Data and text mining

Extrema-weighted feature extraction for functional data

Willem van den Boom¹,*, Callie Mao¹, Rebecca A. Schroeder² and David B. Dunson¹,*

¹Department of Statistical Science, Duke University, Durham, NC 27708, USA and ²Department of Anesthesiology, Duke University School of Medicine, Durham, NC 27710, USA

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on October 3, 2017; revised on February 21, 2018; editorial decision on February 23, 2018; accepted on February 27, 2018

Abstract

Motivation: Although there is a rich literature on methods for assessing the impact of functional predictors, the focus has been on approaches for dimension reduction that do not suit certain applications. Examples of standard approaches include functional linear models, functional principal components regression and cluster-based approaches, such as latent trajectory analysis. This article is motivated by applications in which the dynamics in a predictor, across times when the value is relatively extreme, are particularly informative about the response. For example, physicians are interested in relating the dynamics of blood pressure changes during surgery to post-surgery adverse outcomes, and it is thought that the dynamics are more important when blood pressure is significantly elevated or lowered.

Results: We propose a novel class of extrema-weighted feature (XWF) extraction models. Key components in defining XWFs include the marginal density of the predictor, a function up-weighting values at extreme quantiles of this marginal, and functionals characterizing local dynamics. Algorithms are proposed for fitting of XWF-based regression and classification models, and are compared with current methods for functional predictors in simulations and a blood pressure during surgery application. XWFs find features of intraoperative blood pressure trajectories that are predictive of postoperative mortality. By their nature, most of these features cannot be found by previous methods.

Availability and implementation: The R package ‘xwf’ is available at the CRAN repository: https://cran.r-project.org/package=xwf.

Contact: willem.van.den.boom@duke.edu or dunson@duke.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Trying to relate functional data with a scalar outcome is a common goal in applications. For instance, images and time series data can be interpreted as functions. One might want to understand what differentiates images that belong to different groups, or what aspects of stock index trajectories are unique to developing versus developed economies. Even though pictures and the stock indices are both functional data, their nature is clearly different, showing the diversity in functional data analysis (FDA) applications and consequently methods.

This article presents a new FDA method motivated by clinical questions regarding blood pressure measurements. Intraoperatively, blood pressure is measured either continuously via invasive monitors or at intervals of 30 s to 3 min. See Figure 1 for example trajectories. The effects of various blood pressure levels are extensively studied. Nonetheless, questions from clinicians remain, especially regarding the relevance of various degrees of blood pressure variability (Mancia and Grassi, 2000).

Variability has been established as an important biomarker (Parati et al., 2013), even though often not fully considered.
For instance, it is suggested that high average blood pressure can be a result of blood pressure variability such that established associations between high blood pressure and certain outcomes might be driven by variability. For instance, Rothwell et al. (2010) found that blood pressure variability can predict stroke better than mean blood pressure, and that the effect of variability interacts with the mean. Sternbach et al. (2002) show that anesthesiologists have a big influence on perioperative blood pressure variability.

It is not clear what features of the dynamic variations in blood pressure are most relevant (Aronson et al., 2010). For example, anesthesiologists want to identify those aspects of the dynamics that predict adverse postoperative outcomes, such as postoperative mortality. The main goal is to identify, in real time, those patients that are experiencing adverse blood pressure dynamics. Anesthesiologists can then intervene and modify the dynamics to hopefully prevent subsequent organ injury. It is thought that abrupt changes in blood pressure, especially when associated with severe systolic hypertension and shear stress, significantly increase this risk (Aronson et al., 2011). However, this is mostly based on poor quality evidence and there is no clear idea what types of dynamics are most or least harmful.

Previous studies have used simple metrics of blood pressure variability, such as the standard deviation of measured values, which ignores the order in which the measurements were obtained, and average real variability (Hansen et al., 2010; Mena et al., 2005). These features capture only a very specific type of blood pressure dynamic. Moreover, they do not allow for blood pressure values at the extremes to be treated differently from those in the normal blood pressure ranges.

Trying to relate functions, here blood pressure trajectories, with an outcome, such as postoperative mortality, is often done in FDA. See, for instance, Ramsay and Silverman (2005). Usually, one extracts features from the functional predictors, which are then used inside a regression or classification model. Popular methods for feature extraction are functional principal components analysis (Jones and Rice, 1992) or basis expansions, which include wavelets (Morris et al., 2003, 2008) and Fourier transforms.

Giraldo and Domínguez (2010) describe feature extraction jointly with dimension reduction to motivate a weighting of features extracted from functional data. Chen et al. (2014) consider weighting of functions for FDA. They focus on distances between functions while we focus on weighted feature extraction. Moreover, while also data-dependent, their weighting functions and their motivation differ from those used here. Hall and Hooker (2016) estimate a truncation of the domain on which the functional parameter in functional linear regression is nonzero. This allows varying domains in a limited fashion. Gellar et al. (2014) propose a variable-domain functional regression which builds upon the generalized functional linear model by allowing smooth changes in the functional parameter as the time domain changes. Despite the large literature, existing FDA methods are clearly unsatisfactory for our motivating blood pressure application, as well as for other related applications.

Most existing approaches assume that the various functions share the same domain, while the time and progression of surgery varies substantially across patients, even when stratiﬁing by type of procedure. One could naively normalize the lengths of surgeries, but such standardization removes valuable information on time and has no clinical justification. The reason to place the curves on a common time domain is that usual FDA methods focus on characterizing shared versus individual-specific components of variability in this trajectory specific to each time. If there are similarities in the dynamics over time (e.g. a spike upwards in the curve), then the curves need to be aligned so that the dynamic features (e.g. the spike) occur at the same time to enable detection. Similar limitations hold for the domain-varying method from Gellar et al. (2014) as the functional linear model with smoothing with respect to changes in the domain can still only detect features that are in some way aligned in the various domains. That method, for instance, reduces to standard functional regression in the absence of variation in the domains of the functional predictors. For the blood pressure data, such time alignment is artificial; instead, we want to characterize similarities and differences in the dynamics in the trajectories even if these dynamics occur at varying times during surgeries of very different lengths.

These shortcomings of existing methods for this application and the interest of clinicians in local effects, with the extremes of particular importance, provide motivation for this article. Our focus is on proposing a new approach for feature extraction for functional predictors that address these issues. Its main building blocks are a flexible definition of variability, or more generally local effects, and a framework that allows for extremes to contribute more than average values. Functions measuring local effects, or variability, are averaged over the domain of the functional predictor. This is a weighted average where the weight is influenced at each point in the domain by how extreme the functional predictor is there. This average for a specific local-effect function is referred to as an extrema-weighted feature (XWF).

1.1 The blood pressure data and the outcome

We have access to the blood pressure data from 80 065 surgeries performed at the Duke University Health System between 2000 and 2014. We consider the mean arterial pressure (MAP). All MAP measurements below 10 and above 250 are discarded as measurement errors. Most measurements are 30 s apart but not all. The methods presented can deal with unevenly spaced measurements but might behave unexpectedly in the presence of large gaps in measurements. Therefore, cases with gaps of >5 min in their blood pressure measurements are removed, also because they potentially involve unusual circumstances. Additionally, we remove blood pressure trajectories that are shorter than 30 min.

The outcome of interest is 30-day postoperative mortality. The medical database includes mortality records which indicate whether someone died within 30 days after surgery. Clinically, certain postoperative complications, such as infection or reduced renal function, are more interesting to be related with intraoperative blood pressures. However, mortality data are more reliable and have less missing data issues than other outcomes, and hence provide a clearer application to test our new method on.

To obtain a more homogeneous population of procedures, we restrict ourselves to non-cardiac surgeries. Other features considered are gender, age and body mass index (BMI). Any cases with these missing, or a BMI above 400, most likely a data error, are excluded.
This leaves us with 18,657 surgeries. Figure 1 shows a couple of these blood pressure trajectories.

2 Materials and methods

This section describes the proposed XWF extraction as well as two alternative approaches we will use for comparison.

2.1 Notation and proposed model

Let \( y_i \) denote the response for subject \( i \), for \( i = 1, \ldots, n \). Let \( x_i : T_i \to \mathbb{R} \), where \( T_i \subset \mathbb{T} \subset \mathbb{R} \), denote a functional predictor for subject \( i \), here the MAP trajectory, with \( T_i \), the time domain across which the predictor is observed for that subject. We restrict ourselves for simplicity to functions that are curves over time, though extensions to functions in more than one dimension are possible. Let \( z_i = (z_{i1}, \ldots, z_{iq})^T \) denote \( q \) other covariates that may be important in predicting \( y_i \). For the blood pressure application, we consider gender, age, BMI and duration of surgery such that \( q = 4 \).

We assume dense measurements of \( x_i(t) \) are available at successive locations \( t_i = (t_{i1}, \ldots, t_{in})^T \) within \( T_i \). In the presence of short blocks of time with intermittent missingness within \( T_i \), we use linear interpolation to fill in the gaps. Although we could estimate our features without this interpolation, simply using the times that are available for a patient, we wish to account for the possibility that the observed trajectories at adjacent times to the missing data block are informative about the occurrence of missing values. One can alternatively use more advanced interpolation or model the measurements as noisy observations of the underlying function \( x_i \), though we do not consider such approaches in this article.

We let \( f \) denote the marginal density of \( x_i(t) \). In the application, \( f \) is the marginal density of blood pressure measurements. We define this marginal density to satisfy \( x_i(t) \sim f \), with \( \{x_i, T_i\} \) sampled from the population and \( t \) sampled uniformly from \( T_i \). In addition, let \( F \) denote the corresponding cumulative distribution function (CDF). Motivated by the application in Section 1.1, we want to define a method for upweighting the importance of local features of \( x_i \) for times \( t \) where \( x_i(t) \) is in the tails of \( f \). In the application, these tails correspond to unusually high or low blood pressure values calibrated relative to the distribution of blood pressure values across individuals having non-cardiac surgical procedures. Let \( u(t) = F(x_i(t)) \), for \( t \in T_i \), denote trajectories that are transformed to have standard uniform marginal density.

2.1.1 Local features

Let \( \psi_j(x_i, t) \), for \( j = 1, \ldots, p \), be functions that characterize different aspects of the local dynamics of the trajectory \( x_i \) at time \( t \). Let us define \( |c| = \max(0, c) \) and \( (c) = \max(0, -c) \). In this article, \( p = 4 \), \( \psi_1(x_i, t) = 1 \), \( \psi_2(x_i, t) = x_i(t) \) as the level of the trajectory at that time point, \( \psi_3(x_i, t) = (x_i'(t))_+ \) as the rate of increase, and \( \psi_4(x_i, t) = (x_i'(t))_- \) as the rate of decrease. Here, we use linear interpolation such that the derivative follows as:

\[
x'_i(t) = \frac{x_i(t_k) - x_i(t_{k-1})}{t_k - t_{k-1}} \quad \text{for} \ t \in (t_{k-1}, t_k].
\]

2.1.2 Extrema weighting

Weight functions \( \omega_j(u) \) on \([0, 1] \) allow differential weights of the local functional dynamic features \( \psi_j \) according to the quantile \( u(t) \). This is motivated by the hypothesis that a specific blood pressure local functional feature affects the outcome in varying degrees depending on whether it appears at low, medium or high blood pressures.

Due to the difference in nature of abnormally high versus low blood pressures, we split \( w \) into two parts; one upweighting low \( \omega_{L2} \) and one upweighting high \( \omega_{H2} \) blood pressures. Specifically, we choose

\[
\omega_{L2}(u) = \omega_{L1}(u) + \omega_{L2}(u)
\]

with

\[
\omega_{L1}(u) = \begin{cases} 
1, & 0 \leq b_{L1} \\
\left(1 - u \right) \frac{2}{1 - b_{L1}}, & u > b_{L1}
\end{cases}
\]

and

\[
\omega_{H2}(u) = \begin{cases} 
1, & u \geq b_{H1} \\
\left(\frac{u}{b_{H2}}\right)^2, & u < b_{H1}
\end{cases}
\]

Here, \( b_{L1} \in (0, 1/2] \) and \( b_{H2} \in [1/2, 1) \) are parameters controlling the weighting near zero and one. Figure 2 contains examples. Weight functions that are more concentrated near zero or one will weight values in the further extremes of the tail more highly. We will estimate \( b_{L1} \) and \( b_{H2} \) based on the data as described in Section 2.1.4.

Supplementary Section S2 compares the results from Section 3.3 with this quadratically decaying choice for \( \omega_j \) to linearly and exponentially decaying weighting. The sensitivity analysis reveals that, while the choice is important, our main finding is consistent across these different choices for \( \omega_j \).

2.1.3 Extrema-weighted features

Employing the aforementioned components, we define the XWFs as

\[
w_{u} = \frac{1}{T_i} \int_{T_i} \omega_j(u(t)) \psi_j(x_i, t) dt
\]

\[
= \frac{1}{T_i} \int_{T_i} \omega_j(F(x_i(t))) \psi_j(x_i, t) dt
\]

\[
= \frac{1}{T_i} \int_{T_i} \omega_{L1}(u(t)) \psi_j(x_i, t) dt + \frac{1}{T_i} \int_{T_i} \omega_{H2}(u(t)) \psi_j(x_i, t) dt
\]

where \( T_i = |T_i| \) is the total time across which the trajectory is observed for individual \( i \). Since this time may be informative about the outcome \( y_i \), we included \( T_i \) as one of the covariates in \( z_i \).

The XWFs have an interpretation as weighted averages. If \( \omega_j(t) = 1 \), then the XWFs are time averages of the \( \psi_j \). Since \( \omega_j \) varies, they become weighted: \( u_{L2} \) is a time average of blood pressure values where low values are weighted more. Similarly, \( u_{H2} \) is a time average of increases in blood pressure where times with high blood pressures are weighted more.

![Fig. 2. Examples of \( \omega_{L2} \) and \( \omega_{H2} \) with \( b_{L1} = 0.3 \) and \( b_{H2} = 0.9 \), respectively](https://example-image-url.com)


Finally, the response $y_i$ is modeled as a generalized additive model (GAM) (Hastie and Tibshirani, 1986; Wood, 2017) with predictors $x_{ik}$ in terms of $u_{1k}$ and $u_{2k}$, and $z_i$. That is, $y_i \sim \Pi(\mu_i)$ independently for $i = 1, \ldots, n$ where $\Pi(\mu_i)$ is a distribution from an exponential family with mean $\mu_i$. For some link function $g$,

$$
g(\mu) = \alpha + \sum_{j=1}^{d} \gamma_j(\xi_j) + \sum_{p} \beta_{1p}(u_{1kp}) + \sum_{q} \beta_{2q}(u_{2kq}) \tag{2}
$$

with $\alpha$ an intercept, $\gamma_1, \ldots, \gamma_d$ (non-linear) functions on the covariates $\xi_1, \ldots, \xi_d$, and $\beta_{1p}, \beta_{2p}, \ldots, \beta_{1p}, \beta_{2q}$ functions on the XWFs. In the blood pressure application, we use logistic regression. That means that $\Pi(\mu)$ is a Bernoulli distribution with success probability $\mu$ and $g$ is the logit function.

### 2.2.4 Estimation algorithm

We estimate the density $f$ through a weighted kernel smoothing approach. In particular, we treat the data as arising from a stratified sampling design, and assign each blood pressure measurement $x_i(t_k)$, $k = 1, \ldots, n_i$, $i = 1, \ldots, n$, sampling weight $1/T_i$, which is inversely proportional to the surgery duration. From an estimate of the density $f$, we can obtain an estimate of the CDF $F$.

The integral in defining the XWFs in (1) is one-dimensional and can be accurately approximated using a variety of numerical methods. Then, once accurate approximations of the feature vectors $u_{1k}$ and $u_{2k}$ are produced, conditionally on the unknown $b_1$ and $b_2$ in the weight functions $w$, we can easily obtain penalized maximum likelihood estimates for the coefficients in (2) using the R package `mgcv`. The smoothness of the estimated functions is determined by minimizing the generalized cross-validation (GCV) score, when the scale parameter is unknown, or the unbiased risk estimator (UBRE) score, when the scale parameter is known such as here in logistic regression (Wood, 2017). Therefore, a grid search can find the $b_1$ and $b_2$ that minimize the GAM UBRE score. We employ an adaptive grid search akin to binary search, detailed in Algorithm 1. We use $M = 3$ in this article such that we get $b_{1s}$ and $b_{2s}$ at a resolution of $2^{-4}$.

### Algorithm 1 Adaptive grid search to optimize the weighting functions

**Initialization:** Set $b_{10} = 0.25$ and $b_{20} = 0.75$ for $j = 1, \ldots, P$.

For $m = 1, \ldots, M$, do the following for $j = 1, \ldots, P$:

1. Compute the UBRE score for when $b_{1j}$ is $2^{1-m}$ larger than its current value, equal to its current value, and when it is $2^{1-m}$ smaller than its current value, while not changing any other parameters. Increase or decrease $b_{1j}$ by $2^{2-m}$ if the first or the third UBRE score, respectively, is the smallest. Otherwise, do nothing.

2. Compute the UBRE score for when $b_{2j}$ is $2^{1-m}$ larger than its current value, equal to its current value, and when it is $2^{1-m}$ smaller than its current value, while not changing any other parameters. Increase or decrease $b_{2j}$ by $2^{2-m}$ if the first or the third UBRE score, respectively, is the smallest. Otherwise, do nothing.

**Return:** Values for $b_{1j}$ and $b_{2j}$, $j = 1, \ldots, P$.

### 2.2 Alternative approaches for comparison

To evaluate the XWF extraction in perspective, we consider two other approaches: A measure from the medical literature, average real variability and a method based on the power spectrum of blood pressure trajectories which is closer to traditional FDA approaches.

#### 2.2.1 Average real variability

Mena et al. (2005) proposed average real variability (ARV) as an improvement over standard deviation as a measure of blood pressure variability. ARV has been used to measure the effect of variability on health outcomes (Hansen et al., 2010). We consider the version from Hansen et al. (2010) that takes into account differences in times between subsequent measurements. That is,

$$
\text{ARV}_i = \frac{1}{t_{i_{n_i}} - t_{i_1}} \sum_{k=2}^{n_i} \left( t_{i_k} - t_{i_{(k-1)}} \right) |x_i(t_{i_k}) - x_i(t_{i_{(k-1)}})|
$$

in our notation.

To compare the XWFs with this more straight-forward measure, we fit the data in a generalized additive model with one of the predictors being the ARV. Specifically, the alternative approach consists of replacing the $2p$ XWFs in (2) by the ARV.

### 2.2.2 Power spectrum

Being able to capture local dynamics rather than global effects is an important characteristic of XWFs. An issue with global effects is that they require some kind of alignment across observations of the functional predictor. By transforming the functional predictor to the frequency domain, one also moves away from this alignment requirement. Therefore, we consider principal components analysis on the power spectrum of the functional predictor.

We compute the power spectrum of each $x_i$ via a fast Fourier transform. Since a spectrum is a function itself, we treat it as a functional predictor. Specifically, we compute its supervised principal components (PCs) (Bair and Tibshirani, 2004; Bair et al., 2006). The supervised PC threshold is determined by cross-validation. As detailed in Ramsay and Silverman (2005, Chap. 8), principal component computation for the spectra, being discretized functions, does not differ from vector-valued data. The three largest PCs replace the XWFs in (2). That is, supervised PCs analysis on the power spectrum of the functional predictor provides features for a generalized additive model.

### 2.3 Inference

Our XWFs algorithm estimates the weighting functions by minimizing the UBRE score. The generalized additive model is then fit with the resulting XWFs as predictors. The $P$-values that the GAM provides (Wood, 2013) however do not take into account any estimation uncertainty in the weighting functions. To obtain principled measures of statistical significance, we obtain $P$-values via a permutation test where the weighting functions are re-estimated for each permutation. We reorder the elements in the outcome vector $y$ at random, fit the XWF model to the resulting data, and record the reported $P$-values. By construction, any association between the predictors and the outcome $y$ in these data is by chance. Also, the recorded $P$-values are sample statistics. Let us denote them by $P^s$, $s = 1, \ldots, S$, with $S$ the number of permutations. Let $P^D$ be the $P$-value from the original data. Then, we compute the $P$-value as:
which is conservatively very slightly greater than the exact $P$-value of a permutation test (Phipson and Smyth, 2010). We use $S = 100$.

For consistency, the $P$-values for the alternative approaches, the ARV and the power spectrum, are obtained in the same fashion. An added benefit of this approach is that it does not require assumptions on the data generating process for the inference to be valid.

3 Results

This section starts with two simulation studies to understand the performance and interpretation of all three methods considered. It concludes with the results from the blood pressure application. Supplementary Section S1 contains additional simulations.

3.1 Frequency simulation

Let $n = 200$, and, for $i = 1, \ldots, n$, $x_{it} \sim \text{Uniform}(0, 1)$ independently for $i = 1, \ldots, n$. The functional predictors are defined by:

$$x_i(t) = \begin{cases} \sin \left( \frac{\pi t}{30} \right), & t \leq t_i \\ \frac{\pi (t_i + (t - t_i) \phi_i)}{30} + 2, & t \in (t_i, t_i + 30] \\ \sin \left( \frac{\pi (t_i + 30 \phi_i + t)}{30} \right), & t > 30 \end{cases}$$

for $i = 1, \ldots, n$. That is, the $x_i$ have a fixed frequency and are centered at zero except for a randomly sampled interval of size 30 in their domain where they have a higher, random frequency $\phi_i$ and are centered at two. Let $q = 2$ in (2) and sample the $z_{iq}$ i.i.d. from a standard normal distribution. The outcome $y_i \in \{0, 1\}$ is sampled according to

$$P(y_i = 1) = \frac{1}{1 + \exp \left( \text{median} \left( \phi_i \right)^n_{i=1} - \beta \right)},$$

independently for $i = 1, \ldots, n$.

Supplementary Table S1 contains the $P$-values for the XWFs, ARV and the power spectrum. The $P$-values for the $z_{iq}$ were not significant and were left out. Figure 3 contains the principal components used in the power spectrum method. The weighting functions of the statistically significant XWFs (with $P < 0.05$) were estimated as $b_{R1} = 0.84$, $b_{L2} = 0.03$, $b_{R3} = 0.53$ and $b_{R4} = 0.75$. The corresponding smooth GAM functions are shown in Figure 4. The confidence region is from the GAM and ignores uncertainty in our estimation of the weighting functions. The horizontal axes range from the 1st to the 99th percentile of the XWF.

Since ARV increases with $\phi_i$, ARV finds a trend between variability and the outcome in the simulated data. Similarly, $\phi_i$ directly relates to the frequencies in the power spectrum. Therefore, the power spectrum method also finds an effect. However, it is hard to interpret the first spectrum principal component, even when knowing the data-generating process.

Extrema-weighted feature selection picks up that variability at higher values increases $P(y_i = 1)$: The magnitude of the derivative of $x_i$ is proportional to $\phi_i$, if $x_i(t) > 1$. The XWFs $w_{R3}$ and $w_{R4}$ measure the magnitude of the derivative since $\psi_L(x_i, t) = (x_i(t))^\prime$ and $\psi_R(x_i, t) = (x_i(t))^\prime$. What happens at higher values of $x_i(t)$ contributes more to the values of $w_{R3}$ and $w_{R4}$. According to the data-generating process, $y_i$ is positively associated with $\phi_i$. Figure 4 shows that $y_i$ is positively associated with $w_{R3}$ and $w_{R4}$. This association is thus reflective of the data-generating process. The other two significant XWFs, $w_{R1}$ and $w_{L2}$, do not vary with $\phi_i$. Their association with $y_i$ as implied by their statistical significance is not consistent with the true model where $y_i$ depends on $\phi_i$ only. The results from the XWF-based analysis yield observations consistent with the underlying data-generating process via $w_{R3}$ and $w_{R4}$ with richer and clearer interpretations than the alternative methods.

3.2 XWF simulation

Let $n = 1000$ and $n_i = 50$ for $i = 1, \ldots, n$. The $z_{iq}$ in (2) and $t$ are defined analogous to the frequency simulation from the previous section. We draw $\phi_i$ and $\mu_i$ from a uniform distribution on (0.9, 0.99) and (2(1 - $\phi_i$), 4(1 - $\phi_i$)), respectively, independently for $i = 1, \ldots, n$. Sample the measurements of $x_i$ as follows: Set $x_i(1) = \frac{1}{1 - \phi_i}$ and $x_i(k) \sim N(\mu_i + \phi_i \cdot x_i(k - 1), 0.05^2)$, $k = 2, \ldots, n_i$, independently for $i = 1, \ldots, n$. We set $b_{R1} = 0.2$ and $b_{R3} = 0.7$ and compute the corresponding $w_{R1}$ and $w_{R3}$. Set $\theta_i = w_{R1}/\sigma_{R1} + w_{R3}/\sigma_{R3}$, where $\sigma_{R1}$ and $\sigma_{R3}$ are the sample standard deviations of $w_{R1}$ and $w_{R3}$, respectively. Then, the outcome $y_i \in \{0, 1\}$ is sampled independently for $i = 1, \ldots, n$ according to

$$P(y_i = 1) = \frac{1}{1 + \exp \left( \text{median} \left( \theta_i \right)^n_{i=1} - \theta_i \right)},$$

Supplementary Table S2 of the contains the $P$-values for the three approaches under consideration. As in Section 3.1, the $P$-values for the $z_{iq}$ were not significant and were left out. Figure 5 contains the supervised principal components used in the power spectrum method. The weighting functions of the statistically significant XWFs were estimated as $b_{L1} = 0.03$, $b_{R3} = 0.72$ and $b_{R4} = 0.25$. The corresponding smooth GAM functions are shown in Figure 6.

Average real variability measures variation which is associated with $y_i$ via $w_{R3}$. As a result, ARV finds a trend. The supervised principal components fail to find a signal with the frequency spectrum dominated by the autoregressive process: The autoregressive parameters do not influence $y_i$ directly. Figure 6 shows that $w_{R1}$ and $w_{R3}$
capture most of the variation in $y$. This is consistent with how $y$ was simulated from $w_{1,1}$ and $w_{1,3}$. Also, the estimates for $b_{1,1}$ and $b_{1,3}$ are close to the values used for simulation.

### 3.3 Blood pressure data

Table 1 contains the results for the three methods when applied to the blood pressure data. The $P$-values for the $z_i$s were omitted, being all.01. Figure 7 contains the supervised principal components used in the power spectrum method. For computational stability of the supervised principal components, we truncated the spectra at a frequency of a 1000. The weighting function of the statistically significant XWF was estimated as $b_{1,2} = 0.03$. The corresponding smooth GAM function is shown in Figure 8.

Figure 9 shows the blood pressure trajectories with the most extreme values for the XWFs, out of all 18 657 trajectories. It is striking how the trajectories vary in length: One of the goals of the XWFs was to be well-defined over, and be comparable across, functions with greatly varying sizes of domain. While the XWFs are defined in an interpretable fashion, their values are not always readily gleaned from a trajectory plot. It is for instance hard to eyeball that the black trajectory has a higher XWF value than the gray trajectory for $w_{1,2}$ and $w_{1,3}$ in Figure 9. However, that $w_{1,1}$ and $w_{1,3}$ measure difference in level is clearly visible with the black and gray trajectories separated in terms of average value. Similarly, both trajectories plotted for $w_{1,3}$ and $w_{1,4}$ show a clear difference in variability.

Average real variability and the power spectrum method find an association between the blood pressure trajectories and postoperative mortality. However, only the XWFs reveal that low blood pressure values are predictive. Figure 8 shows that higher and more blood pressure values at the lower extreme ($w_{1,2}$) are associated with increased mortality. This interpretation is different from how one would interpret ARV. This shows the importance of considering richer features than ARV.

#### 3.3.1 Relation with existing findings

Our result for ARV is consistent with previous literature that showed that an increased ARV is associated with adverse outcomes (Hansen et al., 2010; Mena et al., 2005). ARV is, to our knowledge, the most advanced method used previously to assess the effect of the blood pressure variability.

As XWFs provide richer insights, they cannot be contrasted one-to-one with former findings. We can instead look at how the richer insights from the XWF results fit into previously found associations. Having a larger proportion of blood pressure values in the extremes was associated with mortality. Extreme values are more common in the presence of high blood pressure standard deviation or ARV which previous study also found to be associated with increased mortality (Parati et al., 2013).

#### 3.3.2 Predictive performance

Using specific predefined features instead of general methods that are less constrained in how they capture variability in the data, might hurt the overall explanatory power. To assess whether XWFs suffer from this in the motivating application, we look at predictive performance. We randomly select 1000 cases, 100 with $y_{1} = 1$ and 900 with $y_{1} = 0$, as test data leaving training data of size 17 657. We compute the XWFs, ARV and supervised principal components as before on the training data. Three predictive models are considered: A GAM with the $z_i$s and only XWFs, with XWFs and ARV, and with XWFs and supervised principal components. We compare the performance of these on the test data to assess any loss in explanatory power by being constrained to XWFs.

Area under the curve (AUC) of the receiver operating characteristic quantifies predictive performance. Using 10 repeated random splits of the data into training and test set, we obtain 10 AUCs for each of the three models considered. They range from 0.56 to 0.64 while the absolute differences in AUC between the models are <0.01, none of which are statistically significant. Additionally, Supplementary Section S3 similarly does not show a notable difference in the precision-recall curves. The usage of the more
interpretable predefined XWFs does not seem to lessen explanatory power compared to approaches which have a greater focus on predictive performance and less on interpretability.

4 Discussion

Existing functional feature extraction methods focus on functions that share the same domain and mostly extract global features. Motivated by rich blood pressure data from surgeries, we introduced a new feature extraction method, called XWFs, allowing for differing domains while aiming at functional features that are defined in a local nature, like variability. We compared XWFs to ARV, a functional feature from the medical literature, and to supervised principal components analysis on the power spectrum of the functional predictor.

Simulations reveal that XWFs are not only more easily interpretable due to the way they are defined, but they are also more flexible in detecting associations that do not fit traditional models. This was instrumental in finding a relationship between blood pressure trajectories during surgery and the outcome postoperative mortality. Whereas the power spectrum principal components detected a hard cutoff at both extremes of the observed blood pressure range are associated with increased mortality. The XWF results directly address questions of interest to the clinician collaborators. Despite their predefined and specific nature, XWFs had similar explanatory power as competing methods.

This article presented a set of well-defined features applicable to the data at hand. The ideas behind them and the general framework of allowing varying domains in the functional predictors and taking weighted averages of local phenomena can however be applied in a larger variety of ways, making the methods from this article widely applicable to functional data.

Many parts of the XWF model provide choices to adapt the method to other applications. Weighting functions can be chosen and parametrized as desired. Similarly, the generalized additive model shows one way to use XWFs but the features can be broadly used within different models and machine learning methods.

We presented a novel functional feature extraction method motivated by modeling the effect of intraoperative blood pressure trajectories on postoperative mortality. The XWFs proposed are by design suited for functional predictors whose domains vary. The XWFs outperform more traditional methods and FDA tools, both in simulations and in the application of interest. The proposed model constitutes a modifiable framework for interpretable functional feature extraction.

Acknowledgements

We would like to thank the anonymous referees for suggestions that substantially improved the article and Dr Michael W. Manning and Dr Tracy L. Setji for their discussions of the medical data.

Funding

This work was supported by Accenture PLC.

Conflict of Interest: none declared.

References

Aronson, S. et al. (2010) Intraoperative systolic blood pressure variability predicts 30-day mortality in aortocoronary bypass surgery patients. Anesthesiology, 113, 305–312.
Aronson, S. et al. (2011) Does perioperative systolic blood pressure variability predict mortality after cardiac surgery? An exploratory analysis of the ECLIPSE trials. Anesth. Analg., 113, 19–30.
Bair, E. and Tibshirani, R. (2004) Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol., 2, e108.
Bair, E. et al. (2006) Prediction by supervised principal components. J. Am. Stat. Assoc., 101, 119–137.
Chen, H. et al. (2014) Optimally weighted $L^2$ distance for functional data. Biometrics, 70, 516–525.
Gellar, J.E. et al. (2014) Variable-domain functional regression for modeling ICU data. J. Am. Stat. Assoc., 109, 1425–1439.
Giraldo, L.G.S. and Domínguez, G.C. (2010) Weighted feature extraction with a functional data extension. Neurocomput., 73, 1760–1773.
Hall, P. and Hooker, G. (2016) Truncated linear models for functional data. J. R. Stat. Soc. Series B (Stat. Methodol.), 78, 637–653.
Hansen, T.W. et al. (2010) Prognostic value of reading-to-reading blood pressure variability over 24 hours in 8938 subjects from 11 populations. Hypertension, 55, 1049–1057.
Hastie, T. and Tibshirani, R. (1986) Generalized additive models. Stat. Sci., 1, 297–310.
Jones, M.C. and Rice, J.A. (1992) Displaying the important features of large collections of similar curves. Am. Stat., 46, 140.
Mancia, G. and Grassi, G. (2000) Mechanisms and clinical implications of blood pressure variability. J. Cardiovasc. Pharmacol., 34 (7 Suppl 4), S13–19.
Mena, L. et al. (2005) A reliable index for the prognostic significance of blood pressure variability. J. Hypertens., 23, 505–511.
Morris, J.S. et al. (2003) Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. J. Am. Stat. Assoc., 98, 573–583.
Morris, J.S. et al. (2008) Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. Biometrics, 64, 479–489.

Parati, G. et al. (2013) Assessment and management of blood-pressure variability. Nat. Rev. Cardiol., 10, 143–155.

Phipson, B. and Smyth, G.K. (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol., 9, Article 39. doi: 10.2202/1544-6115.1585.

Ramsay, J. and Silverman, B.W. (2005) Functional Data Analysis. 2nd edn. Springer-Verlag, New York.

Rothwell, P.M. et al. (2010) Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. Lancet, 375, 895–905.

Sternbach, Y. et al. (2002) Hemodynamic benefits of regional anesthesia for carotid endarterectomy. J. Vasc. Surg., 35, 333–339.

Wood, S.N. (2013) On P-values for smooth components of an extended generalized additive model. Biometrika, 100, 221–228.

Wood, S.N. (2017) Generalized Additive Models: An Introduction with R. 2nd edn (Chapman & Hall/CRC Texts in Statistical Science). Chapman and Hall/CRC, Boca Raton, Florida.