RESEARCH ARTICLE

Covariate-adjusted functional data analysis for structural health monitoring

Philipp Wittenberg[1] (iD), Lizzie Neumann[1] (iD), Alexander Mendler[2] (iD) and Jan Gertheiss[1] (iD)

[1]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany
[2]Department of Materials Engineering, Technical University of Munich, Munich, Germany
Corresponding author: Philipp Wittenberg; Email: pwitten@hsu-hh.de

Abstract

Structural health monitoring (SHM) is increasingly applied in civil engineering. One of its primary purposes is detecting and assessing changes in structure conditions to increase safety and reduce potential maintenance downtime. Recent advancements, especially in sensor technology, facilitate data measurements, collection, and process automation, leading to large data streams. We propose a function-on-function regression framework for (nonlinear) modeling the sensor data and adjusting for covariate-induced variation. Our approach is particularly suited for long-term monitoring when several months or years of training data are available. It combines highly flexible yet interpretable semi-parametric modeling with functional principal component analysis and uses the corresponding out-of-sample Phase-II scores for monitoring. The method proposed can also be described as a combination of an "input–output" and an "output-only" method.

Impact Statement

Structural health monitoring (SHM) utilizes sensor technology to identify and detect potential damage to critical infrastructure like bridges. However, the available measurements can be affected by external factors such as temperature, requiring data-driven methods to account for these influences. This study presents an interpretable framework based on functional data analysis to effectively utilize the data's structure and characteristics, including daily and seasonal patterns, when adjusting for confounding effects. Many existing methods for removing environmental effects are encompassed as special cases. Additionally, damage-sensitive features are readily accessible and can be monitored using state-of-the-art control charts.

1. Introduction

Structural health monitoring (SHM) employs sensor technologies to collect data from structures such as bridges to detect, localize, or quantify damage (Deraemaeker et al., 2008; Kullaa, 2011; Hu et al., 2012; Magalhães et al., 2012). These field measurements often exhibit missing data and are influenced by

---

environmental and operational factors such as temperature, wind, humidity, or traffic load (Wang et al., 2022). Multiple studies, including Wang et al. (2022), Han et al. (2021) and the references therein, identify temperature as a predominant factor affecting structural stiffness and material properties due to thermal expansion and contraction (Han et al., 2021). Consequently, data-driven methods are essential to take these confounding effects in SHM data analysis into account.

With respect to separating temperature-induced responses from structural responses, it can be distinguished between so-called "input–output" methods and "output-only" approaches (Wang et al., 2022). In the first case, both the sensor measurements and observations of the confounding variables, such as temperature, are considered, while in the latter case, as the name suggests, only the responses of the structures are used, often using projection methods such as principal component analysis (PCA). Among input–output methods, a prevalent approach is regressing sensor measurements on environmental or operational variables, also known under the name *response surface* modeling. Then, following the so-called "subtraction method," the predicted sensor data is subtracted from the actually observed data, and the residuals are used for further analysis. For fitting regression function(s) to the data, various methods have been used in the SHM literature, ranging from simple linear or polynomial regression to advanced machine learning approaches such as artificial neural networks, see Avci et al. (2021). However, often, these methods ignore that output data may exhibit daily or yearly patterns or that error terms are correlated over time, e.g., if estimating unknown parameters through least-squares (Cross et al., 2013; Maes et al., 2022). In other cases, rather restrictive parametric assumptions known from time series analysis are made (Hou and Xia, 2021). Furthermore, input–output and output-only methods are typically considered to be completely different approaches. An overarching framework is still missing where, for instance, it is possible to switch from one approach to the other without changing the downstream monitoring procedure, or the common situation can be handled where measurements are available only for a few confounding variables while others are unobserved or unknown.

In light of this, we will present a novel functional data analysis (FDA) perspective to address these challenges in SHM. FDA has been an area of intensive methodological research over the last two or three decades. For an introduction to FDA in general and an overview of recent developments, the works by Ramsay and Silverman (2005), Wang et al. (2016), and Gertheiss et al. (2024) are recommended. Existing applications of FDA within SHM are reviewed by Momeni and Ebrahimkhanlou (2022). In SHM, so far, FDA has primarily been used for distributional regression and change point detection. For instance, Chen et al. (2020) used FDA to segment data by time and analyze corresponding probability density functions through warping functions and functional principal component analysis (FPCA). Other notable contributions include the works by Chen et al. (2018), Chen et al. (2019), and Chen et al. (2021), who developed methods for imputing missing data using distribution-to-distribution and distribution-to-warping functional regression, and Lei et al. (2023a), Lei et al. (2023c), and Lei et al. (2023b) focused on outlier detection and change-point detection in FDA. Jiang et al. (2021) modeled temperature-induced strain relationships using warping functions and FPCA.

To elucidate our functional perspective, consider Figure 1. Similar plots in SHM studies, such as Xia et al. (2012), Xia et al. (2017), and Zhou et al. (2011), have been used to illustrate the relationship between natural frequencies, strain or displacement, and temperature. However, to the best of our knowledge, this functional nature of the data has never been exploited in SHM when correcting for confounding effects. The data in Figure 1 are from the KW51 railway bridge (Maes and Lombaert, 2020), near Leuven, Belgium, monitored from October 2, 2018 to January 15, 2020, including a retrofitting period from May 15 to September 27, 2019. This dataset contains hourly natural frequencies and steel surface temperature (more details on the bridge will be given in Section 4). The top-right panel of Figure 1 shows the bridge's natural frequency (mode 6) for selected days before retrofitting. It appears plausible to assume that there is some underlying daily pattern common to all profiles, which might be caused by environmental influences that may follow some recurring pattern as well. Those patterns and relationships can be further analyzed through FDA, among other things, by regressing the natural frequency profiles on the temperature curves shown in the bottom-left panel of Figure 1. From the colors, it is already visible that there is a negative association between temperature and the natural frequency, meaning that lower
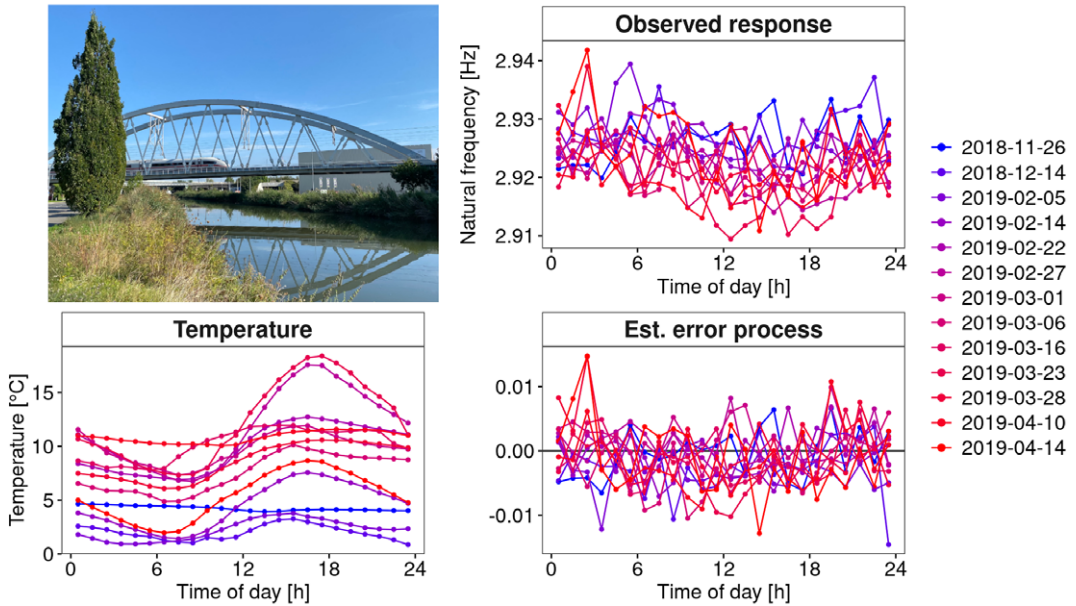
Figure 1. *KW51 Bridge from the south side in September 2024 (top-left). The natural frequency of mode 6 for some selected days before the retrofitting started (top-right) and the corresponding steel temperature curves (bottom-left). If a simple piecewise linear model is used for temperature adjustment, the error profiles (bottom-right) are obtained.*

temperatures lead to a higher frequency; compare, e.g., Xia et al. (2006), Xia et al. (2012). The bottom-right panel shows the resulting error profiles when using a conventional, non-functional piecewise linear model to account for this effect, as done by Maes et al. (2022).

If considering entire daily profiles rather than individual measurements as the sampling instances, this regression task is called "function-on-function" regression. In the statistical process monitoring (SPM) literature, it is part of so-called profile monitoring, see Woodall et al. (2004), Woodall (2007), Maleki et al. (2018) for detailed reviews on that topic. Current state-of-the-art methodologies, such as the (linear) functional regression control chart (FRCC) by Centofanti et al. (2021), have limitations in SHM applications, where the relationship between temperature and natural frequencies is often nonlinear (Han et al., 2021). Additionally, by integrating over the entire domain (i.e., the entire day in Figure 1), the employed functional linear model allows "future observations" to influence the current value of the response (Wittenberg et al., 2024). Specifically, if (material) temperature data is available directly from the structure itself instead of ambient temperature, it appears more plausible to use a so-called concurrent model, where the response data at time point *t* depends only on the steel temperature at time *t*, not the temperature over the entire day. Finally, the assumption of complete datasets, e.g., made in Capezza et al. (2023), is impractical because SHM sensor data often exhibit dropouts due to power or sensor failure. To address these challenges often found in SHM, this paper will consider the very general framework of functional additive mixed models (Scheipl et al., 2015; Greven and Scheipl, 2017) instead of the common functional linear models used by Centofanti et al. (2021) and Capezza et al. (2023). By doing so, this paper also provides methodological novelty concerning profile monitoring in a more general sense. Besides the regression task, the functional perspective on SHM data offers an elegant way of "de-noising" the data and extracting features through functional principal component scores that can be used as inputs to advanced multivariate control charts. The methods presented in this article can be applied directly to sensor measurements (e.g., strains, inclinations, accelerations) or extracted damage-sensitive features (e.g., eigenfrequencies). That is why this paper will use the more general term "system outputs." In summary, using the methods discussed, we can model (i) recurring daily and yearly patterns as well as

(ii) confounding effects flexibly and interpretably. Furthermore, the presented functional approach also accounts for (iii) correlations and (iv) heteroscedasticity within error profiles. Finally, it will (v) extract covariate-adjusted features from the system outputs and (vi) unify all these aspects within a coherent statistical framework called *Covariate-Adjusted Functional Data Analysis for SHM* (CAFDA-SHM). This combined approach includes subtraction methods, PCA-based projections (both often used in SHM), and the FRCC in one universal framework, encompassing multiple aspects that have only been considered partially in previous approaches.

The remainder of the article is structured as follows. In Section 2, our functional data approach is introduced, and our model training and monitoring workflow is presented. The results of a Monte Carlo simulation study validating the developed methods are presented in Section 3. A structural health monitoring application for dynamic response data of the KW51 railway bridge and a second case study for a quasi-static response of a reinforced concrete bridge can be found in Section 4, respectively. Section 5 offers some concluding remarks. Appendix A provides some further modeling options beyond Section 2. For researchers and practitioners who want to apply the newly proposed methods, the code is open source and publicly available at https://github.com/wittenberg/CAFDA-SHM_code.

2. A functional data approach for modeling and monitoring system outputs

2.1. Basic model structure

The model we assume for "in-control" (IC) data has the following basic form. To keep things simple, we first restrict ourselves to a single, functional covariate $z_j(t)$, e.g., denoting the temperature at time $t \in \mathcal{T}, \mathcal{T} = (0\text{h}, 24\text{h}]$, and day j, and a single system output $u_j(t)$. The latter could be a rather raw sensor measurement (yet preprocessed in some sense), such as strain or inclination data, or extracted features, such as natural frequencies. Then, we assume the basic model

$$u_j(t) = \alpha(t) + f(z_j(t)) + E_j(t), \tag{1}$$

where $\alpha(t)$ is a fixed functional intercept, $f(z_j(t))$ is a fixed, potentially nonlinear effect of temperature, and $E_j(t)$ is a day-specific, functional error term with zero mean and a common covariance, i.e., $\mathbb{E}(E_j(t)) = 0$, $\text{Cov}(E_j(s), E_j(t)) = \Sigma(s,t)$, $s, t \in \mathcal{T}$. In the FDA framework used here, sampling instances are days instead of single measurement points, and the daily profiles are considered the quantities of interest. This model has two advantages over scalar-on-scalar(s) regression as typically used for response surface modeling in SHM:

- The functional intercept $\alpha(t)$ captures recurring daily patterns that cannot be explained through the available environmental or operational variables, e.g., because the factors causing them are not recorded/available. In the case of long-term monitoring, we may extend the one-dimensional $\alpha(t)$ to a two-dimensional surface $\alpha(t, d_j)$, where d_j denotes the time/date of the year corresponding to day j in the dataset. The latter accounts for a potential second, i.e., yearly level of periodicity.
- The error term $E_j(t)$ is typically not a white noise process but correlated over time, i.e., in the t-direction. Furthermore, variances may vary over the day. For instance, error variances may be lower at night (e.g., because there is less traffic, no influence by the sun, etc.). In other words, $\Sigma(s,t)$ is not necessarily zero for $s \neq t$, and $\Sigma(t,t)$ is not constant. For illustration, the (estimated) error process for some selected days for the KW51 data is shown in Figure 1 (bottom-right), where a piecewise linear model with one breakpoint was used for temperature adjustment, as suggested in the literature (Moser and Moaveni, 2011; Worden and Cross, 2018; Maes et al., 2022). Apparently, those profiles exhibit some more structure than pure white noise. However, ignoring this correlation when fitting α and f through ordinary least squares or common maximum likelihood, assuming conditional independence between measurements, will typically lead to less accurate estimates. More importantly, if (conditional) independence is assumed but not given, measures of statistical uncertainty, such as confidence and prediction intervals or control limits, will be biased. In SHM, this can be particularly

harmful as these quantities are used to determine if measurements are "out-of-control." For instance, if a memory-based control chart with the assumption of independence is used to detect a mean shift in an error process, as shown in Figure 1, the false positive rate will be substantially increased (Knoth and Schmid, 2004). The big advantage of the FDA framework over standard parametric assumptions known from time series analysis, such as auto-correlation of some specific order, is that the error process can be modeled in a very flexible, semiparametric fashion through functional principal component analysis.

In the case of SHM, there is another important aspect to consider with respect to $E_j(t)$: This process contains the relevant information for the monitoring task since it captures deviations from the system output $\alpha(t) + f(z_j(t))$ that would be expected for a specific, let us say, temperature at time t if the structure is "in-control." For exploiting this information, it is beneficial to further decompose this process into a more structural component $w_j(t)$ and white noise $\epsilon_j(t)$ with variance σ^2, i.e.,

$$E_j(t) = w_j(t) + \epsilon_j(t). \tag{2}$$

Since $\epsilon_j(t)$ is assumed to be pure noise, it does not carry relevant information, and $w_j(t)$ should be the part to focus on for monitoring purposes. To decompose $E_j(t)$ into $w_j(t)$ and $\epsilon_j(t)$, we use functional principal component analysis (FPCA). It is based on the Karhunen–Loeve expansion (Karhunen, 1947; Loève, 1946), which allows us to expand the random function $E_j(t)$ as

$$E_j(t) = \sum_{r=1}^{\infty} \xi_{rj} \phi_r(t), \tag{3}$$

where ϕ_r are orthonormal eigenfunctions of the covariance, i.e., $\int_{\mathcal{T}} \phi_r(t) \phi_k(t) dt = 1$ if and only if $k = r$ and zero otherwise. In particular, according to Mercer's theorem (Mercer, 1909),

$$\text{Cov}\big(E_j(s), E_j(t)\big) = \Sigma(s,t) = \sum_{r=1}^{\infty} v_r \phi_r(s) \phi_r(t) \tag{4}$$

with decreasing eigenvalues $v_1 \geq v_2 \geq \ldots \geq 0$. Furthermore, ξ_{rj} are uncorrelated random scores with mean zero and variance v_r, $r = 1, 2, \ldots$, and are independently normal if E_j is a Gaussian process.

In FPCA, the sum in (3) is typically truncated at a finite upper limit m, which gives the best approximation of E_j with m basis functions (Rice and Silverman, 1991). Looking at the fraction $\left\{ \sum_{r=1}^{m} v_r \right\} / \left\{ \sum_{r=1}^{\infty} v_r \right\}$ allows a quantitative assessment of the variance explained by the approximation. For choosing the concrete value m, there are various methods available (compare, e.g., Wang et al., 2016). One possibility is to use a value of m such that a high proportion, e.g., 95% or 99%, of the overall variance is explained (e.g., Sørensen et al., 2013; Gertheiss et al., 2024), here we will use 99% throughout the paper. By choosing such a large value, it is reasonable to assume that the selected m eigenfunctions account for all relevant features in the data, and the remainder is merely noise. Consequently, we set

$$w_j(t) = \sum_{r=1}^{m} \xi_{rj} \phi_r(t), \tag{5}$$

and use the scores $\xi_{1j}, \ldots, \xi_{mj}$ as damage-sensitive features for monitoring. As the functions $\alpha, f, \phi_1, \ldots, \phi_m$ are estimated from IC data in the model training phase, it is the scores obtained for future data that tell us whether the system outputs deviate from the values that would be expected for an IC structure over the day for given values of the covariate (and the specific time of the year, if $\alpha(t, d_j)$ instead of $\alpha(t)$ is used in model (1)). How to estimate the different model components, such as $\alpha, f, \phi_1, \ldots, \phi_m$ from the training data, and how to estimate the scores on future data, will be described in Section 2.2 below. At this point, it is important to note that if environmental and operational variables are present beyond z, and those variables are measured, model (1) can be extended to include multivariate covariates, see Section 2.4. If there are additional confounding effects that are not measured or not known at all, we can proceed analogously to

the popular output-only method using (multivariate) PCA. That means we assume that the first, let us say, ρ components mainly account for variation induced by latent factors (Cross et al., 2012), so we only use the remaining scores ξ_{jr}, $r > \rho$, for monitoring. An important advantage of model (1) over output-only methods is that available covariate information can still be exploited through f. How to modify our approach in terms of a pure output-only method if no covariate information is available at all will also be described in Section 2.4.

2.2. Model training

Model fitting is carried out on in-control/Phase-I training data only. This will be described in Sections 2.2.1 and 2.2.2. In Section 2.2.3, we will discuss how to obtain the scores for Phase-II data that can be used for monitoring.

2.2.1. Basic model training strategy

For estimating functions such as α or f in (1), we follow an approach popular in functional data analysis and semiparametric regression, compare Greven and Scheipl (2017) and Wood (2017). The unknown function, say f, is expanded in basis functions such that

$$f(z) = \sum_{l=1}^{L} \gamma_l b_l(z). \tag{6}$$

A popular choice for $b_1(z), \ldots, b_L(z)$ is a cubic B-spline basis, which means that f is a cubic spline function (de Boor, 1978; Dierckx, 1993). For being sufficiently flexible with respect to the types of functions that can be fitted through (6), typically, a rich basis with a large L is chosen. A large L, however, often leads to wiggly estimated functions if the basis coefficients $\gamma_1, \ldots, \gamma_L$ are fit without any smoothness constraint. The latter is typically imposed by adding a so-called *penalty* term when fitting the unknown coefficients through least-squares or maximum likelihood. A popular penalty is the integrated squared second derivative $\int_{\mathcal{D}} \left[f''(z) \right]^2 dz$, where $\mathcal{D}$ is the domain of f. An alternative is a (quadratic) penalty on the discrete second- or third-order differences of the basis coefficients γ_l, which gives a so-called P-spline, see Eilers and Marx (1996) for details. If we assume that the eigenfunctions $\phi_1, \ldots, \phi_m$ from (5) are known, model (1) becomes

$$u_j(t) = \sum_{l=1}^{L^{(\alpha)}} \gamma_l^{(\alpha)} b_l^{(\alpha)}(t) + \sum_{l=1}^{L^{(f)}} \gamma_l^{(f)} b_l^{(f)} \left(z_j(t) \right) + \sum_{r=1}^{m} \xi_{rj} \phi(t) + \epsilon_j(t), \tag{7}$$

where $\gamma_l^{(\alpha)}$, $b_l^{(\alpha)}$, $\gamma_l^{(f)}$, and $b_l^{(f)}$ are the basis coefficients and basis functions for α and f, respectively, and ξ_{rj} are day-specific random effects. Given a training dataset of system outputs and covariate information $\{u_j(t_{ji}), z_j(t_{ji})\}$ for days $j = 1, \ldots, J$ and time points $t_{ji} \in \mathcal{T}$, $i = 1, \ldots, N_j$, the unknown parameters $\gamma_l^{(\alpha)}, \gamma_l^{(f)}$ can be estimated through penalized least-squares, whereas the random effects ξ_{rj} can be predicted using (linear) mixed models approaches. Specifying ξ_{rj} as random effects makes this part of the model robust against over-fitting, thanks to the implicit shrinkage of random effects; compare, e.g., Fahrmeir et al. (2013). The trade-off between model complexity concerning α and f and fit to the data is controlled through the strength of the penalty on α and f. Instead of using a penalty, model complexity could also be controlled through the number of basis functions $L^{(\alpha)}$ and $L^{(f)}$, but using a penalty makes the method robust against the concrete choice of $L^{(\alpha)}$ and $L^{(f)}$ as it allows using large numbers (e.g., around 20), while controlling model complexity through smoothing; compare, e.g., Gertheiss et al. (2024) and Wood (2017). There are a couple of options for choosing the strength of the penalty in a specific application, such as (generalized) cross-validation. In our case, the quadratic penalties used can also be incorporated into the mixed model framework, such that the strength of the penalty can be determined through (restricted) maximum likelihood, analogously to the variance components

$v_1, \ldots, v_m$ and σ^2, see Wood (2011) and Wood (2017) for details. An in-depth discussion of these methods is beyond the scope of this paper. However, an important point to note is that model (1) and hence (7) is not identifiable in this general form given above because we may simply add some constant c to α while at the same time subtracting c from f. Then, the two models are equivalent. In other words, α and f are only identifiable up to vertical shifts. That is why an overall constant α_0 is typically introduced in terms of

$$u_j(t) = \alpha_0 + \tilde{\alpha}(t) + f(z_j(t)) + E_j(t), \tag{8}$$

and $\tilde{\alpha}$ and f are centered in some sense. The constraint used in the R package `mgcv` (Wood, 2017), which we use here for model training, is $\sum_{i,j} \tilde{\alpha}(t_{ji}) = 0$ and $\sum_{i,j} f(z_j(t_{ji})) = 0$, respectively. Furthermore, it is worth noting that measurement points t_{ji} neither need to be the same for each day nor on a regular grid. As a consequence, missing values in the output profiles or the covariate curves are allowed.

2.2.2. Estimation of eigenfunctions

In Section 2.2.1, we assumed that the eigenfunctions ϕ_r, $r = 1, \ldots, m$, are known. In practice, those functions need to be estimated from the training data in some way. To do so, we first fit model (1), or a modification from below, with a working independence assumption concerning $E_j(t)$ (Scheipl et al., 2015). Specifically, this means that $E_j(t)$ from (1) is white noise and $w_j(t)$ from (2) is omitted; compare, e.g., Ivanescu et al. (2015). Then, we use the resulting estimates of the error process for FPCA, plug in the estimated eigenfunctions $\hat{\phi}_r$, and fit the final model as discussed above. Such a two-step approach has worked well in the past (Gertheiss et al., 2017). For FPCA, we use an approach based on Yao et al. (2005) that accommodates functions with additional white noise errors and thus works relatively generally (Gertheiss et al., 2024). The idea is to incorporate smoothing into estimating the covariance in (4). The estimation is based on the cross-products of observed points within error curves, $E_j(t_{ji}) E_j(t_{ji'})$, which are rough estimates of $Cov(E(t_{ji}), E(t_{ji'}))$. All cross-products are pooled, and a smoothing method of choice is used for bivariate smoothing. Yao et al. (2005) used local polynomial smoothing, while the `refund` R package (Goldsmith et al., 2022) used here employs penalized splines. If an additional white noise error is assumed (as we do here), the diagonal cross-products approximate the variance of the structural component plus the error variance. The diagonal is thus left out for smoothing. Once the smooth covariance is available, the orthogonal decomposition (4) is done numerically on a fine grid using the usual matrix eigendecomposition.

The entire model training workflow is summarized in the orange part of Figure 2. We first use the Phase-I data (step 0) to fit the initial model with working independence in step 1. On the resulting residuals, the "observed" error process, we carry out FPCA to obtain estimates of the eigenfunctions (step 2). For choosing the number of components, we use the usual threshold of 95% or 99% of the variance explained. Then, the eigenfunctions are used to train the final model in step 3. From this model, we obtain estimates of the fixed and random effects, the variance components, and the residuals. Particularly, the fixed effects, the eigenfunctions, and the variance components will be needed to estimate Phase-II component scores and implement a monitoring scheme as described below.

2.2.3. Estimation of phase-II scores

Once the mixed model for function-on-function regression has been trained, it can be used to monitor future system outputs if the covariates in the trained model are available as well. The corresponding data is denoted as "new covariate data z" and "new system output data u" in Figure 2. An essential input for the monitoring scheme described in the next subsection are the principal component scores $\xi_{1,g}, \ldots, \xi_{m,g}$ for a new day g in Phase II. To estimate those scores from the new data, we first use the fixed effects from the Phase-I model, $\hat{\alpha}(t_{gi})$ and $\hat{f}(z_g(t_{gi}))$ from (1), to obtain a prediction for the system outputs $u_g(t_{gi})$. Here, t_{gi}, $i = 1, \ldots, N_g$, denote the time instances of the covariate $z_g(t_{gi})$ available at day g. Then, those
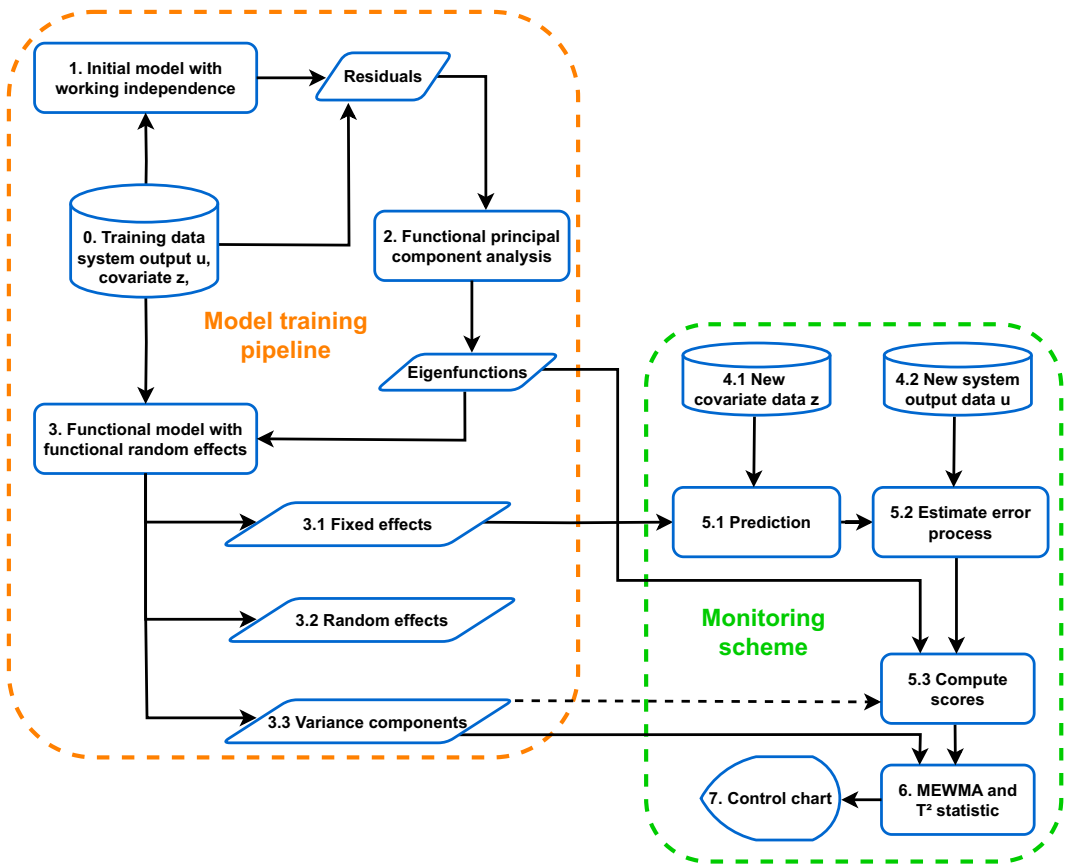
Figure 2. *Visual summary of the CAFDA-SHM framework.*

predictions are subtracted from the actually observed $u_g(t_{gi})$ to obtain estimated measurements $\hat{E}_g(t_{gi})$ of the error process. If those time points are sufficiently dense over the course of day g, which is typically the case if day g is (nearly) over, the scores can be estimated through numerical integration such that

$$\hat{\hat{\xi}}_{r,g} = \int_{\mathcal{T}} \hat{E}_g(t)\hat{\phi}_r(t)dt. \tag{9}$$

However, sometimes there is a substantial amount of missing values in the u or z data, e.g., due to technical problems, or — more importantly — the score estimates have to be available before the new day is over. In the latter case, we may say that all the data points after some time point are "missing." In both situations, the interpretation as a mixed model is helpful, as this also allows for predicting the random effects for a reduced set of measurement points t_{gi}.

For that purpose, let $\boldsymbol{\phi}_{gr} = (\phi_r(t_{g1}),\ldots,\phi_r(t_{gN_g}))^{\mathrm{T}}$ be the rth eigenfunction evaluated at time points t_{gi}, $i = 1,\ldots,N_g$, $r = 1,\ldots,m$, and $\Sigma_{\mathbf{E}_g}$ the covariance matrix of the error vector $\mathbf{E}_g = \left(E_g(t_{g1}),\ldots,E_g(t_{gN_g})\right)^{\mathrm{T}}$. Then, assuming a Gaussian distribution, the conditional expectation of the score $\xi_{r,g}$ given $\mathbf{E}_g$ is $\mathbb{E}\left(\xi_{r,g}|\mathbf{E}_g\right) = v_r\boldsymbol{\phi}_{gr}^{\mathrm{T}}\Sigma_{\mathbf{E}_g}^{-1}\mathbf{E}_g, r = 1,\ldots,m$ (Yao et al., 2005). Due to (2) and (5), the matrix $\Sigma_{\mathbf{E}_g}$ can be estimated as $\widehat{\Sigma}_{\mathbf{E}_g} = \widehat{\boldsymbol{\Phi}}_g\mathrm{diag}\left(\widehat{v}_1,\ldots,\widehat{v}_m\right)\widehat{\boldsymbol{\Phi}}_g^{\mathrm{T}} + \widehat{\sigma}^2\mathbf{I}_{N_g}$, with $\widehat{\boldsymbol{\Phi}}_g = \left(\widehat{\boldsymbol{\phi}}_{g1}|\ldots|\widehat{\boldsymbol{\phi}}_{gm}\right)$. After plugging in the estimates $\hat{\sigma}^2$, $\hat{v}_r$, $\hat{\phi}_r$, $r = 1,\ldots,m$, from the model training phase and $\hat{\mathbf{E}}_g = \left(\hat{E}_g(t_{g1}),\ldots,\hat{E}_g(t_{gN_g})\right)^{\mathrm{T}}$ from above, we obtain the estimated scores

$$\hat{\xi}_{r,g} = \hat{v}_r \hat{\phi}_{gr}^{\mathrm{T}} \hat{\Sigma}_{\mathbf{E}_g}^{-1} \hat{\mathbf{E}}_g, \ r = 1, \ldots, m. \tag{10}$$

These scores can then be used as input to a control chart. The workflow of estimating Phase-II scores is also summarized in Figure 2 (steps 4.1–5.3). The dashed arrow towards the scores (5.3 in Figure 2) indicates that variance components $\hat{v}_1, \ldots, \hat{v}_m$ and $\hat{\sigma}^2$ are needed if using (10) but not for (9).

2.3. Control charts

After accounting for environmental influences through appropriate modeling, the system outputs can be monitored by applying a control chart. A multivariate Hotelling control chart is often used to quickly detect a shift in the structural condition (Deraemaeker et al., 2008; Magalhães et al., 2012; Comanducci et al., 2016). Here, we employ a memory-type control chart, the Multivariate Exponentially Weighted Moving Average (MEWMA) introduced by Lowry et al. (1992) to the Phase-II scores $\xi_{1,g}, \ldots, \xi_{m,g}$, $g = 1, 2, \ldots$, from Section 2.2.3. The MEWMA contains the Hotelling chart as a special case. The MEWMA chart assumes serially independent, normally distributed vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ of dimension m with $\boldsymbol{\xi}_g \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. We follow Knoth (2017) and define a mean vector $\boldsymbol{\mu}$ that follows a change point model $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ for $g < \tau$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ for $g \geq \tau$ for an unknown time point τ and by definition $\boldsymbol{\mu}_0 = \mathbf{0}$, see Section 2.1. For in-control data, the scores $\xi_{1,g}, \ldots, \xi_{m,g}$ are assumed to be uncorrelated with variances $v_1, \ldots, v_m$, see also Section 2.1. Hence, the covariance matrix $\boldsymbol{\Lambda}$ is diagonal with $\boldsymbol{\Lambda} = \mathrm{diag}(v_1, \ldots, v_m)$. Then, we apply the following smoothing procedure to compute the MEWMA statistic (step 6 in Figure 2)

$$\boldsymbol{\omega}_g = (1 - \lambda)\boldsymbol{\omega}_{g-1} + \lambda\hat{\boldsymbol{\xi}}_g, \quad \boldsymbol{\omega}_0 = \mathbf{0} \tag{11}$$

with $g = 1, 2, \ldots$, smoothing constant $0 < \lambda \leq 1$, and the estimated scores $\hat{\xi}_{r,g}$ from (9) or (10) collected in $\hat{\boldsymbol{\xi}}_g$ such that $\hat{\boldsymbol{\xi}}_g = \left(\hat{\xi}_{1,g}, \ldots, \hat{\xi}_{m,g} \right)^{\mathrm{T}}$. The smoothing parameter λ controls the sensitivity of the shift to be detected. Smaller values of λ such as $\lambda \in \{0.1, 0.2, 0.3\}$, are usually selected to detect smaller shifts (Hunter, 1986), while $\lambda = 1$ results in the Hotelling chart. In this study, we use $\lambda = 0.3$. The control statistic is the Mahalanobis distance

$$T_g^2 = (\boldsymbol{\omega}_g - \boldsymbol{\mu}_0)^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{\omega}}^{-1} (\boldsymbol{\omega}_g - \boldsymbol{\mu}_0), \tag{12}$$

with an asymptotic covariance matrix of $\boldsymbol{\omega}_g$, $\boldsymbol{\Lambda}_{\boldsymbol{\omega}} = \lim_{g \to \infty} \mathrm{Cov}(\boldsymbol{\omega}_g) = \{\frac{\lambda}{2-\lambda}\}\boldsymbol{\Lambda}$. Note, if the scores in $\hat{\boldsymbol{\xi}}_g$ are estimated on a reduced set of measurement points (e.g., due to missing values), the model-based estimates $\hat{v}_1, \ldots, \hat{v}_m$ (step 3.3 in Figure 2) may not give the correct variances of the scores. In a case like this, it is recommended to run steps 5.1–5.3 from Figure 2 on the training data but with the reduced set of measurement points to obtain adjusted estimates of the variances (through the empirical versions), which can then be used in $\boldsymbol{\Lambda}$.

The MEWMA chart issues an alarm if $T_i^2 > h_4$, i.e., the control statistic is above the threshold value h_4. The stopping time $N = \min\{g \geq 1 : T_g^2 > h_4\}$, also known as average run length (ARL), is often used to measure the control chart's performance. It is defined as the average number of observations until the chart signals an alarm. If the process is in-control, the ARL (ARL$_0$) should be high to avoid false alarms. If there is a change in the underlying process, the ARL (ARL$_1$) should be low to detect changes quickly. To determine the threshold value, the ARL must be calculated when the process is in-control, usually applying a grid search or a secant rule. This ARL$_0$ can be calculated as described in Knoth (2017) and is implemented in R package `spc` (Knoth, 2022). An evaluation of our CAFDA framework on artificial data is demonstrated in Section 3.

2.4. Modified and extended models

Our basic model (1) is a particular case of a functional additive mixed model, as introduced and discussed by Scheipl et al. (2015), Scheipl et al. (2016), and Greven and Scheipl (2017), which can be simplified,

modified, or extended in various ways depending on the available data and prior knowledge. Some alternative specifications that may be an option with the data available in our case studies in Section 4 are recapped as follows.

- *Standard linear models:* If we have reason to believe that potential daily patterns can be explained entirely through variations in z, we can replace $\alpha(t)$ with a constant α_0. Similarly, if it is believed that the association of the confounder z and system output u is linear, we can replace $f\big(z_j(t)\big)$ in (1) with $\beta z_j(t)$. So, the standard linear regression model is a special case in our broader framework. Typically, if fitting the more flexible model (1) to the data, a nearly constant α and a close to linear f will indicate that the simpler model is sufficient.
- *Unmeasured covariates:* If covariate information is unavailable, $f\big(z_j(t)\big)$ can be omitted, and our approach turns into an output-only method using FPCA. In that case, replacing $\alpha(t)$ with $\alpha(t,d_j)$ is highly recommended. The reason for this is that variables such as temperature typically vary over the day and year, imposing a specific yearly pattern on u as well. The latter can be modeled through $\alpha(t,d_j)$.
- *Multiple covariates:* If p covariates $z_{j1}(t),\ldots,z_{jp}(t)$, e.g., temperature, relative humidity, wind speed, solar radiation, have an effect, or if several temperature sensors are used to account for the local temperatures at different spatial locations in the construction material, their effects can be combined additively in (1) by replacing the term $f\big(z_j(t)\big)$ with $\sum_{k=1}^{p} f_k\big(z_{jk}(t)\big)$.
- *Covariate interactions:* If $z_{jk}(t), z_{jl}(t)$ are believed to have an interacting effect on the system output, e.g., temperature and relative humidity, model (1) can also contain (two-way) interactions in terms of $f_{kl}\big(z_{jk}(t), z_{jl}(t)\big)$. In theory, interactions of higher order are possible as well. However, it can be challenging to estimate the corresponding parameter functions due to the computing resources and amount of data needed to learn those interactions reliably, compare Section 2.2.

While all those models are so-called *concurrent* models, which are useful to adjust for the temperature of the structure itself, the framework of functional additive mixed models is flexible enough to cope with delayed effects (as, e.g., plausible for ambient temperature) using so-called *historical* functional effects (Scheipl et al., 2016). Also, the very popular linear function-on-function regression approach (Centofanti et al., 2021) is included as a special case. An overview of various modeling options beyond those discussed above is provided in Appendix A. The decision about the model to use should be based on subject matter knowledge, the available data, the goodness-of-fit, and the parameter estimates. As pointed out, concurrent models are particularly useful if the structure's temperature is measured. If more than one (potential) confounder is measured, the model should be able to include multiple covariates. To distinguish between additive effects and interactions, the model fits should be compared, and the estimated (two-dimensional) surface f_{kl} should be inspected visually. The less complex (additive) model should be chosen if the two model fits are similar and the surface looks like two additive effects; see also Section 4.1. Sometimes, it may even be known from prior studies that a particular confounder effect is (almost) linear. In a case like this, however, choosing a more complex model like (1) is usually fine, too, because due to the penalty, the estimated f will typically turn out to be (nearly) linear if the data point in that direction (compare Section 4.2).

　　The resulting modifications concerning model fitting can be summarized as follows. If model (1) is simplified, parameterization (7) simplifies, as well. For instance, if an output-only approach based entirely on FPCA is used, the part in (7) referring to the regression function f vanishes. If the latter is assumed to be linear, only the corresponding β has to be estimated. If β is allowed to change over the course of the day in terms of $\beta(t)$, this function can be expanded in basis functions as it was done with f in (6) and estimated analogously. If, on the other hand, multiple covariates are given, and model (1) is extended to $\sum_{k=1}^{p} f_k\big(z_{jk}(t)\big)$ instead of a single $f\big(z_j(t)\big)$, we have to add basis representations $\sum_{l_k=1}^{L^{(f_k)}} \gamma_{l_k}^{(f_k)} b_{l_k}^{(f_k)}\big(z_{jk}(t)\big)$ for each component f_k, $k=1,\ldots,p$, in (7). If we want to account for both daily and yearly patterns in a

purely additive way, that is, $\alpha(t,d_j) = \alpha_0 + \tilde{\alpha}_1(t) + \tilde{\alpha}_2(d_j)$, the procedure is completely analogous. However, if $\alpha(t,d_j)$ is supposed to be a surface, or if a two-way interaction, say $f(z_{j1}(t), z_{j2}(t))$ in (1), is to be fitted, some minor modifications are necessary. The unknown functions (α and f, respectively) now have two arguments, meaning that the basis needs to be chosen appropriately, e.g., as a so-called tensor product basis. For instance, if considering f, equation (6) turns into

$$f(z_1, z_2) = \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \gamma_{l_1, l_2} b_{l_1}(z_1) b_{l_2}(z_2), \tag{13}$$

and smoothing is typically done in both the z_1- and z_2-directions. Having said that, the model complexity increases in terms of the number of basis coefficients γ_{l_1, l_2} that need to be fitted. As pointed out earlier, higher-order interactions can also be estimated, but the number of coefficients increases even further. Bivariate (or even higher dimensional) smoothers allowing for terms such as (13) are available in `mgcv`, and corresponding estimates will be shown in the real-world data evaluations in Section 4. The approach for bivariate $\alpha(t,d_j)$ or $f(z_j(t),t)$ cases, where the potentially nonlinear covariate effect is allowed to change over the course of the day, is analogous. In summary, in any of the cases considered here, the final model is linear in the basis coefficients and random effects, and all those quantities can be estimated/predicted through penalized least squares in a mixed model framework.

3. Illustration on artificial data

The aim of this section is to validate the methodology proposed in Section 2 on artificial data in a Monte Carlo simulation study. To this end, we first describe the data-generating process (DGP) of the functional data in the following Section 3.1. We then apply the modeling part of our framework to the artificially generated data in Section 3.2 to provide insight into the modeling performance. In the last Section 3.3, we evaluate the proposed monitoring scheme (the "green part" in Figure 2) in the CAFDA-SHM framework.

3.1. Data generation

We consider the basic model (1) with one system output and one covariate. To generate the data, we consider each component of the model separately. The eigenfunctions ϕ_r, $r = 1, 2, 3$, are obtained using orthonormal Legendre Polynomials (Abramowitz and Stegun, 1964) up to an order of two. Following (2), we split E_j into two parts. First, the structural component $w_j(t)$, $t \in \mathcal{T}$, $\mathcal{T} = (0, 24)$, is generated by $w_j(t) = \xi_{1j}\phi_1(t) + \xi_{2j}\phi_2(t) + \xi_{3j}\phi_3(t)$, where the individual scores are obtained through $\xi_{r,j} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, v_r)$ $r = 1, 2, 3$, $j = 1, \ldots, J$, and mean $\mu = 0$. The eigenvalues $v_r = \exp\left(-\frac{r+1}{2}\right)$ are decreasing exponentially towards zero (Happ-Kurz, 2020). The second component of E_j, the white noise at the measurement points t_i, $i = 1, \ldots, 24$, is generated by $\varepsilon_j(t_i) \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.2)$. That means the covariance of the error process is (implicitly) given by (4) (with "∞" replaced by "3") plus the additional variance of 0.2 on the diagonal. A functional intercept α is also created, separated into α_0 and $\tilde{\alpha}(t)$. While the functional component is $\tilde{\alpha}(t) = \sin\left(\frac{\pi t}{48}\right) + \cos\left(\frac{\pi t}{6}\right)$, $t \in \mathcal{T}$, and centered by subtracting its mean value, the overall intercept is set to $\alpha_0 = 5$. To form a functional covariate that can represent different shapes of cyclic daily profiles analogous to temperature profiles at different times of the year, a sine function in $z_j(t) = \zeta_{0j} + \zeta_{1j}\sin\left(\frac{\pi t}{12} + 0.3\right)$, $t \in \mathcal{T}$, with $\zeta_{0j}, \zeta_{1j} \overset{\text{iid}}{\sim} U(a, b)$ is used, where $U(a, b)$ denotes the uniform distribution over the interval (a, b). Thus, the covariate profile level is set by ζ_{0j} with $a = 2$ and $b = 12$, and its shape by ζ_{1j}, here with $a = 0$ and $b = 4$. Subsequently, the covariate data $z_j(t)$ is transformed by the regression function $f(z) = \exp\left(-\frac{11}{5}z\right) - 0.5$ and the output u_j for day j at the measurement points t_i is generated through $u_j(t_i) = \alpha_0 + \tilde{\alpha}(t_i) + f(z_j(t_i)) + w_j(t_i) + \epsilon_j(t_i)$, see Equations (2) and (8). Figure 3 shows a sample of 20 generated profiles of the output u_j, the covariate z_j, and the error process E_j similar to the real data in Figure 1. For our evaluations below, the sample size was $J = 300$.
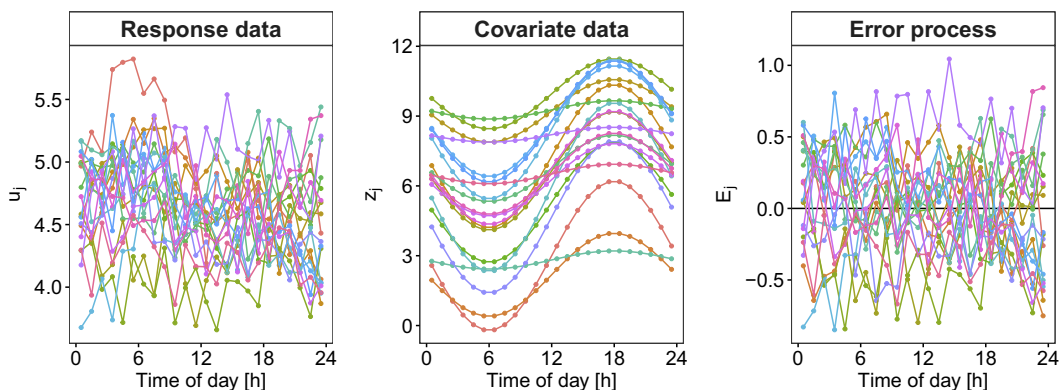
Figure 3. *Sample of 20 generated output profiles (left), covariate profiles (middle), and error profiles (right).*

3.2. Model training simulation results

Next, we perform the model training pipeline of the CAFDA-SHM framework (steps 0–3, the "orange part" in Figure 2). We repeated the procedure of generating data (according to Section 3.1) and model training 100 times to gain insights into the model performance and possible parameter estimation variability. Figure 4 illustrates the results comparing the true functions $\tilde{\alpha}(t)$, $\alpha_0 + f(z)$, and $\phi_r(t)$, $r = 1, \ldots, 3$, drawn in blue, to the estimates $\widehat{\tilde{\alpha}}(t)$, $\hat{\alpha}_0 + \hat{f}(z)$, $\hat{\phi}_r(t)$ from the 100 simulation runs (gray), where each run consisted of a training set of $J = 300$ u- and z-profiles. The respective mean functions of the estimates are shown in red. The top-left figure shows the results for the centered functional intercept
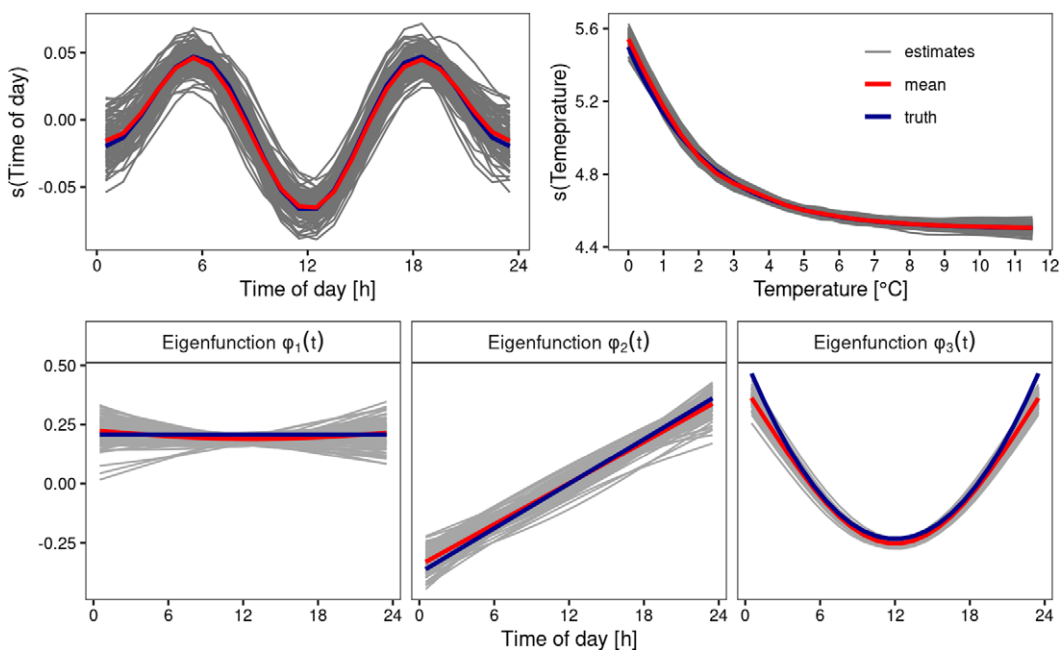
Figure 4. *The estimates of the functional intercept (top-left) and the nonlinear covariate effect (top-right), as well as the estimated eigenfunctions of the structural component of the error process (bottom row). Shown in gray are the results for each of the 100 simulation runs, while the blue curves give the true functions, and the red curves the mean across the 100 runs.*

$\tilde{\alpha}(t)$ and the top-right figure for the covariate effect plus the overall intercept $\alpha_0 + f(z)$. The three figures in the bottom row show the results for eigenfunctions $\phi_r(t)$, $r = 1,2,3$. The red and blue curves are close, meaning that the trained models and their estimated functional components can approximate the true underlying functions well on average. In other words, the estimates are (nearly) unbiased. Also, the variation is relatively small, see the gray curves.

3.3. Monitoring simulation results

In this section, the monitoring scheme (green part in Figure 2) is evaluated concerning the ARL performance in a Monte Carlo simulation. Specifically, for each of the 100 models from Section 3.2, the out-of-sample ARL is calculated by taking the average of the observed run lengths N over 10,000 replicates of the procedure with steps 1–7 given below. The control limit h_4 is determined based on a prespecified $ARL_0 = 100$ (days), and the ARL is computed by the following procedure for Phase-II days $g = 1, 2, \ldots$:

1. Generate new covariate data $z_g(t)$ according to Section 3.1 (which produces the data in step 4.1 of Figure 2).
2. Add a shift δ in one of the three components of $\boldsymbol{\mu}_0$ to create $\boldsymbol{\mu}_1$ and draw scores from the corresponding normal distribution to simulate the functional random effect $w_g(t)$.
3. Generate the output data $u_g(t_i)$ at measurement points $t_1, \ldots, t_{24}$ by adding the functional intercept, the fixed effect of temperature $f(z_g(t_i))$, the functional random effect $w_g(t_i)$, and white noise error $\epsilon_g(t_i)$, analogously to Section 3.1. This produces the data in step 4.2 of Figure 2.
4. Predict the system output using the model from Phase I (compare Section 3.2) and compute the resulting $\hat{E}_g(t_i)$ by subtracting the predicted from the observed system output.
5. Calculate the scores $\hat{\tilde{\zeta}}_{r,g}$ according to Eq. (10).
6. Compute MEWMA and T^2 statistic according to Eqs. (11) and (12).
7. If $T^2 > h_4$, note the run-length N, else repeat steps 1–6 for the next profile (day).

The upper panel of Figure 5 shows ARL profiles on a logarithmic scale for shifts in the scores' mean on a fine grid for $\delta \in [0,3]$ with three different smoothing parameters $\lambda \in \{0.1, 0.3, 1\}$. For a fair comparison, all three charts are calibrated to the same $ARL_0 = 100$. However, note that the ARL of all three charts is slightly above 100 for the in-control scenario. This can be explained by the fact that the prediction (10) of the scores in the mixed model framework imposes a Ridge-type penalty and, hence, a form of shrinkage toward zero. Furthermore, it can be seen that the choice of the smoothing parameter λ influences the ARL_1 profiles and, thus, the detection speed of the out-of-control scenario. As mentioned in Section 2.3, small values of λ lead to quicker detection of small changes, while charts using, e.g., $\lambda = 1$ (the pink curves in Figure 5) have a shorter run length on average for substantial shifts. Comparing, for example, the crossing points of the ARL profiles for different shift sizes δ for the three components in the upper row, one could conclude that a shift in the third score would be detected more quickly than in the first component. However, this effect is blurred as the variances v_r of the principal components decrease for larger r, see Section 3.1. Hence, a version with standardized shift size $\frac{\delta}{v_r}$ is plotted in Figure 5 (bottom row) to give a clearer picture of the detection speed in regards to each component. It shows that the detection speed is comparable for all three components.

For comparison, we show the results for two competing/existing approaches popular in SHM in brown and green in Figure 5. Specifically, the effect of temperature (compare the top-right panel of Figure 4) is captured through a so-called "bilinear" model, i.e., a piecewise linear model with one breakpoint. Then, the so-called "misfits" are calculated by subtracting the predicted values (hourly) from the measured eigenfrequencies. Finally, the (squared) normalized misfits can be used for monitoring, with control limits typically obtained as quantiles of a χ^2 distribution. This approach is also called "Shewhart control chart" (see, e.g., Montgomery, 2009). Following Maes et al. (2022), we consider two versions: (i) hourly data are monitored directly (the brown lines in Figure 5), and (ii) hourly misfits are averaged over 24 h (i.e., days)
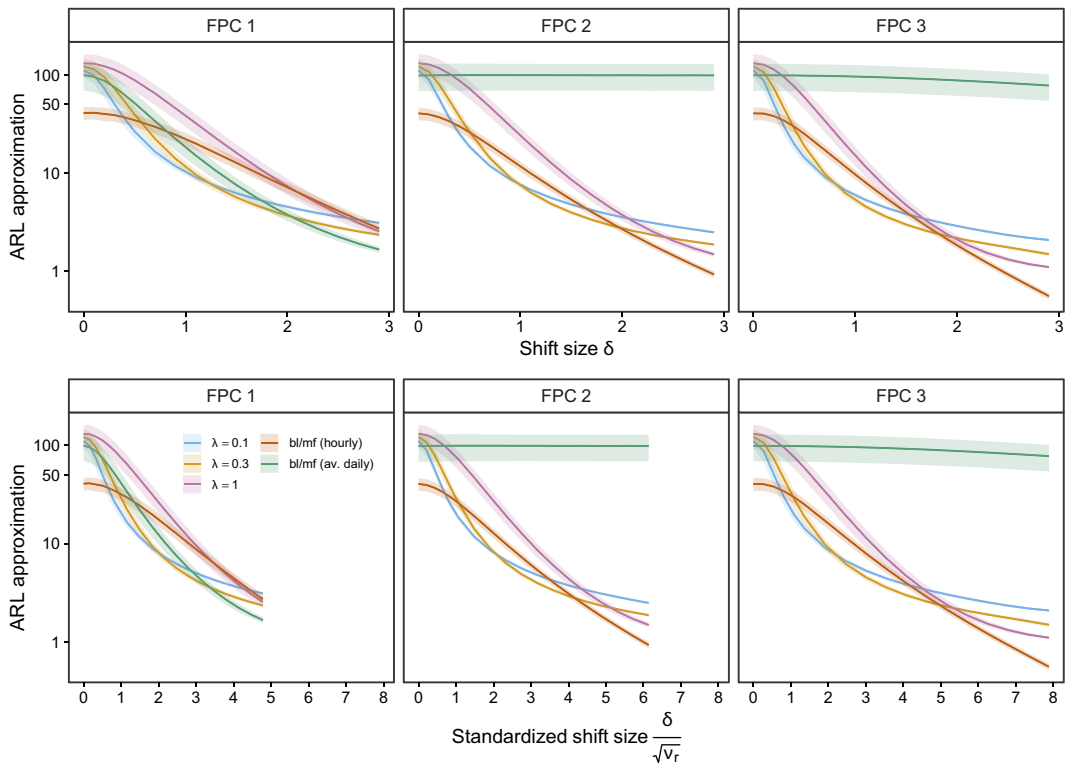
Figure 5. *ARL profiles (averaged over the 100 models/training datasets from Section 3.2) on a logarithmic scale for a shift in the scores' mean. Considered are MEWMA control charts as proposed in Section 2.3 with different smoothing parameter λ (blue, orange, pink) and Shewhart charts based on hourly (brown) or 24 h averaged (green) misfits of a piecewise linear model, displayed for absolute shift size δ (top row) and standardized shift size $\frac{\delta}{\sqrt{v_r}}$ (bottom row). The shaded regions give the uncertainty concerning the ARL approximation in terms of ± one standard deviation across the 100 models/training datasets from Section 3.2.*

before calculating the normalized misfits (the green lines in Figure 5). It is seen that averaging misfits over 24 h works very well for detecting shifts in the first (functional) principal component. The reason is that the averaged misfit approximates the first score well, i.e., it is nearly proportional to the first score; compare equation (9) and the first eigenfunction in Figure 4 (bottom-left), which is constant over the day. However, if the shift occurs in the second or third component, the averaged misfits cannot detect this (compare the results/green lines for FPC1 and FPC2/FPC3 in Figure 5). This is because the eigenfunctions are orthogonal (see the bottom row of Figure 4). The approach (i), which monitors the hourly misfits, on the other hand, can detect shifts in all three components.[1] However, it is severely miscalibrated, as the average time-to-signal on the in-control data (i.e., for a shift size $\delta = 0$) is below 50 days (instead of 100). The reason is that the common assumption of uncorrelated misfits was used to determine the control limits. The hourly misfits, however, are strongly correlated. Of course, one could try to account for this using some parametric model, such as first-order autocorrelation, when calculating the control limits. In practice, however, it is typically unclear which (parametric) model to use. As pointed out, one advantage of the FPCA-based approach is that the correlation structure is estimated semiparametrically from the

---

[1] Note, an ARL below one according to the brown curves for the hourly misfits means that (on average) a shift is already detected within the first day. But this only happens for extremely large shifts δ of more than 6(!) standard deviations.

data. So, generally speaking, it can be concluded that the proposed FDA-based approach can extract much more information from the data in a statistically sound way than simpler approaches considering misfits.

4. Applications to structural health monitoring data

In what follows, we will apply our CAFDA-SHM framework to the data from the KW51 bridge in a first case study in Section 4.1. We focus on the system's dynamic response in terms of natural frequencies. Section 4.2 provides a second case study considering displacement measurements from a reinforced concrete motorway bridge, the so-called "Sachsengraben viaduct."

4.1. The KW 51 railway bridge

KW51 is a steel railway bridge of the bowstring type, with two curved, ballasted, electrified tracks. It is 115 m long, 12.4 m wide, and located between Leuven and Brussels, Belgium, on the railway line L36N. The bridge was monitored from October 2, 2018 to January 15, 2020, with a retrofitting period from May 15 to September 27, 2019. Various quantities, such as the steel surface temperature or relative humidity, were measured on an hourly basis (Maes and Lombaert, 2020; Maes et al., 2022). Due to the large amounts of data, only fragments of the raw acceleration and inclination data are available for reuse/download. Maes and Lombaert (2020) employed operational modal analysis (OMA) to determine the modal parameters with the reference-based covariance-driven stochastic subspace (SSI-cov/ref) algorithm on an hourly basis as well. The natural frequencies of 14 modes are available for the entire monitoring period and are described in detail in Maes and Lombaert (2021). Here, we will focus on the data for mode 6, with parts of it already shown in Figure 1. In addition, we will consider two potentially confounding variables, temperature at the bridge deck level and relative humidity, both measured directly at the bridge. In a preprocessing step, some extreme outliers were removed from the dataset. Those outliers correspond to some data points that resulted from abnormal bridge behavior on particularly cold days (Maes and Lombaert, 2021; Maes et al., 2022). Eventually, 225 daily profiles were available for modeling before the retrofitting started. In Sections 4.1.1–4.1.2 below, we will discuss the CAFDA results for different model specifications, including the basic model with one covariate "temperature" only and two versions of an extended model with both temperature and relative humidity included as covariates. All those models allow for nonlinear covariate effects, whereas earlier analyses of the KW51 data (Anastasopoulos et al., 2021; Maes et al., 2022) focused on linear modeling. Finally, some results for monitoring Phase-II data will be shown in Section 4.1.3. To prevent the end of the in-control phase from coinciding with the start of the retrofitting, Phase I is limited to 200 observed profiles (between October 2, 2018, and April 19, 2019) for the model training. Out of those 200 profiles of natural frequencies, 129 profiles are observed over 24 h of the day, while the rest have between 4% and 100% of missing values. Concerning the temperature and humidity curves, 57 and 66 profiles have 4%–100% missing values, respectively. Overall, the percentage of missing values is high, at 63%.

4.1.1. Results for the basic model

We start with the basic model and the parametrization in (8). It includes an overall intercept α_0, a functional intercept $\tilde{\alpha}(t)$, a potentially nonlinear temperature effect $f(z)$, and the structural component $w_j(t)$ of the error process. After estimating the fixed effects of the initial model in step 1 of Figure 2, we apply the FPCA on 148 residual profiles (those with sufficient data points of both the natural frequencies and temperature curves available) to extract the eigenfunctions of the functional random effects as described in Section 2.2.2. The number of components to extract is chosen such that 99% of the variance with respect to the smoothed covariance matrix is explained. In step 3 of our modeling framework, we refit the basic model, incorporating the eigenfunctions to account for the functional random effects.

Figure 6 shows the estimates $\hat{\phi}_1(t),\ldots,\hat{\phi}_4(t)$ of the eigenfunctions. Note that the first two eigenfunctions already explain about 85.6% of the variance of the structural part $w_j(t)$ of the error process. The first three eigenfunctions have straightforward and physically interpretable shapes. The first eigenfunction has
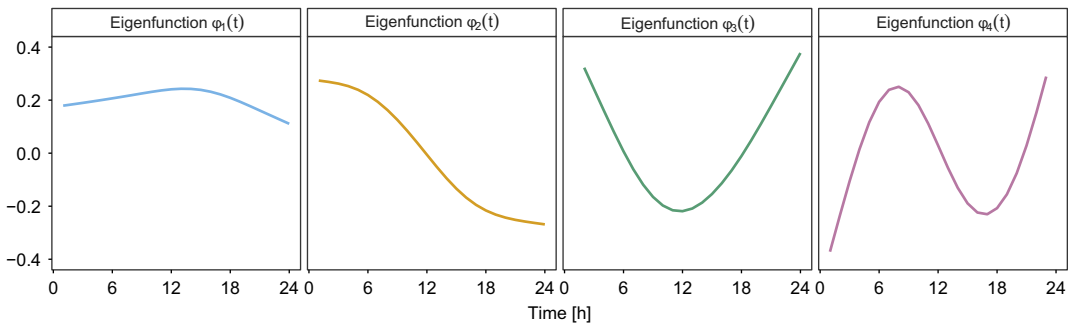
Figure 6. *Estimates of the eigenfunctions of the functional random effects in the basic model for the KW51 data.*

almost the shape of a horizontal line, which means that the first principal component represents the overall extent of the daily error in terms of a weighted average with maximum weights in the early afternoon. The second component describes the difference in the errors between the morning and evening hours, whereas the third component corresponds to the contrast between night and day. The last component, component 4, is less clear, but this component accounts for less than 5.6% of the variance of $w_j(t)$.

Figure 7 shows the centered functional intercept $\tilde{\alpha}$ and the nonlinear fixed effect $f(z)$ of temperature z on the natural frequency of mode 6, which is also centered across the data observed. The functional intercept can be interpreted as a recurring daily pattern. The gray-shaded areas represent the estimated effects' uncertainty in terms of pointwise 95% confidence intervals. Comparing both effects, we find that the effect of the functional intercept is flat over the day, which means that a recurring daily pattern, if any, is very weak if the steel temperature is taken into account. In contrast, the temperature effect shows a pronounced nonlinear shape with a kink between 2°C and 3°C. This confirms statements in the literature that the influence of temperature on the dynamic response is stronger at lower temperatures (Xia et al., 2012; Han et al., 2021). Figure 7 (right) also shows another benefit of the nonlinear approach compared to classical linear regression by Maes et al. (2022), who simply omitted the data below 2°C. Our method can also utilize the data from the colder days and thus capture the nonlinear temperature effect well.

4.1.2. Results for extended models
Next, we extend the basic model by including relative humidity as a second covariate in an additive way. We now have $u_j(t) = \alpha_0 + \tilde{\alpha}(t, d_j) + f_1(z_1(t)) + f_2(z_2(t)) + E_j(t)$, $j = 1, \ldots, J$, $t \in \mathcal{T}$, where f_1 and f_2
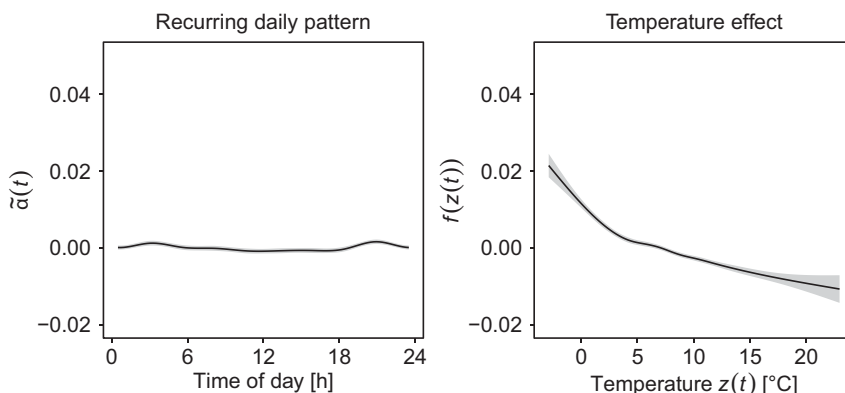

Figure 7. *Results of the basic functional modeling approach (1) for the functional intercept (left) and the nonlinear effect of temperature (right) on the natural frequency (mode 6) of the KW51 bridge.*
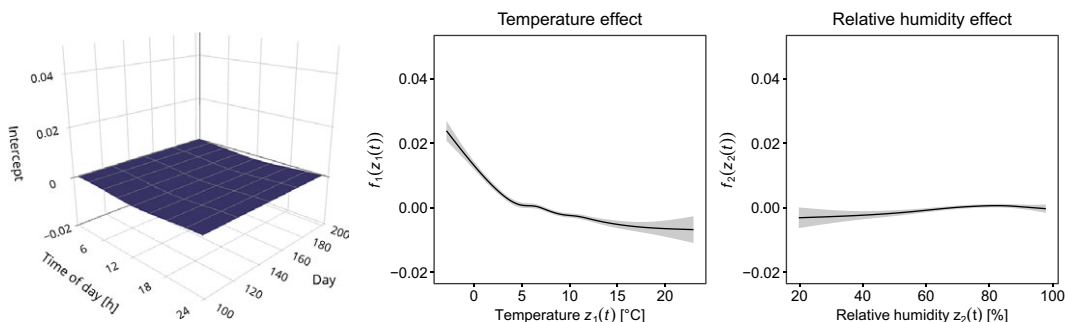
Figure 8. *Results of the extended functional modeling approach showing a two-dimensional functional intercept (left) and the additive, potentially nonlinear effects of temperature (middle) and relative humidity (right) on the natural frequency (mode 6) of the KW51 bridge.*

denote the potentially nonlinear effects of temperature and relative humidity, respectively. Due to the limited number of datasets where the two covariates are measured jointly, the number of daily profiles that can be used for parameter estimation reduces to $J = 100$ days. As before, we proceed with the model training steps 1–3 from Figure 2. Figure 8 shows the results for the final model, the (centered) two-dimensional functional intercept (left), and the effects of temperature (middle) and relative humidity (right). The flat intercept indicates that there are no daily and yearly patterns, such as seasonal effects, that cannot be explained through the covariates in the model. The (explicit) temperature effect f_1 shows a similar nonlinear behavior as seen with the basic model in Figure 7. The effect f_2 for relative humidity is much smaller but still present, as can also be seen from the increased R^2 of 0.52, compared to 0.43 for the basic model with temperature as the only covariate, see Table 1.

As a second extension, we replace the additive effects from above by the term $f_{12}\big(z_{j1}(t), z_{j2}(t)\big)$ for the two covariates, allowing for an interacting effect on the system outputs. Figure 9 shows the resulting (centered) two-dimensional functional intercept $\tilde{\alpha}(t, d_j)$ and the surface f_{12} taking interactions of temperature and relative humidity into account in a smooth, nonparametric way. As we can see, the strongest effect is along the temperature axis, but humidity is also relevant, with its most substantial effect around 10°C. As before, the functional intercept is flat, indicating no further daily or seasonal patterns. The reason why the yellow color of the surface in the right plot of Figure 9 is only seen for low humidity is that no data points with negative temperatures and high humidity are available. That is why f_{12} cannot be fitted in that area, whereas data points with negative temperature and low humidity are found in the dataset. For the latter, the highest natural frequency values (of mode 6) are predicted as indicated by the yellow color. The overall model fit ($R^2 = 0.54$) is slightly better than with the additive model (compare Table 1) and, as such, considerably better than the basic model. In summary (see also Table 1), the temperature appears to be the most influential covariate. For comparison, using the two-dimensional intercept only in terms of an output-only approach leads to an $R^2 = 0.25$, which is substantially lower than the $R^2 = 0.43$ obtained with the basic model. Further note: If the monitoring as proposed here (the green part of Figure 2) is to use the results of the output-only model, training data

Table 1. *Summary of the R^2, overall number of observations, and number of profiles used in Phase I for the different compared models*

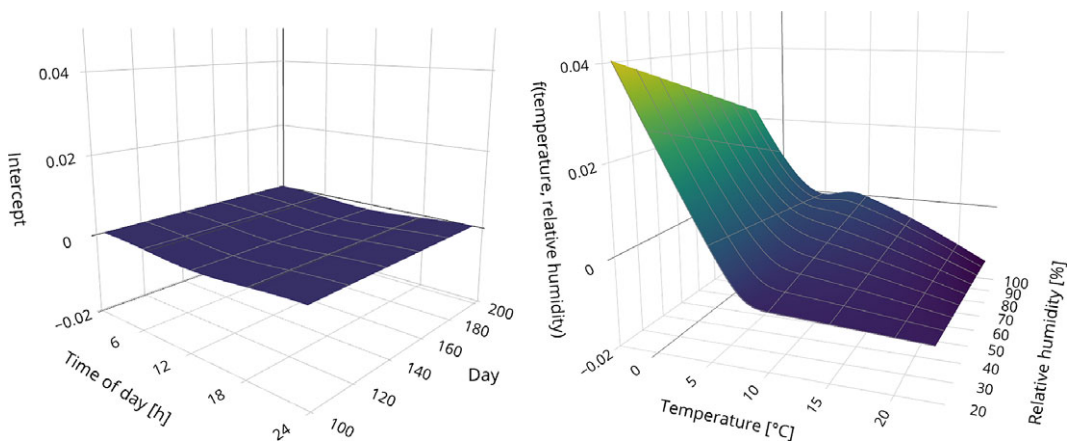|	Basic	Output only	Additive	Interactions
R^2	0.43	0.25	0.52	0.54
No. of observations	3,411	4,488	2,268	2,268
No. of profiles	148	195	100	100

Figure 9. *Results of the extended functional modeling approach showing a two-dimensional functional intercept $\alpha(t, d_j)$ (left) and a two-dimensional functional interaction of temperature and relative humidity $f_{12}(z_{j1}(t), z_{j2}(t))$ (right) on the natural frequency (mode 6) of the KW51 bridge.*

over an entire year must be available because otherwise, for some future Phase-II days, no predictions and, thus, no error profiles for calculating the scores are available. In general, this is also true for other models like Figures 8 and 9 with two-dimensional intercepts, but in the specific case of KW51 considered here, those intercepts are flat, meaning that they are irrelevant for prediction. Consequently, those models can be used for monitoring, as illustrated in Section 4.1.3 below. Monitoring with an output-only model will be considered in Section 4.2.

4.1.3. Anomaly detection

Finally, we apply the second part of the CAFDA-SHM framework, the monitoring scheme, to the natural frequency data (mode 6) of the KW51 bridge, including the retrofitting data for online monitoring as in Maes et al. (2022). Two MEWMA chart configurations with different λ are chosen, and two functional models are compared: the basic model (Section 4.1.1) and the best-performing model according to Table 1, the extended model with interaction term. As mentioned, Phase I consists of the first 200 days. However, only 3411 and 2268 observations, or 148 and 100 profiles, could be used for the basic and the extended model, respectively, due to data availability, compare Table 1. The control charts are calibrated to an $ARL_0 = 370.4$ applying thresholds $h_4 = 16.25$ and $h_4 = 15.83$ for $\lambda = 1$ and $\lambda = 0.3$, respectively. Utilizing the estimated parameters of the basic model and the extended model, respectively, and the incoming data, the scores $\hat{\xi}_{r,g}$ from (10), $m = 4$, are used in (12) to calculate the control statistic T_g^2 for day g (compare Section 2.3).

Figure 10 displays the resulting control charts. In all MEWMA charts, online monitoring begins in Phase II after the dashed line, with the retrofitting period marked through the gray shaded area. If using the basic model and the control chart with $\lambda = 1$ (top-left), the first signal is seen on June 8, 2019, while the chart, which uses the extended model (top-right) issues the first signal 6 days earlier on June 2, 2019, thus showing a higher sensitivity. Thereafter, both charts indicate a clear shift in the natural frequency of mode 6. The control charts with $\lambda = 0.3$ (bottom-row) show a false alarm on January 24–27, 2019, for the basic model, and on January 26 and 27, 2019 for the extended model, respectively. The bottom-row panels also reveal that compared to $\lambda = 1$, an alarm was triggered on May 26, 2019, so 11 days earlier in both charts. However, comparing the different scenarios, it is noticeable that all four control charts issue several signals in a row, indicating a sustained change in the process within the retrofitting (gray shaded area).
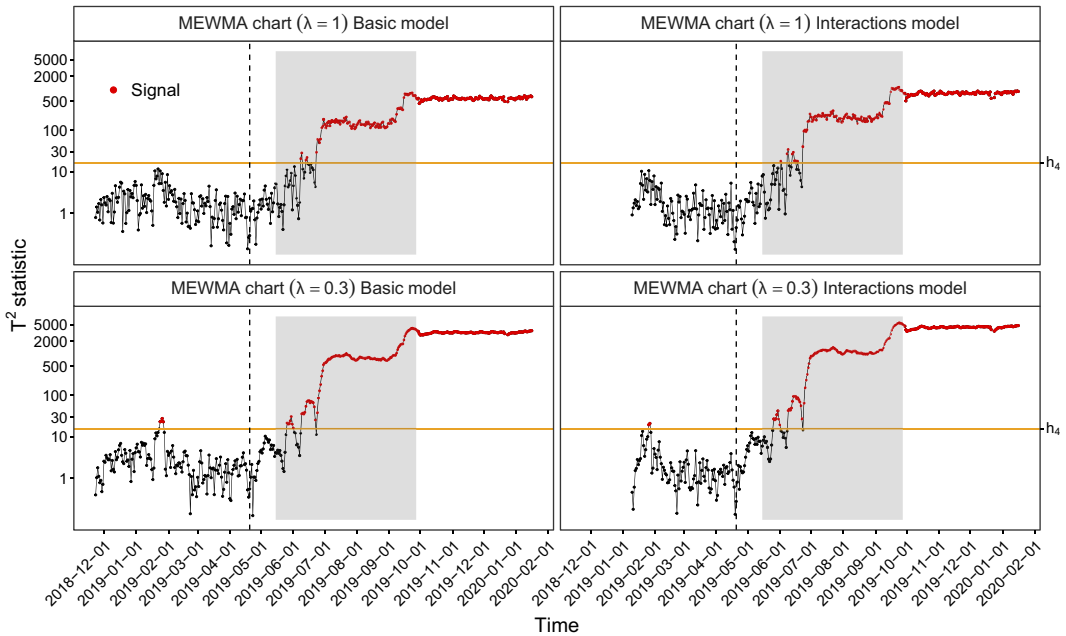
Figure 10. *MEWMA control charts for $\lambda = 1$ (top row) and $\lambda = 0.3$ (bottom row), on a logarithmic y-axis, using the basic (left column) and extended functional model with interactions (right column) for the KW51 bridge, including the retrofitting period (gray shaded area).*

In summary, employing our CAFDA-SHM framework, the control charts detect a change in the bridge's behavior caused by retrofitting, with faster detection being achieved through MEWMA parameterizations with $\lambda < 1$. Note, if using the misfits from the piecewise linear model, analogously to the simulation study in Section 3 and Maes et al. (2022), not only the hourly (as in Section 3) but also the 24 h averaged misfits are highly correlated in time, leading to miscalibrated control limits. That is why we only showed the results for CAFDA-SHM here. Also note that we considered mode 6 for illustration and evaluation because the corresponding eigenfrequencies show the typical, nonlinear behavior reported in the literature (Han et al., 2021; Maes et al., 2022; Peeters and De Roeck, 2001). Similar behavior is found for modes 3, 9, and 13 of the KW51, and results obtained through CAFDA-SHM are also very similar.

4.2. The Sachsengraben viaduct

This section presents an application of the CAFDA-SHM framework on a second real-world SHM dataset, the OSIMAB[2] dataset (Bundesanstalt für Straßenwesen (BASt), 2023), which contains measurements from the Sachsengraben bridge, see Figure 11. The bridge is located on the A45 motorway in Germany and was constructed in 1971. It spans 98 m and is made of prestressed concrete. The bridge's cross-section consists of a single-cell box girder with a construction height of 2.8 m. The webs, floor, and deck slab also have coves; the end and support cross girders are stiffened. There are two separate substructures, and we will only focus on the northern one. The superstructure has three fields, divided by pillars. The two outer fields are 30 m long, while the middle field is 38 m long (Bundesanstalt für Straßenwesen (BASt), 2021).

Displacement was measured with eight displacement sensors per superstructure for 18 months, from January 1, 2020 to August 1, 2021 and with a sampling frequency of 100 Hz for the first 191 days and a 40-minute integration interval every hour. On day 191, the measurement switched to a 1 Hz sampling

---
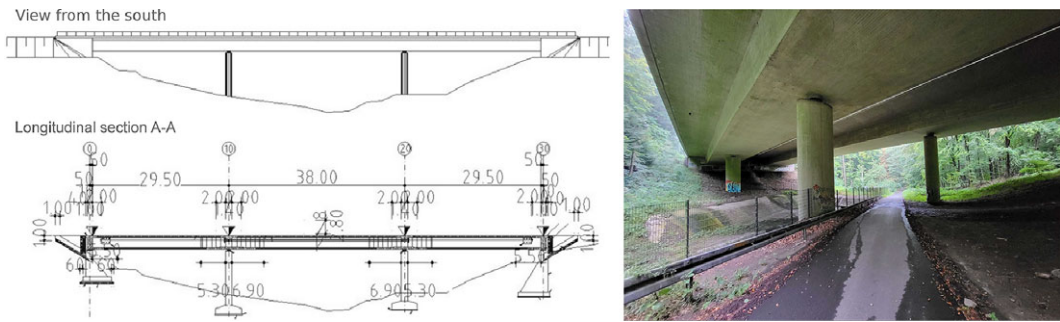
[2] Online-Sicherheits-Managementsystem für Brücken

Figure 11. *Schematic representation (left) of the Sachsengraben viaduct from the OSIMAB report (Bundesanstalt für Straßenwesen (BASt), 2021) and a photo of the bridge in 2023 (right).*
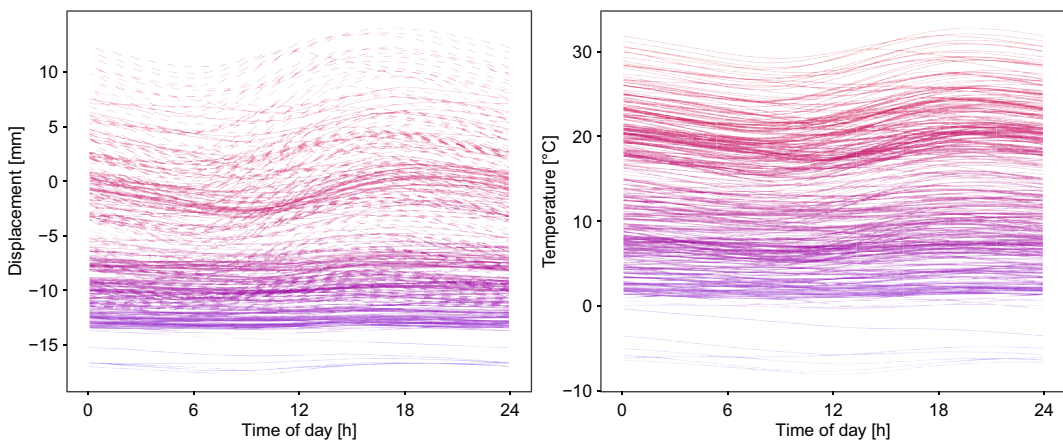

Figure 12. *Phase-I profiles of displacement sensor N_F3_WA_NO (left) and temperature sensor N_F1_T_1 (right) with a 10-minute sampling rate. The profiles are highlighted in color according to their average daily temperature.*

scheme, recording the data continuously. Ambient air temperature data recorded by a weather station and data from ten material temperature sensors with a sampling frequency of 1 Hz are also available. The entire dataset was collected in the OSIMAB project. A sample of the recorded measurement data is publicly available (Bundesanstalt für Straßenwesen (BASt), 2020), and the complete dataset is accessible via the Bundesanstalt für Straßenwesen (BASt) (2023).

Figure 12 shows the Phase-I profiles of one displacement sensor (N_F3_WA_NO, left) and measurements from the closest temperature sensor (N_F1_T_1, right) for 369 days. The dashed lines represent the first 191 days, from February 10, 2020 to August 19, 2020. The displacement and temperature measurements were resampled from the higher frequencies (100 and 1 Hz, respectively) by taking the median of the measurements within a 10-minute interval. This leads to a substantially larger number (144) of measurement points per day than with the KW51 bridge data, where the eigenfrequencies were obtained at 1-hour intervals. Furthermore, the 1.5-year measurement period means that data from each time of the year is available, and a two-dimensional intercept can be fitted that spans the entire year. The temperature values observed range from –10°C to 30°C. Out of 538 profiles in total, 287 displacement profiles and 461 temperature profiles did not exhibit any missing values, and overall, 20% of the data is missing.[3] This makes clear again that for SHM data, the methods used must be able to deal with missing data.

---

[3] 17 days (out of 538) have less than 13 observations per profile. Those days were excluded from model training.

As mentioned above, the measurement recording process changed during Phase I. Nevertheless, we can still exploit the complete information available and capture a complete annual cycle in our framework. For doing so, we assume that changing the measurement systems' sampling rate does not impact the external influence of the covariates — the fixed effects in our model $u_j(t) = \alpha_0 + \tilde{\alpha}(t, d_j) + f(z_j(t)) + E_j(t)$, $j = 1, \ldots, J$, $t \in \mathcal{T}$, where $\tilde{\alpha}$ is the two-dimensional intercept (potentially) varying over both the course of the day and the year, and f denotes the effect of the temperature. However, the functional random effects may show different behavior for the two time periods, which means that the eigenfunctions and variance components may differ. That is why two sets of residuals, before and after switching the sampling rate, were extracted to estimate two sets of eigenfunctions. It turned out that the eigenfunctions from the two time periods are very similar, but the variances differ substantially. Therefore, data were pooled to estimate a joint set of eigenfunctions, but variance components were allowed to differ between the two time periods when refitting the model in step 3 of Figure 2. Furthermore, as described in Sections 2.1 and 2.4, our model framework contains an output-only method as a special case in which only the system output of interest and its recorded timestamp are used, but no covariate data has to be recorded. In this case, the $f(z_j(t))$ from the model above is omitted, and the only fixed effects left are the overall intercept α_0 and the two-dimensional surface $\tilde{\alpha}(t, d_j)$, taking seasonal patterns into account. The latter function $\tilde{\alpha}$ is parameterized using a tensor product basis analogously to (13) with cyclic cubic regression splines in the d_j-direction, which means that the function's ends at the beginning and end of the year match up to the second derivative (Wood, 2017). Using this reduced model, we could use all available Phase I data from the displacement sensor, without being limited by possibly missing covariate information.

The left panel in Figure 13 visualizes $\tilde{\alpha}(t, d_j)$ of the final output-only model with an $R^2 = 0.70$ for the fixed effects. We can clearly see a yearly pattern with (on average) lower output values in the winter and higher values in the summer. The reason for these seasonal effects in the output-only model becomes clear from the middle and right panels of Figure 13, where the resulting estimates of the fixed effects in the extended model with the temperature included are shown. Now the two-dimensional intercept (middle) is rather flat over the year, indicating that the seasonal effects seen before were mainly caused by temperature variation. The explicit temperature effect on displacement in terms of the estimated f is shown in the right panel of Figure 13. This effect is almost linear with a tiny kink at around 15°C, which confirms similar results found in the literature, compare, e.g., Han et al. (2021). In summary, it can be concluded that by specifying a two-dimensional functional intercept in the output-only model, we can already account for some of the environmental effects without the need to record corresponding covariate information. Consequently, if using an output-only approach to adjust for environmental influences, such a functional intercept should always be included if training data are available over a sufficiently long period.
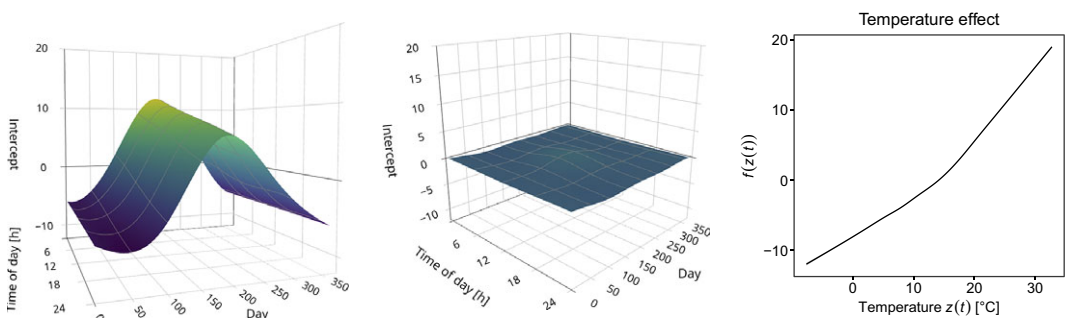

Figure 13. *Results of the functional output-only model (left) and the extended functional modeling approach (middle) showing a two-dimensional functional intercept $\tilde{\alpha}(t, d_j)$ and the (non)linear effect of temperature sensor N_F1_T_1 (right) on the displacement sensor N_F3_WA_NO.*
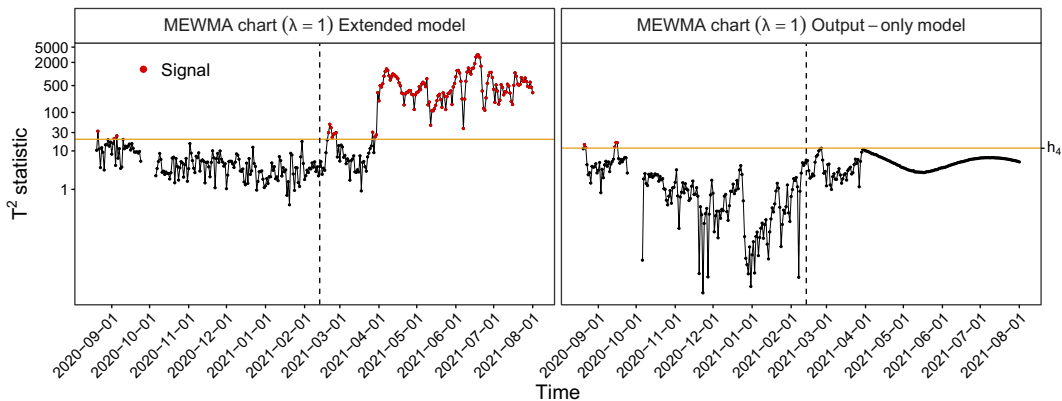
Figure 14. *MEWMA Control chart for $\lambda = 1$ on a logarithmic y-axis, using an extended functional model (left) and the output-only functional model (right) for the Sachsengraben viaduct.*

Similar to applying the control chart for anomaly detection in Section 4.1.3 and as shown in Figure 2, we set up the control charts, calibrating them to an $ARL_0 = 370.4$. For monitoring purposes, only the second set of variance components is used, as those refer to the reconfigured measurement system with the new sampling rate. Figure 14 displays the control charts for $\lambda = 1$ and a chart threshold $h_4 = 20.06$ for the extended model and $h_4 = 11.83$ for the output-only model. It is visible from both control charts that, besides the three false alarms in August and September 2021 in the Phase I data, the measurements before the dashed vertical line seem to be in control. The dashed line marks the beginning of the online monitoring on February 13, 2021. Bundesanstalt für Straßenwesen (BASt) (2021) reported that a sensor malfunction induced by a rust film blocking the sensor was noticed and confirmed at the end of March. The control chart utilizing the extended model in Figure 14 (left) detects this anomaly. However, it additionally detects some irregularities at the end of February. An explanation for this could be that the sensor malfunction already started before the end of March, and the control chart detected this. In contrast, the control chart using the output-only model (right) slightly fails to detect the anomaly. However, a modified control chart setup with $\lambda < 1$ (not shown here) would increase the chart's detection sensitivity and lead to signals in Phase II. In summary, it can be stated that although the output-only model takes seasonal patterns caused by environmental effects into account, it is beneficial to account for these effects more explicitly.

5. Concluding remarks

The modeling and monitoring methodology presented in this paper combines several parts of a structural health monitoring system in one unified framework (CAFDA-SHM). It is of high practical relevance, as it tackles important challenges in SHM. The discussed functional data approach offers the flexibility to model recurring daily and yearly patterns alongside environmental or operational influences in a highly adaptable and interpretable semiparametric manner. It can handle missing observations and effectively accounts for variations and correlations in the error process through functional principal components. Furthermore, FPCA extracts interpretable, data-based features, which are then utilized for monitoring within a sound statistical framework, thus providing a reliable basis for decision-making. Our real-world data analyses showed that CAFDA-SHM can be used for covariate adjustment and monitoring of different types of response variables: dynamic responses, as shown in Section 4.1, and quasi-static responses, as demonstrated in Section 4.2. It is hence planned to implement and test the methods presented in this article as part of a larger online SHM system (Kessler et al., 2024).

Although the proposed modeling approach is an input–output method, it contains an output-only FPCA-based version as a particular case. The latter has advantages over existing approaches as it accounts for daily and yearly patterns typically ignored by common PCA-based methods. Utilizing state-of-the-art

functional additive mixed models enables nonlinear modeling, as often needed in SHM. Our approach is flexible in handling diverse data types, from sparse and aggregated to high-resolution, dense data. Although the focus of this article was on concurrent models, the framework of functional additive mixed models includes various other options, such as historical functional effects or the common linear function-on-function approach (compare Appendix A).

Finally, there are three noteworthy limitations. First, if the error process becomes close to white noise, e.g., because the covariates' explanatory power is very high, the functional approach no longer provides much of an advantage over a standard, non-functional response surface model. Second, in the functional additive mixed model, we implicitly assume that the error process and the functional random effects are (at least approximately) Gaussian. However, if outliers are present in the data (compare, e.g., Capezza et al., 2024), an additional step to identify these outliers would need to be included in a preprocessing step in our framework. Third, the framework presented in this paper is designed for univariate functional responses. Future research will involve modeling multiple system outputs simultaneously, such as natural frequencies of multiple modes.

References

Abramowitz M and Stegun IA (1964) *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* New York: Dover Publications.

Anastasopoulos D, **De Roeck G and Reynders EPB** (2021) One-year operational modal analysis of a steel bridge from high-resolution macrostrain monitoring: Influence of temperature vs. retrofitting. *Mechanical Systems and Signal Processing 161*, 107951.

Avci O, **Abdeljaber O**, **Kiranyaz S**, **Hussein M**, **Gabbouj M and Inman DJ** (2021) A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical Systems and Signal Processing 147*, 107077.

Bundesanstalt für Straßenwesen (BASt) (2020) *OSIMAB – Online-Sicherheits-Managementsystem für Brücken.* Data licence Germany – attribution – Version 2.0.

Bundesanstalt für Straßenwesen (BASt) (2021) *Verbundprojekt OSIMAB (Online Sicherheits-Mangementssystem für Brücken).* Gesamtabschlussbericht.

Bundesanstalt für Straßenwesen (BASt) (2023) OSIMAB – Online-Sicherheits-Managementsystem für Brücken. (accessed 15 October 2023).

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** (2024) Robust multivariate functional control chart. *Technometrics 66*(4), 531–547.

Centofanti F, **Lepore A**, **Menafoglio A**, **Palumbo B and Vantini S** (2021) Functional regression control chart. *Technometrics 63*(3), 281–294.

Chen Z, **Bao Y**, **Li H and Spencer BF** (2018) A novel distribution regression approach for data loss compensation in structural health monitoring. *Structural Health Monitoring 17*, 1473–1490.

Chen Z, **Bao Y**, **Li H and Spencer BF** (2019) LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data. *Mechanical Systems and Signal Processing 121*, 655–674.

Chen Z, **Bao Y**, **Tang Z**, **Chen J and Li H** (2020) Clarifying and quantifying the geometric correlation for probability distributions of inter-sensor monitoring data: A functional data analytic methodology. *Mechanical Systems and Signal Processing 138*, 106540.

Chen Z, **Lei X**, **Bao Y**, **Deng F**, **Zhang Y and Li H** (2021) Uncertainty quantification for the distribution-to-warping function regression method used in distribution reconstruction of missing structural health monitoring data. *Structural Health Monitoring 20*(6), 3436–3452.

Comanducci G, **Magalhães F**, **Ubertini F and Cunha Á** (2016) On vibration-based damage detection by multivariate statistical techniques: Application to a long-span arch bridge. *Structural Health Monitoring 15*(5), 505–524.

Cross EJ, **Koo K**, **Brownjohn JMW and Worden K** (2013) Long-term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing 35*(1), 16–34.

Cross EJ, **Manson G**, **Worden K and Pierce SG** (2012) Features for damage detection with insensitivity to environmental and operational variations. *Proceedings of the Royal Society A 468*(2148), 4098–4122.

de Boor C (1978) *A Practical Guide to Splines*. New York: Springer-Verlag.

Deraemaeker A, **Reynders E**, **De Roeck G and Kullaa J** (2008) Vibration-based structural health monitoring using output-only measurements under changing environment. *Mechanical Systems and Signal Processing 22*(1), 34–56.

Dierckx P (1993) *Curve and Surface Fitting with Splines*. New York: Oxford University Press.

Eilers PHC and Marx BD (1996). Flexible smoothing with B-splines and penalties. *Statistical Science 11*(2), 89–121.

Fahrmeir L, **Kneib T**, **Lang S and Marx B** (2013) *Regression – Models, Methods and Applications*. Heidelberg: Springer Berlin.

Gertheiss J, **Goldsmith J and Staicu AM** (2017) A note on modeling sparse exponential-family functional response curves. *Computational Statistics & Data Analysis 105*, 46–52.

Gertheiss J, **Rügamer D**, **Liew BXW and Greven S** (2024) Functional data analysis: An introduction and recent developments. *Biometrical Journal 66*(7), e202300363.

Goldsmith J, **Scheipl F**, **Huang L**, **Wrobel J**, **Di C**, **Gellar J**, **Harezlak J**, **McLean MW**, **Swihart B**, **Xiao L**, **Crainiceanu C and Reiss PT** (2022) *refund: Regression with Functional Data*. R Package version 0.1–26.

Greven S and Scheipl F (2017) A general framework for functional regression modelling. *Statistical Modelling 17*(1–2), 1–35.

Han Q, **Ma Q**, **Xu J and Liu M** (2021) Structural health monitoring research under varying temperature condition: A review. *Journal of Civil Structural Health Monitoring 11*, 149–173.

Happ-Kurz C (2020) Object-oriented software for functional data. *Journal of Statistical Software 93*(5), 1–38.

Hou R and Xia Y (2021) Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019. *Journal of Sound Vibration 491*, 115741.

Hu W-H, **Moutinho C**, **Caetano E**, **Magalhães F and Cunha Á** (2012) Continuous dynamic monitoring of a lively footbridge for serviceability assessment and damage detection. *Mechanical Systems and Signal Processing 33*, 38–55.

Hunter JS (1986) The exponentially weighted moving average. *Journal of Quality Technology 18*(4), 203–210.

Ivanescu AE, **Staicu A-M**, **Scheipl F and Greven S** (2015) Penalized function-on-function regression. *Computational Statistics 30*, 539–568.

Jiang H, **Wan C**, **Yang K**, **Ding Y and Xue S** (2021) Modeling relationships for field strain data under thermal effects using functional data analysis. *Measurement 177*, 109279.

Karhunen K (1947) Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae. Series A I Mathematica-Physica 37*, 1–79.

Kessler S, **Mendler A**, **Klemm A**, **Jaelani Y**, **Köhncke M and Gündel M** (2024) A highway bridge—A database for digital methods. *Structural Concrete 25*(55), 3046–3049.

Knoth S (2017) ARL numerics for MEWMA charts. *Journal of Quality Technology 49*(1), 78–89.

Knoth S (2022) *spc: Statistical Process Control—Calculation of ARL and Other Control Chart Performance Measures*. R package version 0.6.7.

Knoth S and Schmid W (2004) Control charts for time series: A review. In Lenz H-J and Wilrich P-T (eds.), *Frontiers in Statistical Quality Control 7*. Heidelberg. Physica-Verlag HD, pp. 210–236,

Kullaa J (2011) Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring. *Mechanical Systems and Signal Processing 25*(8), 2976–2989.

Lei X, **Chen Z and Li H** (2023a) Functional outlier detection for density-valued data with application to robustify distribution-to-distribution regression. *Technometrics 65*(3), 351–362.

Lei X, **Chen Z**, **Li H and Wei S** (2023b) A change-point detection method for detecting and locating the abrupt changes in distributions of damage-sensitive features of SHM data, with application to structural condition assessment. *Structural Health Monitoring 22*(2), 1161–1179.

Lei X, **Chen Z**, **Li H and Wei S** (2023c) Detecting and testing multiple change points in distributions of damage-sensitive feature data for data-driven structural condition assessment: A distributional time series change-point analytic approach. *Mechanical Systems and Signal Processing 196*, 110344.

Lowry CA, **Woodall WH**, **Champ CW and Rigdon SE** (1992) A multivariate exponentially weighted moving average control chart. *Technometrics 34*(1), 46–53.

Loève M (1946) Fonctions aléatoires du second ordre. *La Revue Scientifique 84*, 195–206.

Maes K and Lombaert G (2020) Monitoring data for railway bridge KW51 in Leuven, Belgium, before, during, and after retrofitting. Zenodo. https://doi.org/10.5281/zenodo.3745914.

Maes K and Lombaert G (2021) Monitoring railway bridge KW51 before, during, and after retrofitting. *Journal of Bridge Engineering 26*(3), 04721001.

Maes K, **Van Meerbeeck L**, **Reynders EPB and Lombaert G** (2022) Validation of vibration-based structural health monitoring on retrofitted railway bridge KW51. *Mechanical Systems and Signal Processing 165*, 108380.

Magalhães F, **Cunha A and Caetano E** (2012) Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. *Mechanical Systems and Signal Processing 28*, 212–228.

Maleki MR, **Amiri A and Castagliola P** (2018) An overview on recent profile monitoring papers (2008–2018) based on conceptual classification scheme. *Computers and Industrial Engineering 126*, 705–728.

Mercer J (1909) Functions of positive and negative type, and their connection with the theory of integral equations. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 83*(559), 69–70.

Montgomery DC (2009) *Statistical Quality Control: A Modern Introduction*. Hoboken, NJ: Wiley.

Momeni H and Ebrahimkhanlou A (2022) High-dimensional data analytics in structural health monitoring and non-destructive evaluation: A review paper. *Smart Materials and Structures 31*(4), 043001.

Moser P and Moaveni B (2011) Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mechanical Systems and Signal Processing 25*(7), 2336–2357.

Muggeo VMR (2008) Segmented: An R package to fit regression models with broken-line relationships. *R News 8/1*, 20–25.

Peeters B and De Roeck G (2001) One-year monitoring of the Z24-Bridge: Environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics 30*, 149–171.

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramsay JO and Silverman BW (2005) *Functional Data Analysis*. New York: Springer.

Rice JA and Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society Series B (Methodological) 53*(1), 233–243.

Scheipl F, **Gertheiss J and Greven S** (2016) Generalized functional additive mixed models. *Electronic Journal of Statistics 10*(1), 1455–1492.

Scheipl F, **Staicu AM and Greven S** (2015) Functional additive mixed models. *Journal of Computational and Graphical Statistics 24*(2), 477–501.

Sørensen H, **Goldsmith J and Sangalli L** (2013) An introduction with medical applications to functional data analysis. *Statistics in Medicine 32*(30), 5222–5240.

Wang JL, **Chiou JM and Müller HG** (2016) Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Wang Z, **Yang DH**, **Yi TH**, **Zhang GH and Han JG** (2022) Eliminating environmental and operational effects on structural modal frequency: A comprehensive review. *Structural Control and Health Monitoring 29*(11), e3073.

Wittenberg P, **Knoth S and Gertheiss J** (2024) Structural health monitoring with functional data: Two case studies. *arXiv:stat.AP/2406.01262. (accepted for publication).*

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 73*(1), 3–36.

Wood SN (2017) *Generalized Additive Models: An Introduction with R*, 2nd Edn. Boca Raton: CRC Press.

Woodall WH (2007) Current research on profile monitoring. *Production, 17*(3), 420–425.

Woodall WH, **Spitzner DJ**, **Montgomery DC and Gupta S** (2004) Using control charts to monitor process and product quality profiles. *Journal of Quality Technology 36*(3), 309–320.

Worden K and Cross EJ (2018) On switching response surface models, with applications to the structural health monitoring of bridges. *Mechanical Systems and Signal Processing 98*, 139–156.

Xia Q, **Zhang J**, **Tian Y and Zhang Y** (2017) Experimental study of thermal effects on a long-span suspension bridge. *Journal of Bridge Engineering 22*(7), 04017034.

Xia Y, **Chen B**, **Weng S**, **Ni YQ and Xu YL** (2012) Temperature effect on vibration properties of civil structures: A literature review and case studies. *Journal of Civil Structural Health Monitoring 2*, 29–46.

Xia Y, **Hao H**, **Zanardo G and Deeks A** (2006) Long term vibration monitoring of an RC slab: Temperature and humidity effect. *Engineering Structures 28*(3), 441–452.

Yao F, **Müller HG and Wang JL** (2005) Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association 100*(470), 577–590.

Zhou RR, **Serban N and Gebraeel N** (2011) Degradation modeling applied to residual lifetime prediction using functional data analysis. *Annals of Applied Statistics 5*(2B), 1586–1610.

Appendix

A. Further modeling options

This Section provides an overview of various modeling options beyond the models discussed in the main paper. All models considered in Sections 2, 3, 4.1, and 4.2 were so-called *concurrent* models, where it is assumed that the system output at time t is influenced by covariate z measured at the same time t only. This makes sense for covariates such as the temperature of the structure itself, as available for the KW51 bridge and the Sachsengraben viaduct. However, if only ambient temperature is given, it seems more reasonable to assume that the temperature over the recent past, let us say three hours, is relevant. Appropriate models can be specified through so-called *historical* (functional) effects.

- In the simplest case of a *linear effect* that does not change over the day, we have

$$u_j(t) = \ldots + \int_0^h \beta(s) z_j(t-s) ds + \ldots,$$

  where h denotes the time limit to look into the past, e.g., three hours. Typically, it is recommended to choose a rather large value here because, if z in that time region is not relevant for u anymore, $\beta(s)$ will be fit to tend towards zero for $s \to h$ (given there is enough data available to learn from).
- If the effect of z is allowed to *change over the course of the day* $\int_0^h \beta(s) z_j(t-s) ds$ turns into $\int_0^h \beta(s,t) z_j(t-s) ds$.
- In the *nonlinear setting*, we have $\int_0^h f(z_j(t-s), s) ds$ if the (historical) effect is constant across t (i.e., the course of the day) and $\int_0^h f(z_j(t-s), s, t) ds$ otherwise.

All those models fit into the framework of functional additive mixed models as well. Instead of the historical functional effects, we could also write

$$u_j(t) = \ldots + \int_{\mathcal{T}} \beta(s,t) z_j(s) ds + \ldots,$$

which is typically denoted as "linear function-on-function regression." In fact, this is the most popular model for function-on-function regression, and, e.g., used by Centofanti et al. (2021). The corresponding nonlinear/smooth version would be

$$u_j(t) = \ldots + \int_{\mathcal{T}} f(z_j(s), s, t) ds + \ldots.$$

However, both models do not seem sensible approaches for SHM data as considered here because "future" observations $z_j(s)$, with $s > t$, would be allowed to affect "present" $u_j(t)$.