice, is constant over the entire domain, the functional coefficient shows only vertical bands along the direction of  $t$ . The most important predictors are those associated with the first and third

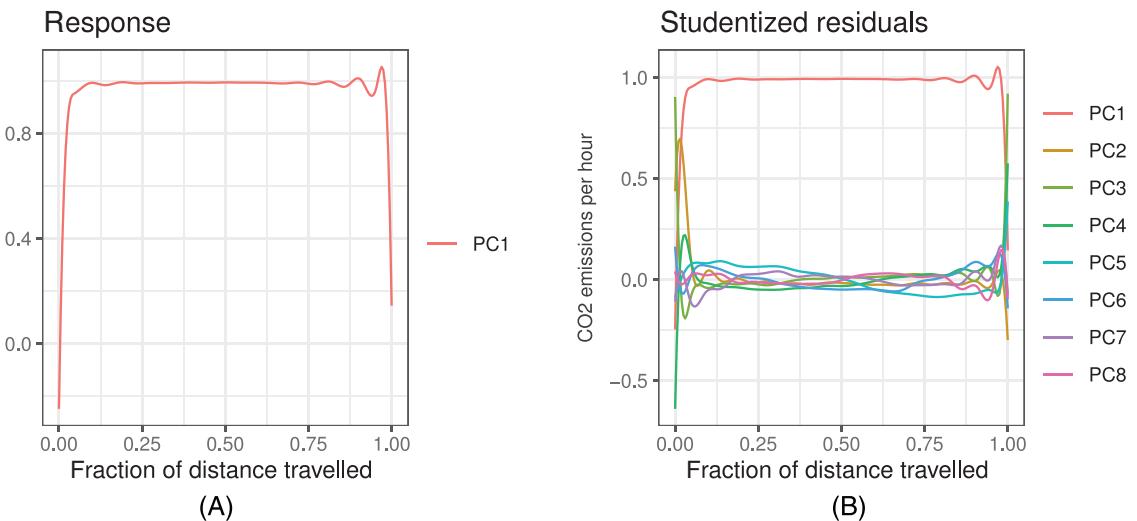


FIGURE 8 Eigenfunctions of the covariance operator of the functional response (A) and the studentized residuals (B)

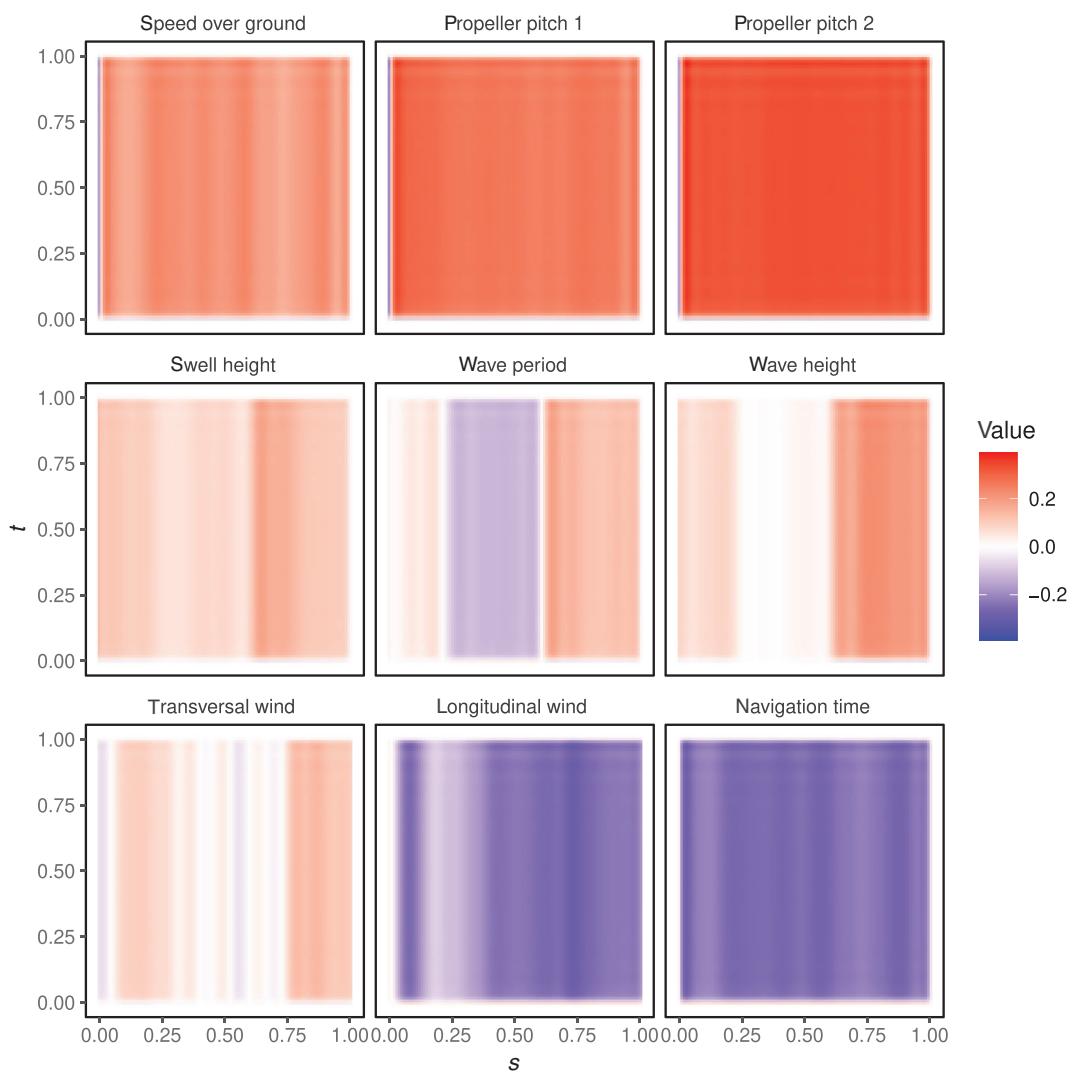
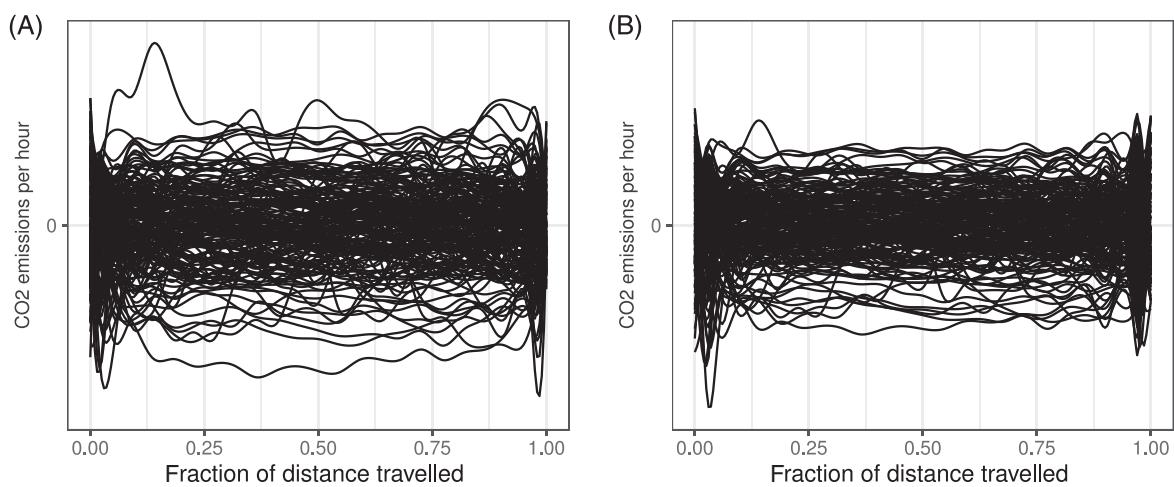


FIGURE 9 Estimate of the regression coefficient obtained by MFLR as in Equation (6), based on the reference data set



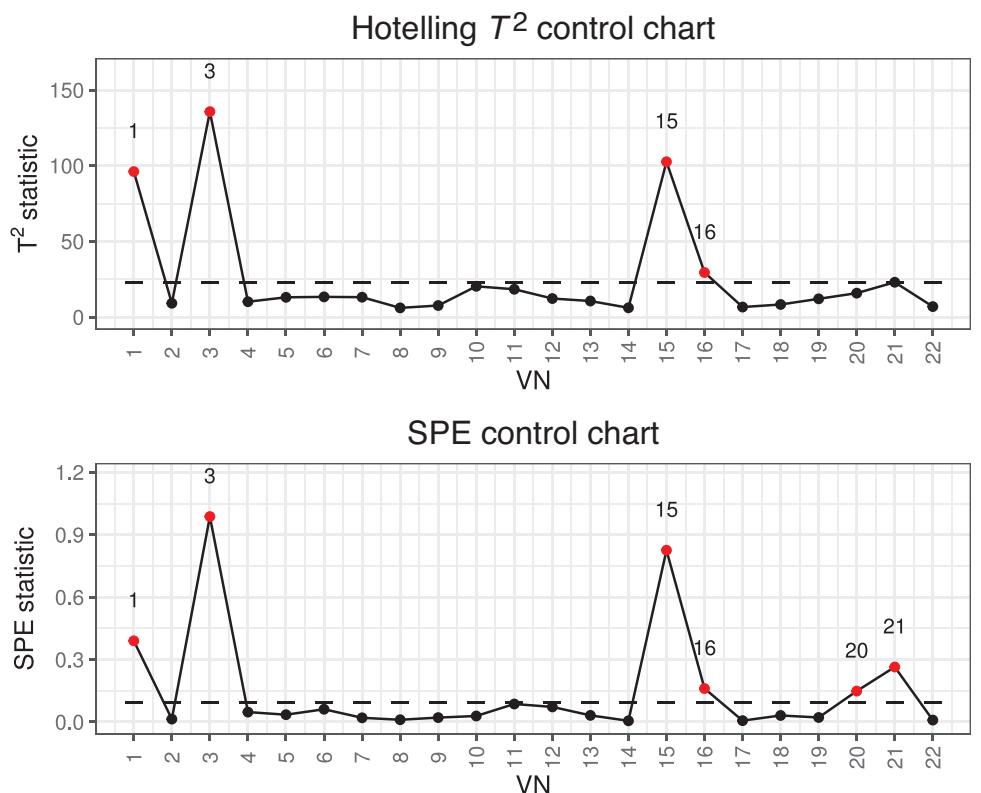
**FIGURE 10** (A) Raw versus (B) studentized functional residuals in the reference data set

PCs in the functional covariates, that is, the two propeller pitch variables, the SOG, with positive coefficients and the navigation time and longitudinal wind component, with negative coefficients. The other environmental variables, which are strongly associated to the second PC in the functional covariates, seem to have a relatively lower weight in the regression. We can then state that voyages with larger value of the propeller pitch and SOG variables have a larger expected value of the functional fuel consumption, whereas those with larger values of the navigation time and longitudinal wind component have lower expected fuel consumption.

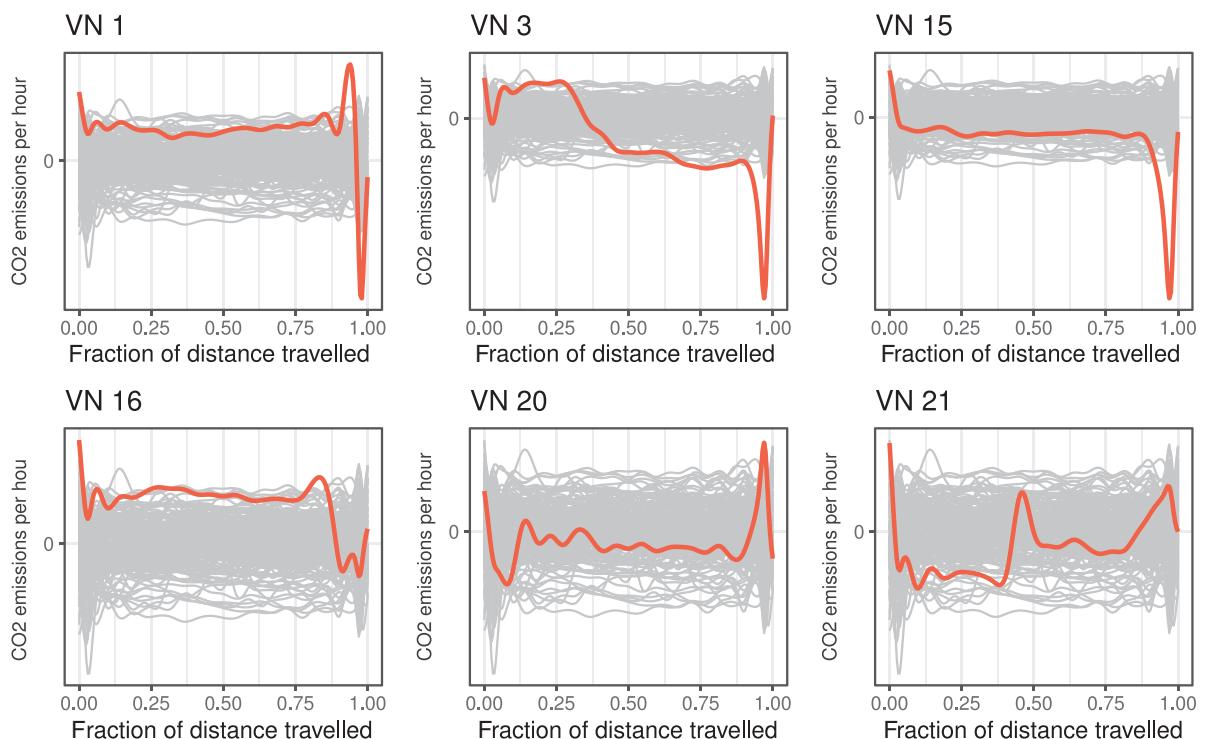
A final consideration on the obtained MFLR model regards the choice (iii) (see Section 3), where studentized functional residuals are used instead of the basic raw residuals. Figure 10 shows the effect of this choice, where some of the more extreme raw functional residuals, shown in Figure 10A, are properly attenuated by the studentization (Figure 10B), when, as in this case, they correspond to more extreme profiles of covariates as described in Section 3.

## 4.2 | Phase II

Figure 11 shows the FRCC used in Phase II, that is, the actual monitoring phase. The points correspond to 22 subsequent voyages, each of them is denoted by its VN. For simplicity, we count and label voyages again with VN 1 through 22, even though these voyages must not be confused with the first 22 ones in Phase I in Figure 4. Some Phase II voyages are signalled as OC in Figure 11. In particular voyages 1, 3, 15 and 16 are OC in both Hotelling  $T^2$  and SPE control charts, while voyages 20 and 21 are OC in the SPE control chart only. OC voyages are generally characterized by some unexpected behaviour in the  $\text{CO}_2$  emissions, because they have not been properly predicted by the MFLR model in some specific part of the domain, or because the prediction error was overall moderately large for a considerable part of the voyage. Functional studentized residuals for these OC voyages are plotted in Figure 12 against the studentized residuals in the reference data set. Voyages 1, 3 and 15 plot far above the upper control limits in both the Hotelling's  $T^2$  and SPE control charts. In particular, the functional studentized residual of voyage 1 is signalled as OC because it is larger than usual on average and at the end of the voyage. Voyage 3 shows an unusual studentized residual that is positive in the first part of the voyage, negative in the second part and extremely large at the very end of the voyage with respect to the Phase I reference profiles (depicted in grey lines). Voyage 15 shows the same unusual negative deviation at the end of the voyage as in voyage 3 and the residual profile remains negative for the whole voyage. Voyages 16, 20 and 21 are signalled as OC in Figure 11, even though they are close to the upper control limits and show less dramatic behaviour, as shown by the plots reported in the second row of panels in Figure 12. Studentized residual profiles corresponding to these voyages have in fact less prominent peaks/valleys. In particular, it is worth noting that voyage 16 is signalled as anomalous plausibly because its residual profile achieves large positive values for almost the entire voyage. Residual profile of voyage 20, even if more regular, as it is near zero in the middle of the voyage, has mild peak and valley at the extremes of the voyage. Residual profile of voyage 21, if considered pointwise, is instead almost entirely in the range of the Phase I studentized residuals. However, it is signalled as OC plausibly because of a sudden profile jump in the middle of the voyage.



**FIGURE 11** FRCC in Phase II. Each point corresponds to a voyage and the values of the Hotelling  $T^2$  and SPE statistics are reported. Horizontal dashes indicate upper control limits. Red points denote OC statistics



**FIGURE 12** Profiles of the Phase II studentized residuals signalled as OC by the FRCC, plotted as red lines against the studentized residuals in the reference data set, plotted in grey in each panel

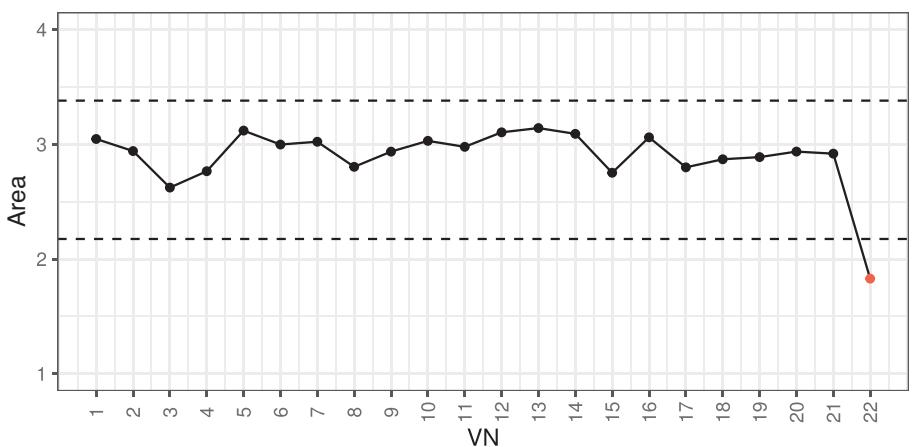


FIGURE 13 INBA control chart

Finally, we want to highlight the importance of using both the Hotelling's  $T^2$  and the  $SPE$  monitoring statistics to detect OC voyages, by discussing their difference. With the FPCA model on the functional residuals, we use the first  $K = 8$  components to approximate the data as in Equation (11), since they explain 95% of their variability (Figure 6C). In particular, note that the first component, which alone explains about 75% of the variability in the data (Figure 6C), in practice looks at shifts in mean during the entire voyage. Then, the Hotelling's  $T^2$  statistic in Equation (12), which is calculated on the scores of the first eight components, monitors the deviation within the FPCA model and from the discussion above we can state that a functional residual with a particularly large shift in mean will have a large  $T^2$ . On the other hand, the  $SPE$  statistic in Equation (13) monitors the remaining part, that is, the error (11) of approximating the residual function with its projection based on FPCA, which should be small when the profile is consistent with the FPCA model. Therefore, a large  $SPE$  value indicates that a profile is clearly different than the underlying FPCA model. In Figure 11, all the voyages signalled have  $SPE$  above the upper control limit, but voyages 20 and 21 have the  $T^2$  IC. These OC voyages show smaller shifts in mean than the others in Figure 12, then it makes sense that the  $T^2$  statistic is IC, however they still show an anomalous behaviour, as the other OC voyages, which is well captured by the  $SPE$  statistic.

#### 4.3 | Comparison with other methods

In this section, we try to show if it is actually convenient to use the FRCC rather than simpler approaches. Centofanti et al<sup>8</sup> showed that the FRCC is more powerful than the index based (INBA) control chart, which monitors the area under the response variable, and the RESP control chart, which monitors the coefficients coming from the functional PC decomposition of the response via Hotelling's  $T^2$  and  $SPE$  control charts. We compare the FRCC with these two control charts and discuss if there are different results in terms of detection of OC voyages in this specific application. Figure 13 shows the INBA control chart, which is not able to detect any of the voyages signalled by the FRCC and seems not appropriate for this type of application. Moreover, this control chart only detects voyage 22 as OC, which is signalled also by the RESP control chart in Figure 14, but, on the other hand, is IC in the FRCC. Apparently, voyage 22 could be an anomalous one that the FRCC is not able to correctly identify. By observing the functional response profile plotted in the left panel of Figure 15, it is plausible that this voyage is signalled as OC by the INBA and RESP control charts because, marginally, the  $\text{CO}_2$  emissions were particularly low during the entire voyage. However, by observing that the corresponding functional studentized residual plotted in the right panel of Figure 15, we can conclude that these low values of the response variable are predicted well by the MFLR model. Therefore, conditionally on the functional covariate profiles, the response variable profile of voyage 22 is correctly judged as IC by the FRCC. This example highlights the convenience in using the FRCC with respect to simpler approaches, when the interest is in monitoring a functional quality of interest conditionally on functional covariates having influence on it. The RESP control chart seems to correctly detect some of the voyages identified by the FRCC, that is, voyages 1, 3, 15 and 16, but it misses voyages (i.e. 20 and 21), while it signals voyage 11 that is IC in the FRCC. Note that voyage 11 is close the upper control limits in both the RESP control chart and the FRCC.

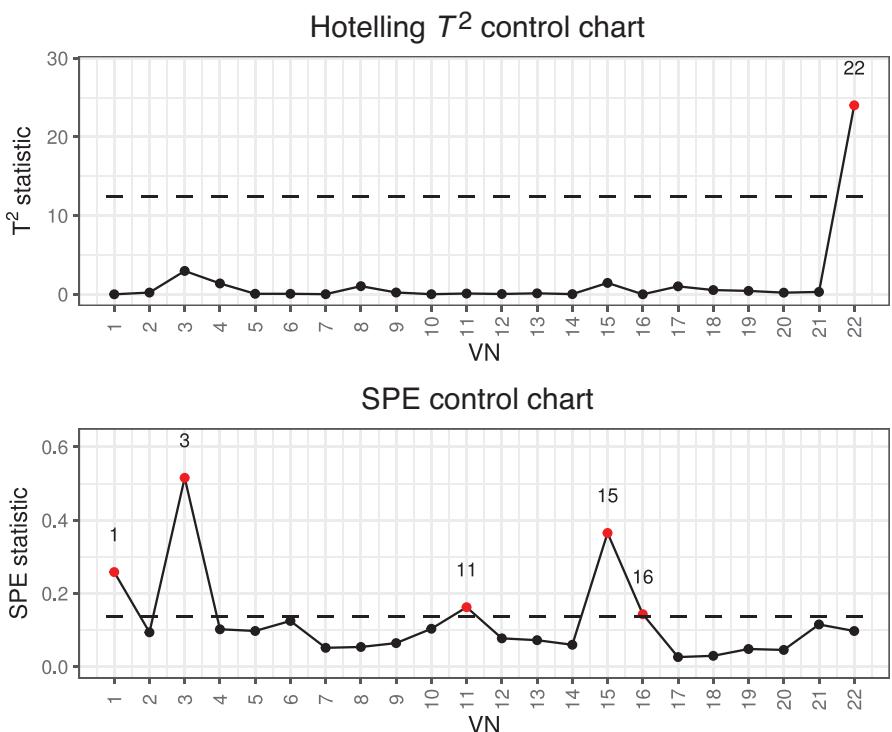


FIGURE 14 RESP control chart

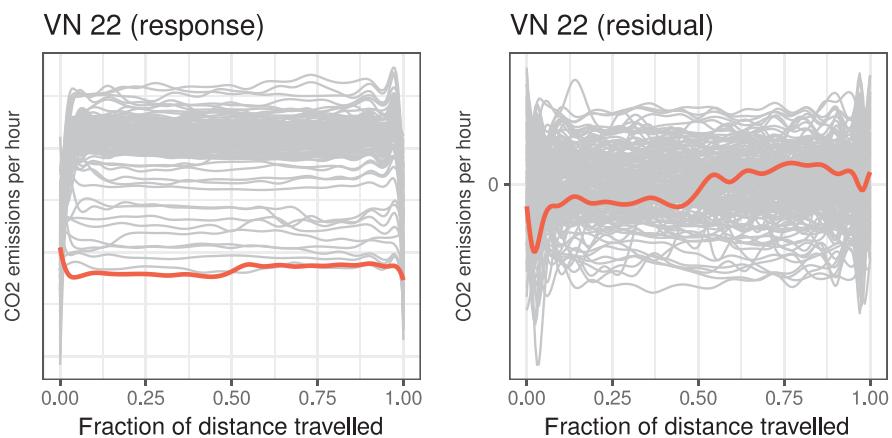


FIGURE 15 Profile of voyage 22 of the Phase II data set (red), plotted against the profile in the reference data set (grey). Both the original functional response and the studentized residual are reported

## 5 | CONCLUSIONS

A particular implementation of the FRCC framework proposed in Centofanti et al<sup>8</sup> is applied in this paper to monitor the ship CO<sub>2</sub> emission profiles, in order to identify any special causes at given values of the functional covariates that may have an influence on them. The CO<sub>2</sub> emission profiles are adjusted for the effect of these covariates by means of an MFLR model estimated via multivariate FPCA. That is, the residuals from the MFLR model are monitored by using jointly the Hotelling's  $T^2$  and the SPE control charts built on their functional PC decomposition. The specific implementation of the FRCC relies on the use of the studentized functional residual to take into account the different residual variance at different covariate values. The proposed FRCC demonstrated to be effective in the identification of anomalous voyages over the real-case study presented, which is concerned with the data collected during 2020 on a Ro-Pax ship operating in the Mediterranean Sea. Moreover, the particular implementation of the FRCC framework was compared against alternative

approaches available in the literature, which, however, only look at the marginal distribution of the functional response, or only to some specific features. All these competing methods showed a lack of ability to signal some important OC and, in other situations, provided false alarms.

Since the monitoring statistics are calculated on the entire profiles acquired at the end of each voyage, this work does not directly allow for a real-time feedback control, that is, to guide actions during an ongoing voyage. Instead, the FRCC can be used to track automatically (and possibly the whole fleet) patterns and trends that may identify malfunctioning in the engines, the need for hull cleaning, or for any other energy efficiency initiative. This information can help to diagnose and prevent possible problems in future voyages, or to schedule maintenance operations.

Finally, one important output achieved in our research is the technological transfer of the FRCC tool to the shipping company Grimaldi Group. The very practical applicability of these statistical tools is in fact further investigated by providing the energy saving department of the company with R code that is able to automatically import new data from the company server, to envelop the mathematical and numerical details provided in the paper, and to routinely produce automatic voyage reports for some Ro-Pax ships of interest from their fleet.

## ACKNOWLEDGEMENTS

The authors are deeply grateful to the anonymous referees and to the editors for their suggestions and help in significantly improving the manuscript. The authors are also extremely grateful to the Grimaldi Group's Energy Saving Department engineers Dario Bocchetti, Andrea D'Ambra and Rosa Di Matteo for the access to observational data, the maritime domain insight and the general support over the course of these activities.

Open Access Funding provided by Universita degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

[Correction added on 11 May 2022, after first online publication: CRUI-CARE funding statement has been added.]

## DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are available in the supplementary material of this article

## ORCID

Christian Capezza  <https://orcid.org/0000-0002-8358-4800>

Fabio Centofanti  <https://orcid.org/0000-0002-5007-525X>

Antonio Lepore  <https://orcid.org/0000-0002-8739-5310>

Biagio Palumbo  <https://orcid.org/0000-0003-1036-8127>

Simone Vantini  <https://orcid.org/0000-0001-8255-5306>

## REFERENCES

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. Wiley Online Library; 2005.
2. Horváth L, Kokoszka P. *Inference for Functional Data with Applications*. Springer Science & Business Media; 2012.
3. Kokoszka P, Reimherr M. *Introduction to Functional Data Analysis*. CRC Press; 2017.
4. Noorossana R, Saghaei A, Amiri A. *Statistical Analysis of Profile Monitoring*. Vol. 865. Wiley; 2012.
5. Colosimo BM, Pacella M. A comparison study of control charts for statistical monitoring of functional data. *Int J Prod Res*. 2010;48(6):1575-1601.
6. Grasso M, Menafoglio A, Colosimo BM, Secchi P. Using curve-registration information for profile monitoring. *J Qual Technol*. 2016;48(2):99-127.
7. Menafoglio A, Grasso M, Secchi P, Colosimo BM. Profile monitoring of probability density functions via simplicial functional PCA with application to image data. *Technometrics*. 2018;60(4):497-510.
8. Centofanti F, Lepore A, Menafoglio A, Palumbo B, Vantini S. Functional regression control chart. *Technometrics*. 2020.
9. Mandel B. The regression control chart. *J Qual Technol*. 1969;1(1):1-9.
10. Palumbo B, Centofanti F, Del Re F. Function-on-function regression for assessing production quality in industrial manufacturing. *Qual Reliab Eng Int*. 2020;36(8):2738-2753.
11. Centofanti F, Lepore A, Menafoglio A, Palumbo B, Vantini S. Adaptive smoothing spline estimator for the function-on-function linear regression model. *arXiv preprint arXiv:2011.12036* 2020.
12. Centofanti F, Fontana M, Lepore A, Vantini S. Smooth Lasso estimator for the function-on-function linear regression model. *arXiv preprint arXiv:2007.00529* 2020.
13. Chiou JM, Chen YT, Yang YF. Multivariate functional principal component analysis: a normalization approach. *Stat Sin*. 2014;24:1571-1596.
14. Montgomery DC, Johnson LA, Gardiner JS. *Forecasting and Time Series Analysis*. McGraw-Hill Companies; 1990.
15. Box GE, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*. Wiley; 2015.
16. Montgomery DC, Mastrangelo CM. Some statistical process control methods for autocorrelated data. *J Qual Technol*. 1991;23(3):179-193.

17. Montgomery DC. *Introduction to Statistical Quality Control*. Wiley; 2007.
18. Capezza C, Lepore A, Menafoglio A, Palumbo B, Vantini S. Control charts for monitoring ship operating conditions and CO<sub>2</sub> emissions based on scalar-on-function regression. *Appl Stoch Models Bus Ind*. 2020;36(3):477-500.
19. Lepore A, Palumbo B, Capezza C. Analysis of profiles for monitoring of modern ship performance via partial least squares methods. *Qual Reliab Eng Int*. 2018;34(7):1424-1436.
20. IMO. Air Pollution and Greenhouse Gas (GHG) Emissions from International Shipping, MARPOL Annex 6. London, UK; 2012.
21. IMO. Guidelines on the method of calculation of the attained Energy Efficiency Design Index (EEDI) for new ships, MEPC.212 Annex 8. London, UK; 2012.
22. IMO. 2014 Guidelines on survey and certification of the Energy Efficiency Design Index (EEDI). London, UK; 2014.
23. Lepore A, Reis MS, Palumbo B, Rendall R, Capezza C. A comparison of advanced regression techniques for predicting ship CO<sub>2</sub> emissions. *Qual Reliab Eng Int*. 2017;33(6):1281-1292.
24. Erto P, Lepore A, Palumbo B, Vitiello L. A procedure for predicting and controlling the ship fuel consumption: Its implementation and test. *Qual Reliab Eng Int*. 2015;31(7):1177-1184.
25. Bocchetti D, Lepore A, Palumbo B, Vitiello L. A statistical approach to ship fuel consumption monitoring. *J Sh Res*. 2015;59(3):162-171.
26. Happ C. MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains. R Package Version 1.1. 2017.
27. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *J Qual Technol*. 2004;36(3):309-320.
28. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
29. Lackenby H. *ITTC Dictionary of Ship Hydrodynamics*. Royal Institution of Naval Architects; 1978.
30. Happ C, Greven S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J Am Stat Assoc*. 2018;113(522):649-659.
31. Chiou JM, Yang YF, Chen YT. Multivariate functional linear regression and prediction. *J Multivar Anal*. 2016;146:301-312.
32. Pini A, Vantini S, Colosimo BM, Grasso M. Domain-selective functional analysis of variance for supervised statistical profile monitoring of signal data. *J R Stat Soc Series C Appl Stat*. 2018;67(1):55-81.
33. Reiss PT, Ogden RT. Smoothing parameter selection for a class of semiparametric linear models. *J R Stat Soc Series B Stat Methodol*. 2009;71(2):505-523.

## AUTHOR BIOGRAPHIES

**Christian Capezza** is a postdoc researcher at the Department of Industrial Engineering of the University of Naples Federico II. He works on advanced statistical methodologies for engineering applications and his research project regards the development of interpretable statistical methods for the analysis of complex systems in industry 4.0.

**Fabio Centofanti** is a PhD student at the Department of Industrial Engineering of the University of Naples Federico II, Italy. His main research interests include functional data analysis and statistical process monitoring for industrial applications.

**Antonio Lepore** is an Assistant Professor at the Department of Industrial Engineering of the University of Naples Federico II, Italy. His main research interests include the industrial application of statistical techniques to the monitoring of complex measurement profiles from multisensor acquisition systems, with particular attention to renewable energy and harmful emissions.

**Alessandra Menafoglio** is an Assistant Professor at MOX, Department of Mathematics, Politecnico di Milano. Her research interests focus on the development of innovative statistical models and methods for the analysis and statistical process control of complex observations (e.g. curves, images and functional signals), possibly characterized by spatial dependence.

**Biagio Palumbo** is an Associate Professor in ‘Statistics for experimental and technological research’ at the Department of Industrial Engineering of the University of Naples Federico II, Italy. His major research interests include reliability, design and analysis of experiments, statistical methods for process monitoring and optimization and data science for technology.

**Simone Vantini** is an Associate Professor of Statistics at the Politecnico di Milano, Italy. He has been publishing widely in Functional and Object-Oriented Data Analysis. His current research interests include: permutation testing, non-parametric forecasting, process control, non-Euclidean data and in general statistical methods and applications motivated by business or industrial problems.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Capezza C, Centofanti F, Lepore A, Menafoglio A, Palumbo B, Vantini S. Functional regression control chart for monitoring ship CO<sub>2</sub> emissions. *Qual Reliab Eng Int.* 2022;38:1519–1537.

<https://doi.org/10.1002/qre.2949>