Taylor & Francis
Taylor & Francis Group

Wavelet Feature Screening

Rodney Fonseca[a] ⬤, Pedro Morettin[b] ⬤, and Aluísio Pinheiro[c] ⬤

[a]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel; [b]Institute of Mathematics and Statistics, University of São Paulo, São Paulo-SP, Brazil; [c]Department of Statistics, University of Campinas, Campinas, SP, Brazil

ABSTRACT

An initial screening of which covariates are relevant is a common practice in high-dimensional regression models. The classic feature screening selects only a subset of covariates correlated with the response variable. However, many important features might have a relevant albeit highly nonlinear relation with the response. One screening approach that handles nonlinearity is to compute the correlation between the response and nonparametric functions of each covariate. Wavelets are powerful tools for nonparametric and functional data analysis but are still seldom used in the feature screening literature. We propose a wavelet feature screening method that can be easily implemented, and we prove that, under suitable conditions, it captures the true covariates with high probability. Simulation results also show that our approach outperforms other screening methods in highly nonlinear models. We apply feature screening to two datasets about ozone concentration and epilepsy. In both applications, the proposed method selects features that match findings in the literature of their respective research fields, illustrating the applicability of feature screening. Supplementary material for this article is available online.

1. Introduction

The number of explanatory variables is commonly much larger than the sample size in modern data science. In the current era of big data, many applications in biology have thousands of covariates, reaching even hundreds of thousands in genetic studies (Fan et al. 2020, sec. 1.1). A key objective in applications is to use all data to determine which features/covariates are associated with characteristics of interest, that is, the response variables. Some classic methods to deal with this type of problem where $p \gg n$ are sub-sampling covariates, ridge regression, and Lasso, for example (Hastie, Tibshirani, and Friedman 2009). However, if p is substantially larger than n, the performance of many variable selection methods may also be compromised (Fan and Lv 2008). In such scenarios, practitioners often do an initial screening to discard the covariates most likely irrelevant to the analysis. This approach works as an initial step in supervised learning. It reduces the problem's dimension to only $p_n \ll p$ covariates and facilitates the learning step of a regression or classification model.

Feature screening consists of computing a marginal utility measure for each feature and selecting the ones with the highest scores. The most popular utility measure is the Pearson correlation between the response variable and each covariate. Fan and Lv (2008) studied this approach and derived its theoretical guarantees. Since Pearson correlation is tailored to linear relationships, further works proposed using nonparametric methods to capture nonlinear relations. Hall and Miller (2009) proposed

using a generalized correlation, where the utility measure is the Pearson correlation of the response variable and an estimated smooth function of the covariate. Likewise, Fan, Feng, and Song (2011) proposed to fit regression splines of each feature to the response and use the norm of the fitted function as the utility score. Many other screening methods have been proposed since then; to name a few examples, we cite the following: model-free screening (Zhu et al. 2011; Xue and Liang 2017), screening based on distance-correlations (Li, Zhong, and Zhu 2012), screening for categorical response (Mai and Zou 2013), conditional marginal screening (Barut, Fan, and Verhasselt 2016), and screening for dependent data (Yousuf 2018). Fan et al. (2020, chap. 8) provides a thorough literature review and further details about feature screening. Our article focuses on nonparametric utility measures (Hall and Miller 2009).

1.1. Generalized Correlation

Let Y be a response variable of interest and $(X_1, \ldots, X_p)^T$ a set of covariates to be used as proxies to Y. Assume we have a sample $\{(\mathbf{X}_{i.}, Y_i)\}_{i=1}^n$ where $\mathbf{X}_{i.} = (X_{i1}, \ldots, X_{ip})$ and $p \gg n$. We aim to reduce the data's dimension by selecting a subset of covariates relevant to the study. Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$ be the design matrix, where $\mathbf{X}_j = (X_{1j}, \ldots, X_{nj})^T$ denotes the observations from the jth covariate.

The standard screening approach uses the Pearson correlation between $\mathbf{Y}$ and $\mathbf{X}_j$ as a utility measure. This approach is widely used by practitioners in all fields that require data analysis

and counts with a solid theoretical foundation (Fan and Lv 2008). These two aspects explain the popularity of this screening approach. However, the Pearson correlation does not capture covariates with a highly nonlinear relation with the response variable. Therefore, Hall and Miller (2009) proposes to tackle this problem by estimating the functional relations between Y and each X_j separately. They compute a generalized correlation as follows

$$\rho(X_j, Y) = \sup_{f_j \in \mathcal{H}} \frac{\text{cov}\{f_j(X_j), Y\}}{\sqrt{\text{var}\{f_j(X_j)\}\text{var}(Y)}}, \quad j = 1, \ldots, p,$$

where $\mathcal{H}$ is some functional space and $f_j \in \mathcal{H}$ is the function that fits X_j to Y. When $\mathcal{H}$ consists of the space of linear functions, $\rho(X_j, Y)$ becomes the Pearson correlation. Hall and Miller (2009) uses cubic splines basis to represent f_j. Given the sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$, the generalized correlation can be estimated by

$$\hat{\rho}(\mathbf{X}_j, \mathbf{Y}) = \sup_{f_j \in \mathcal{H}} \frac{\sum_{i=1}^n \{f_j(X_{ij}) - \bar{f}_j\}\{Y_i - \bar{Y}\}}{\sqrt{\sum_{i=1}^n \{f_j(X_{ij}) - \bar{f}_j\}^2 \sum_{i=1}^n \{Y_i - \bar{Y}\}^2}},$$
$$j = 1, \ldots, p,$$

where $\bar{f}_j = \sum_{i=1}^n f_j(X_{ij})/n$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$. The idea is to use $\hat{\rho}(\mathbf{X}_j, \mathbf{Y})$ to measure the importance of $\mathbf{X}_j$ for fitting $\mathbf{Y}$. The screening approach selects only $p_n \ll p$ covariates with the highest $\hat{\rho}(\mathbf{X}_j, \mathbf{Y})$ in absolute value.

The same screening results can be achieved using $|\|\hat{f}_j\| - \|f_j\||$ as utility measure, where $\|\cdot\|$ denotes some norm. For example, consider the case of Pearson correlation and let $\mathbf{X}$ and $\mathbf{Y}$ be two vectors with zero-mean and unit norm. Since $\|\mathbf{X} - \mathbf{Y}\|^2 = 2 - \langle \mathbf{X}, \mathbf{Y} \rangle$, by maximizing the correlation $\langle \mathbf{X}, \mathbf{Y} \rangle$ we also minimize $|\|\mathbf{X}\| - \|\mathbf{Y}\||^2 \leq \|\mathbf{X} - \mathbf{Y}\|^2$, where the latter follows from the triangular inequality. Therefore, we will focus on the distance $|\|\hat{f}_j\| - \|f_j\||$ to devise a variable selection scheme.

1.2. Proposed Method

We use wavelet methods to estimate the function f_j that links Y to X_j, $j = 1, \ldots, p$. Wavelets form a basis for the space of square-integrable functions, similarly to cosines and sines in Fourier analysis. The main advantages of wavelets are their local adaptivity and regularization properties, making them a powerful tool to represent functional data (Morettin, Pinheiro, and Vidakovic 2017). Other properties are that wavelet coefficients can be quickly computed for equally spaced sampled observations (regular design), and such coefficients characterize quite general functional spaces, such as Hölder and Besov spaces, for example (Vidakovic 1999, p. 87). One challenge in our problem is that multiple covariates are generally irregularly spaced, a design for which often-used wavelet methods are unsuitable.

Our main contribution is a feature screening method that uses *warped wavelets*, which were introduced by Kerkyacharian and Picard (2004) to deal with random covariates. Our screening idea is described in Algorithm 1. It can be easily applied within any statistical software that computes wavelet transforms (e.g., R, Python, Matlab). The paradigm of wavelet feature screening has been suggested previously (Fan, Feng, and Song 2011, p. 552). The closest work to ours is the paper by Zhao, Chen, and

---

Algorithm 1 Warped wavelet screening

Input: response variable $\mathbf{Y} \in \mathbb{R}^n$, features $\mathbf{X}_1, \ldots, \mathbf{X}_p \in \mathbb{R}^n$, threshold $\nu > 0$, wavelet basis.
Output: Indices of screened features $\hat{\mathcal{M}} \subset \{1, \ldots, p\}$.
1: **for** $j = 1, \ldots, p$ **do**
2:      Sort $(Y_{\pi(1)}, X_{j,\pi(1)}), \ldots, (Y_{\pi(n)}, X_{j,\pi(n)})$ such that $X_{j,\pi(1)} \leq \cdots \leq X_{j,\pi(n)}$, where $\{\pi(1), \ldots, \pi(n)\}$ denotes a permutation of $\{1, \ldots, n\}$.
3:      Apply the discrete wavelet transform to $(Y_{\pi(1)}, \ldots, Y_{\pi(n)})$ using the chosen wavelet basis, obtaining wavelet coefficients $\{\hat{\beta}_k\}$.
4:      Apply thresholding on the wavelet coefficients, obtaining $\{\tilde{\beta}_k\}$.
5:      Compute $w_j = \{\sum_k \tilde{\beta}_k^2\}^{1/2}$ using the thresholded coefficients.
6: **end for**
7: Set $\hat{\mathcal{M}} = \{j; 1 \leq j \leq p, w_j > \nu\}$.
8: **return** $\hat{\mathcal{M}}$

Ogden (2015), who proposes a screening step for a wavelet-based Lasso estimator of a linear regression model. However, this work assumes that predictors are sampled at equally spaced points, a condition hardly satisfied in practice. The current article is the first to propose a wavelet algorithm that handles irregular designs and to study its properties. The main component of our algorithm is the application of warped wavelets, whose details are presented in Section 2. Warped wavelet regression estimators adapt well to features with irregular sampling designs (Morettin and Porto 2022), which allows us to apply wavelets to run feature screening. Theoretical properties of the proposed screening approach are presented in Section 3, where we prove that Algorithm 1 can capture the true covariates with high probability.

Feature screening methods are commonly tested with simulations of scenarios where the link function f_j is either smooth or has very few discontinuities. Differently, Section 4 presents simulations of challenging scenarios with highly nonlinear examples of f_j. Our results show that model-free screening methods do not perform well when f_j is very irregular, but Algorithm 1 can still detect true features in many such cases. The wavelet approach also outperforms splines for scenarios where f_j contains local traits like peaks. This analysis illustrates the well-known advantage of wavelets for regression problems where the link function is irregular, containing spikes or discontinuities (Reiss et al. 2017, sec. 2.1).

Another contribution of this article is the application of feature screening to two scientific problems. The first uses ozone concentration data in the largest city in Brazil. Ozone is a pollutant related to various health problems; thus, monitoring which atmospheric features are related to ozone concentration is of practical interest (see Section 5). The second problem uses data of a study about temporal lobe epilepsy. These data were collected to understand how brain white matter measurements are related to brain degeneracy in epileptic patients (see Section 6). In both applications, we show that feature screening can detect covariates previously reported as important factors in their respective fields. The supplementary material contains an

application with the cardiomyopathy data (Segal, Dahlquist, and Conklin 2003), a high-dimension example commonly used in the feature screening literature. Some final remarks in Section 7 close the article.

Notation: We use log for the natural logarithm. Vectors and matrices are represented by bold letters. The ith row and jth column of a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ are denoted as $\mathbf{X}_{i.} \in \mathbb{R}^p$ and $\mathbf{X}_j \in \mathbb{R}^n$, respectively. The Euclidean norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is denoted as $\|\mathbf{x}\| = (\sum_{i=1}^n x_i^2)^{1/2}$. For a random function $f : \mathbb{R} \to \mathbb{R}$, we define the L^2 and L^∞-norms in terms of the expected values $\|f\|_{L^2}^2 = \mathrm{E}\left\{\int |f(x)|^2 dx\right\}$ and $\|f\|_{L^\infty} = \mathrm{E}\left\{\max_{x \in \mathbb{R}} |f(x)|\right\}$, respectively. These norms reduce to their usual definitions when f is not random.

2. Warped Wavelets

Kerkyacharian and Picard (2004) introduces warped wavelets to allow randomly sampled covariates in wavelet regression estimators. We employ this method to perform feature screening, separately analyzing each pair (X_j, Y). The data for the jth covariate is $\{(X_{ij}, Y_i)\}_{i=1}^n, j = 1, \ldots, p$. We fit the following model for each covariate:

$$Y_i = f_j(X_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n, \tag{1}$$

where $\epsilon_{1j}, \ldots, \epsilon_{nj} \overset{\text{iid}}{\sim} N(0, \sigma^2)$, with independent X_{ij} and ϵ_{ij}. We assume that $f_j \in L^2([a, b])$, and one of our goals is to estimate the coefficients of its representation in a wavelet basis. For convenience, let us focus on the case where $\sigma^2 = 1$, $a = 0$ and $b = 1$.

A common assumption in wavelet regression is that $X_{ij} = i/n$, that is, the covariates are equally spaced in the support of f_j. This assumption can be relaxed with warped wavelets, which allow X_{ij} to be irregularly spaced on the interval $[0, 1]$. Assume that X_{ij} follows some continuous distribution with density g_j and distribution function G_j. Without loss of generality, consider that $X_{1j} < X_{2j} < \cdots < X_{nj}$ and denote their empirical distribution function by $\hat{G}_j$. The idea of warped wavelets is to recover the usual regular design using the fact that $\hat{G}_j(X_{ij}) = i/n$. Plugging $X_{ij} = \hat{G}_j^{-1}(\hat{G}_j(X_{ij}))$ in (1), we have

$$Y_i = f_j(X_{ij}) + \epsilon_{ij} = f_j\left(\hat{G}_j^{-1}(\hat{G}_j(X_{ij}))\right) + \epsilon_{ij}$$
$$= f_j\left(\hat{G}_j^{-1}(i/n)\right) + \epsilon_{ij},$$

which resumes the classical equally spaced design when looking at the function composition $f_j\left(\hat{G}_j^{-1}(\cdot)\right)$. We shall focus on the estimators that take G_j as known to avoid some technicalities. Estimators that use the empirical distribution function $\hat{G}_j$ are also proposed by Kerkyacharian and Picard (2004) and have similar properties to the estimator computed assuming G_j known. Kerkyacharian and Picard (2004) show that both approaches attain (up to logarithmic factors) the minimax convergence rate from regular designs.

Another idea of warped wavelets is to take the wavelet decomposition of $f_j\left(G_j^{-1}(\cdot)\right)$ instead of $f_j(\cdot)$. Consider a wavelet basis $\{\psi_{lk}; l \geq -1, k \in \mathbb{Z}\}$ where $\psi_{-1,k}$ is a scaling function

(Härdle et al. 1998, eq. 3.4). Expanding $f_j\left(G_j^{-1}(.)\right)$ in this basis we get

$$f_j\left(G_j^{-1}(y)\right) = \sum_{l \geq -1} \sum_{k \in \mathbb{Z}} \beta_{lk} \psi_{lk}(y), \quad \text{or equivalently}$$

$$f_j(x) = \sum_{l \geq -1} \sum_{k \in \mathbb{Z}} \beta_{lk} \psi_{lk}\left(G_j(x)\right),$$

where the wavelet coefficients are given by

$$\beta_{lk} = \int_0^1 \psi_{lk}\left(G_j(x)\right) f_j(x) g_j(x) dx. \tag{2}$$

Here we omit the index j from β_{lk}, but later we shall write β_{jlk} to let it clear it is a coefficient from f_j. Equation (2) reduces to the usual wavelet coefficient when $g_j(x)$ is the uniform density. In general, the properties of the warped wavelet transform depend on weights defined by $g_j(x)$ and how far they are from uniform weights. We shall assume that g_j is bounded as follows:

C1 There is a finite constant $C > 0$ such that $C^{-1} < g_j(x) < C$ for all $x \in [0, 1]$ and all $j \in \{1, \ldots, p\}$.

Under C1, the density g_j in (4) defines a weighting scheme that belongs to a class known as *Muckenhhoupt weights* (Kerkyacharian and Picard 2004, sec. 4.1). This condition is sufficient to guarantee good convergence rates for warped wavelet estimators. Moreover, C1 has already been used for nonparametric feature screening (Fan, Feng, and Song 2011, condition B). We also assume that the following conditions hold:

C2 The scaling function $\psi_{-1,k}$ and wavelet functions $\psi_{lk}, l \geq 0$, have compact support.

C3 The functions $f_j : [0, 1] \to \mathbb{R}, j = 1, \ldots, p$, are periodic and there is an absolute constant $C > 0$ such that $\|f_j\|_{L^\infty} < C$ for all j.

These are common assumptions for warped wavelets (Porto et al. 2016; Gómez, Porto, and Morettin 2021). C2 allows us to write the wavelet transform as finite sums, and it is satisfied by Daubechies wavelets, for example (Härdle et al. 1998, p. 127). The periodicity assumed in C3 permits us to apply the wavelet transform for functions with support in $[0, 1]$. See (Vidakovic 1999, sec. 5.6) for other transforms of signals in an interval. This assumption also justifies periodization heuristics frequently used in practice to deal with boundary handling and sample sizes that are not a power of two (Ogden 1997). The assumption that f_j is bounded is technical and is used later in Proposition 1.

We consider the warped wavelet estimator given by

$$\hat{f}_j(x) = \sum_{l=-1}^{J} \sum_{k=0}^{2^l-1} \hat{\beta}_{lk} I\{|\hat{\beta}_{lk}| \geq \kappa t_n\} \psi_{lk}\left(G_j(x)\right),$$

$$t_n = \left(\frac{\log n}{n}\right)^{1/2}, \tag{3}$$

where $\kappa > 0$ is a tuning parameter and J satisfies $2^J \leq C t_n^{-1}$, for some absolute constant $C > 0$. The empirical coefficients are

$$\hat{\beta}_{lk} = \frac{1}{n} \sum_{i=1}^n \psi_{lk}(G_j(X_{ij})) Y_i. \tag{4}$$

Note that $\hat{f}_j(x)$ in (3) is computed via hard thresholding. Other thresholding approaches can also be applied (Vidakovic 1999, chap. 6). See Fan and Li (2001) for a detailed analysis of different thresholding options, such as SCAD and soft thresholding. The following result, due to (Kerkyacharian and Picard 2004, Prop. 3), will be useful to deduce theoretical guarantees for the warped wavelet feature screening.

Proposition 1. (Kerkyacharian and Picard 2004, p. 1070) Assume that conditions C1–C3 hold and let $\alpha > 1$. Then, there exist constants C_α and C'_α, and for any $\gamma > 0$ there exists a constant κ_γ, with

$$E(|\hat{\beta}_{lk} - \beta_{lk}|^\alpha) \le C_\alpha \frac{1 + \|f_j\|_{L^\infty}^\alpha}{n^{\alpha/2}}, \quad \text{for } 2^l \le n,$$

$$P\left(|\hat{\beta}_{lk} - \beta_{lk}| > \kappa \sqrt{\frac{\log n}{n}}\right) \le C'_\alpha n^{-\gamma\alpha},$$

$$\text{for } \kappa \ge \kappa_\gamma, \ 2^l \le \sqrt{\frac{n}{\log n}}.$$

3. Feature Screening via Warped Wavelet Regression

In this section, we use the properties of the warped wavelets estimator to establish theoretical guarantees for our Algorithm 1. Assume we can fit model (1) to all covariates but that f_j is null for most of them, that is, only a few X_j are relevant to predict Y. Let $\boldsymbol{\beta}_j = \{\beta_{jlk}, 0 \le l \le J, 0 \le k \le 2^l - 1\}$ be the wavelet coefficients of f_j as in (2). Notice that if f_j is null, then $\boldsymbol{\beta}_j = \mathbf{0}$. Therefore, we represent the set of relevant features in terms of their wavelet coefficients as $\mathcal{M}_* = \{j; 1 \le j \le p, \|\boldsymbol{\beta}_j\| > 0\}$.

Our goal is to find a set $\hat{\mathcal{M}} \subset \{1, \dots, p\}$ that contains $\mathcal{M}_*$ with high probability and has cardinality much lower than p. However, the functions f_j are unknown, and we must use estimators $\hat{f}_j$ to screen the relevant features. This approximation works well if the wavelet transform of $\hat{f}_j$ is close enough to that of f_j. Let $\hat{\boldsymbol{\beta}}_j$ be the wavelet coefficients of $\hat{f}_j$ estimated as in (4). Then, the proposed set of screened variables is as follows:

$$\hat{\mathcal{M}} = \left\{j; \|\hat{\boldsymbol{\beta}}_j\| > \nu_n, 1 \le j \le p\right\},$$

where $\nu_n > 0$. The key point to recover $j \in \mathcal{M}_*$ is that the energy of f_j must be higher than the noise level, that is, $\|\hat{\boldsymbol{\beta}}_j\|$ has to be higher than the norm of wavelet coefficients of pure noise. In this case, we can screen the correct variables with high probability if we choose a suitable ν_n. The next proposition formalizes this idea.

Proposition 2. Let κ, C' and γ be defined as in Proposition 1, with $\gamma > 1/4$. Assume that C1–C3 hold and that $\min_{j \in \mathcal{M}_*} \|\boldsymbol{\beta}_j\| > 2\nu_n$, where $\nu_n = 2^{(J+1)/2}\sqrt{\kappa}((\log n)/n)^{1/4}$. If $\hat{\mathcal{M}}$ is computed with this ν_n, then

$$P\left(\hat{\mathcal{M}} \supset \mathcal{M}_*\right) \ge 1 - \frac{s2^{J+1}C'}{n^{2\gamma}}, \quad s = |\mathcal{M}_*|. \tag{5}$$

Additionally, for fixed s and J, $P\left(\hat{\mathcal{M}} \supset \mathcal{M}_*\right) \to 1$ as $n \to \infty$.

Proof. See Appendix B. □

Another important aspect of any screening set is that its cardinality must be much lower than the dimension p. Proposition 3 shows that, on average, the number of elements on $\hat{\mathcal{M}}$ is much lower than p.

Proposition 3. Assume that all conditions of Proposition 2 hold and that $\hat{\mathcal{M}}$ is defined with $\nu_n = 2^{(J+1)/2}\sqrt{\kappa}((\log n)/n)^{1/4}$. Then, the expected cardinality of $\hat{\mathcal{M}}$ satisfies

$$E(|\hat{\mathcal{M}}|) \le s + \frac{2^{J+1}C'(p-s)}{n^{2\gamma}}, \quad \text{where } s = |\mathcal{M}_*|.$$

Proof. See Appendix C. □

4. Simulation

This section presents simulation results of Algorithm 1. Similarly as in Example 4 of Hall and Miller (2009), we simulate data from the model

$$Y_i = \sum_{j=1}^{5} \frac{6-j}{5}(f(X_{ij}) + f(X_{i,j+5})) + \epsilon_i, \tag{6}$$

where ϵ_i follows a standard normal distribution. The function $f: [0,1] \to \mathbb{R}$ is set as one of the test functions of Donoho and Johnstone (1994) (see Figure 1) or the identity function $f(x) = x$. The pairs $(X_{ij}, X_{i,j+5})$ are independently generated as

$$X_{ij} \sim U(0,1), \quad X_{i,j+5} = \frac{\rho X_{ij} + U_i\sqrt{1-\rho^2}}{\rho + \sqrt{1-\rho^2}},$$

$$j = 1, \dots, 5, \ i = 1, \dots, n, \tag{7}$$

where $\rho = 0.85$ and $U_1, \dots, U_n \overset{iid}{\sim} \text{Unif}(0,1)$. This generating scheme guarantees that some covariates are correlated, with $\text{corr}(X_{ij}, X_{i,j+5}) = \rho$. When true covariates X_j and X_{j+5} are highly correlated, many variable selection methods tend to choose only one of them. This simulation study also checks the ability of screening methods to select both.

We set the dimension as $p = 5000$ and consider sample sizes $n \in \{128, 256, 512\}$. The relevant covariates are $\mathbf{X}_1, \dots, \mathbf{X}_{10}$, that is, $\mathcal{M}_* = \{1, \dots, 10\}$. We also generate $\{X_{i,j}\}_{j=11}^{15}$ correlated with $\{X_{i,j}\}_{j=1}^{5}$ as outlined in (7). This way, we can analyze a case where nonsupport and support covariates are correlated. The remaining $\mathbf{X}_{16}, \dots, \mathbf{X}_{5000}$ have all entries sampled independently from $\text{Unif}(0,1)$. This section presents the numbers of true covariates detected by different methods among their 15 largest screening utility scores. In the supplementary material, we present the number of pairs of support covariates $\{(X_j, X_{j+5})\}_{j=1}^{5}$ that are selected together, and the number of selected nonsupport covariates $\{X_j\}_{j=11}^{15}$. All results are averages from 500 replicates.

Our Algorithm 1 is applied with the Haar wavelet basis and a single decomposition level, which are the default choices in our codes.[1] This wavelet method is compared with the generalized

---

[1]The supplementary material shows that similar results are obtained with different wavelet basis, and that Algorithm 1 outperforms a wavelet screening applied with the method of Kovac and Silverman (2000), which is implemented as the `irregwd` function in the R package `wavethresh` (Nason et al. 2022).
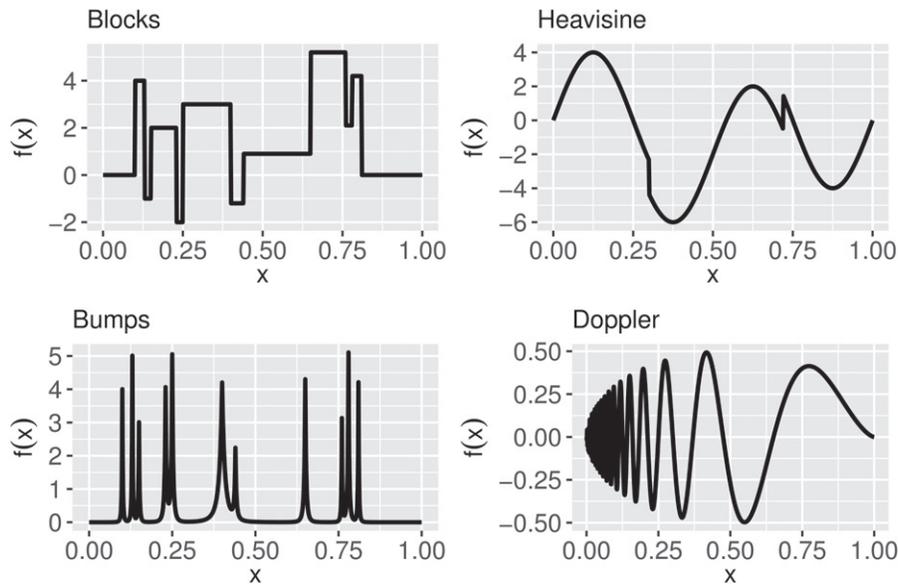
Figure 1. Test functions of Donoho and Johnstone (1994).

correlation screening of Hall and Miller (2009) implemented with cubic splines. We also apply the model-free methods SIRS (Zhu et al. 2011) and DC-SIS (Li, Zhong, and Zhu 2012), which are implemented in the R package `VariableScreening` (Li, Huang, and Dziak 2022).

Simulation results are presented in Table 1. We note that, albeit no method outperforms all others in all scenarios, the warped wavelets provided competitive results. When f is the Blocks or Heavisine functions, generalized correlation with splines gives the best results, and SIRS and DC-SIS have the best performances when f is the identity function. Apart from DC-SIS, all considered methods take on average only a few seconds to run. In general, wavelet methods have competitive results, outperforming the other approaches for the functions Bumps and Doppler, for example. This simulation highlights the good adaptation properties of wavelets for link functions f_j with many localized traits like peaks and sharp oscillations.

5. Application on Ozone Concentration

Ozone is a pollutant related to various respiratory problems, and its efficient monitoring greatly impacts public health (Zhang, Wei, and Fang 2019). However, various chemical and meteorological factors affect ozone concentration, making its monitoring a challenging task (Monks et al. 2015).

Our goal in this application is to identify the main factors that influence ozone concentration in São Paulo–Brazil (Santolaya et al. 2019) through wavelet feature screening. The relations between ozone level and related variables are known to be nonlinear for a long time (Lin, Trainer, and Liu 1988), which makes nonparametric screening methods appropriate tools for variable selection. We use the R package `qualR` (Gavidia-Calderón, Schuch, and de Fatima Andrade 2022) to collect data from the Pinheiros air quality station in São Paulo. The data consists of daily observations of 11 variables collected from January 2004 until December 2022. Due to many missing values, we analyze monthly averages and use only observations from April 2006

Table 1. Average number of true features (#feat) selected among the 15 most relevant according to each screening method, and average runtime in seconds (time).

	WWS		gen. corr./splines		SIRS		DC-SIS	
n	#feat	time	#feat	time	#feat	time	#feat	time
				Blocks				
128	1.1(0.93)	3.48	1.4(0.97)	1.57	0(0.23)	1.09	0.1(0.41)	2.75
256	3.3(0.95)	3.86	3.5(1.12)	1.83	0.1(0.29)	2.39	0.4(0.71)	11.86
512	5.4(0.8)	4.3	5.5(0.89)	2.36	0.2(0.53)	6.04	1.4(1.08)	74.76
				Heavisine				
128	3.3(0.88)	3.47	3.4(0.96)	1.57	0.2(0.55)	1.08	0.5(0.81)	2.71
256	5.2(0.86)	3.84	5.3(0.92)	1.8	0.7(0.92)	2.36	2.2(1.01)	11.77
512	6.5(0.77)	4.47	6.7(0.77)	2.46	2.1(1.13)	6.09	4.3(0.85)	76.74
				Bumps				
128	0.2(0.43)	3.47	0.1(0.23)	1.57	0.1(0.27)	1.08	0.1(0.28)	2.74
256	0.9(0.84)	3.87	0.2(0.48)	1.83	0.2(0.52)	2.38	0.3(0.59)	11.84
512	3.0(1.02)	4.33	0.7(0.88)	2.35	0.5(0.81)	5.78	0.8(1.01)	74.54
				Doppler				
128	0.2(0.45)	3.46	0.1(0.37)	1.56	0.1(0.41)	1.08	0.1(0.44)	2.72
256	0.8(0.84)	3.85	0.4(0.66)	1.81	0.2(0.57)	2.34	0.3(0.63)	11.62
512	2.3(1.05)	4.23	1.6(1.12)	2.3	0.5(0.87)	5.54	1(1.03)	71.22
				Identity				
128	2.2(1.11)	3.55	2.3(1.16)	1.6	3.1(1.15)	1.11	3.2(1.17)	2.77
256	4.1(1.05)	3.83	4.5(1.0)	1.81	5.1(0.98)	2.33	5.1(0.98)	11.67
512	5.4(0.89)	4.26	6(0.78)	2.3	6.3(0.75)	5.5	6.3(0.7)	71.2

NOTE: The standard deviation of #feat is presented in parentheses. The functions of true covariates are the four test functions of Donoho and Johnstone (1994) and the identity function. WWS denotes the warped wavelet feature screening.

until September 2010. Moreover, we compute the cosine and sine of circular variables related to wind direction. The time series are displayed in Figure 2. Notice that most time series present a seasonal pattern; thus, the observations are not independent.

The properties of feature screening methods discussed in this article assume independent observations. Thus, we computed differences in the variables shown in Figure 2 to reduce their dependence on past values. After plots of autocorrelations and partial autocorrelations, we employ the following:
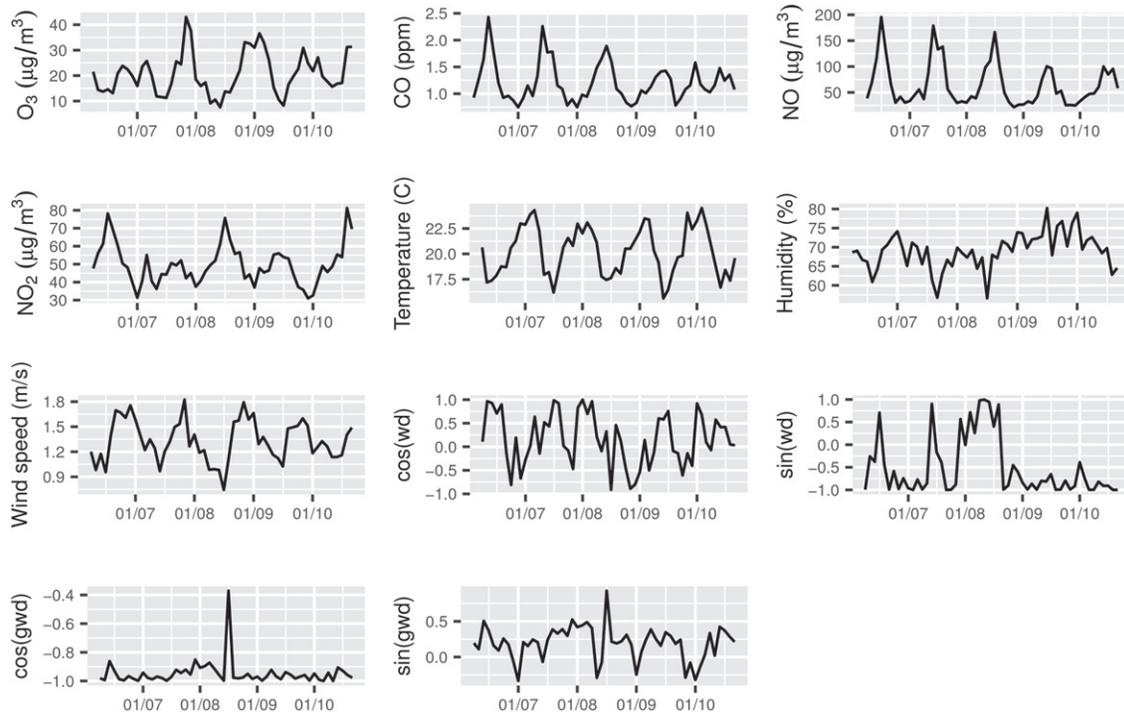
Figure 2. Time series of the meteorological variables recorded at Pinheiros air quality station in São Paulo. The abbreviations wd and gwd refer to wind direction and global wind direction, respectively.

- Differences at lags 1 and 11 for the variables ozone, carbon monoxide, nitrogen monoxide, and temperature,
- Differences at the lag 1 for the variables nitrogen dioxide, humidity, wind speed, cos(wd), and sin(wd), where wd stands for wind direction,
- No transformation for the variables cos(gwd), and sin(gwd), where gwd stands for global wind direction.

The temporal dependence of all the variables was reduced considerably after these transformations, as shown in the supplementary material.

The next step uses the differenced time series to identify the most important features to model ozone concentration. We apply the generalized correlation with cubic splines, SIS, DC–SIS, and SIRS. The warped wavelet feature screening is applied using Daubechies' extremal phase wavelets with filter number six.[2] Table 2 presents the five most relevant variables selected by each method. All methods identify the following features as most relevant: humidity, wind speed, and cos(wd). Hence, there is a consensus that humidity and wind direction and speed are relevant variables to model ozone concentration. This conclusion is in accordance with well-known results in the literature (Thompson et al. 2001, Sec 2.2.2). Wind direction and speed are associated with the transport of ozone in the atmosphere (Camalier, Cox, and Dolwick 2007), and weather factors such as humidity and temperature are frequently used to study peaks of ozone concentration (Oliveira, Drumond, and Rizzo 2022).

Figure 3 shows curves estimated using warped wavelets for some features. The vertical axis represents the differenced time series of ozone concentration minus its average. We subtract the average $\bar{Y}$ of this variable before estimating the curves (Hastie,

Table 2. The five most relevant variables according to different feature screening methods.

Method	1st	2nd	3rd	4th	5th
WWS	Humidity	Temperature	Wind speed	cos(wd)	sin(gwd)
gen. corr.	Humidity	Temperature	Wind speed	cos(wd)	NO$_2$
SIS	Humidity	cos(wd)	Wind speed	sin(wd)	NO$_2$
SIRS	Humidity	cos(wd)	sin(wd)	Wind speed	NO
DC-SIS	Humidity	sin(wd)	Wind speed	cos(wd)	NO

NOTE: WWS denotes the warped wavelet feature screening. Variables are ordered in decreasing order of importance.

Tibshirani, and Friedman 2009, p. 298). The covariates are the differenced time series of features normalized to the interval $(0, 1)$. We see in Figure 3 that Humidity, Wind Speed, and cos(wd) present a nonlinear relationship with the ozone concentration. Understanding the importance of these features and how they influence ozone concentration can aid scientists, for example, in developing simulation models (Sánchez-Ccoyllo et al. 2006; Franco et al. 2019).

6. Application on Epilepsy

Epilepsy is a brain disorder characterized by a predisposition and occurrence of multiple seizures (Fisher et al. 2014). A common form of this disorder is temporal lobe epilepsy (TLE), which can happen on one side of the brain. The progress of TLE is associated with abnormalities in the brain's gray and white matter (Yasuda et al. 2010). Many studies focus on characteristics of white matter to study some aspects of TLE. One particular problem is to understand brain degeneration in epileptic patients. Comparisons with healthy controls have shown that epileptic patients seem to have older brains (Hwang et al. 2020).

[2]See the `wavethresh` manual for details (Nason et al. 2022).
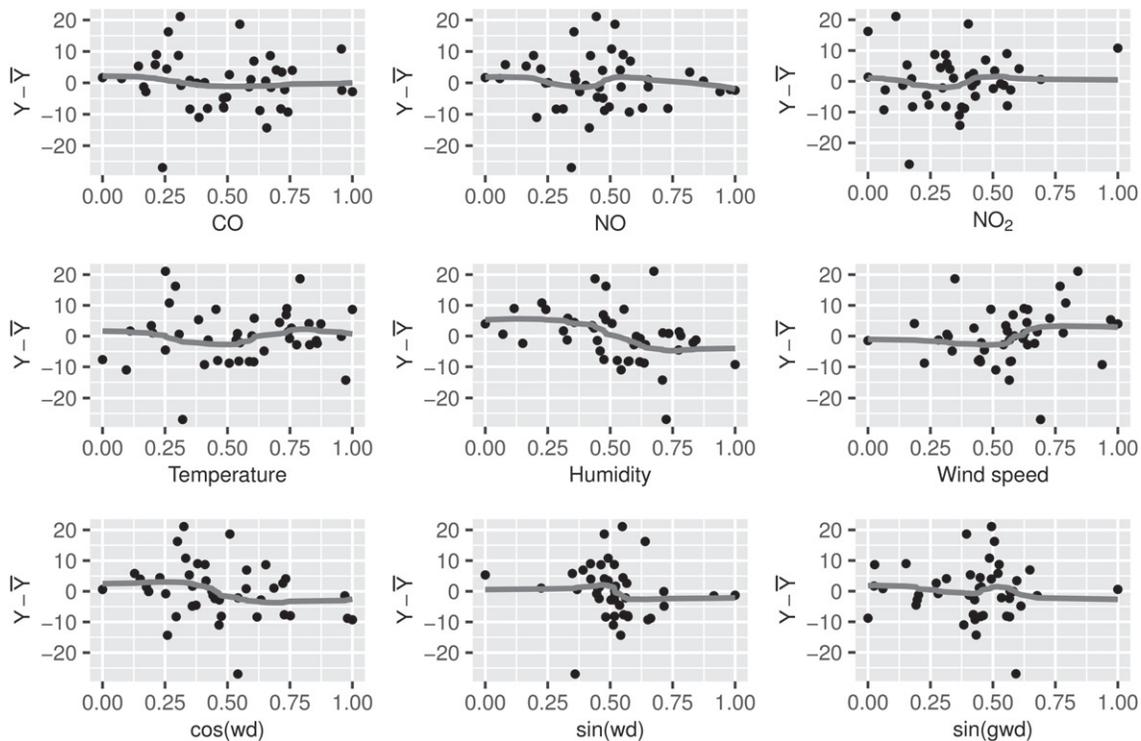
Figure 3. Plots of the response variable Y (differenced and centered ozone concentrations) and each covariate (differenced and normalized to the unit interval). The curves are warped wavelet regression estimates. The observations are considerably noisy, and only a few features display a noticeable effect on Y, like humidity and wind speed.
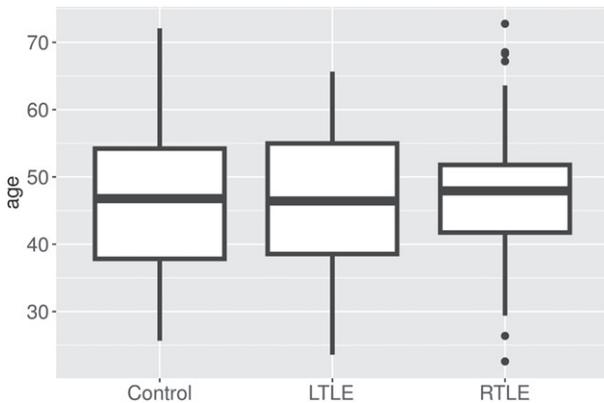

Figure 4. Boxplot of the observed ages for each group in the epilepsy study.

Further studies have been conducted to understand the relation between brain structural age and neurological measurements under the TLE disorder (Yasuda et al. 2023).

In this section, we apply feature screening on a dataset from a study about TLE. The data was collected by the Neuroimaging Laboratory at the University of Campinas-Brazil and consists of 88 features and ages obtained from 226 individuals (89 controls and 137 patients). The patients are divided by the side of their TLE: 76 have left TLE (LTLE) and 61 have right TLE (RTLE). The features are brain white matter measurements acquired via magnetic resonance imaging. We analyze pre-processed measurements in the interval [0, 1]. Further details about the dataset can be checked in Yasuda et al. (2023).

Our task is to identify which features can be used to model the patients' age. The research hypothesis is that epileptic patients have feature values similar to older healthy controls. Figure 4

Table 3. The five most relevant variables according to different feature screening methods for each group of participants in the epilepsy study.

Group	Method	1st	2nd	3rd	4th	5th
	WWS	BODY_FA	Genu_FA	Genu_MD	Genu_AD	Genu_RD
	gen. corr.	BODY_FA	BODY_MD	BODY_AD	BODY_RD	Genu_FA
Control	SIS	BODY_FA	BODY_RD	Genu_FA	Genu_MD	Genu_RD
	SIRS	BODY_FA	BODY_RD	Genu_FA	Genu_MD	Genu_RD
	DC-SIS	BODY_FA	BODY_RD	Genu_FA	Genu_MD	Genu_RD
	WWS	BODY_FA	Genu_FA	Genu_RD	SPLEN_AD	SPLEN_RD
	gen. corr.	BODY_FA	Genu_FA	Genu_MD	Genu_RD	SPLEN_MD
RTLE	SIS	BODY_FA	BODY_RD	Genu_FA	Genu_MD	Genu_RD
	SIRS	BODY_FA	BODY_RD	Genu_FA	Genu_MD	Genu_RD
	DC-SIS	BODY_FA	Genu_FA	Genu_MD	Genu_RD	SPLEN_RD
	WWS	BODY_FA	BODY_MD	BODY_AD	BODY_RD	Genu_FA
	gen. corr.	BODY_FA	BODY_MD	BODY_RD	Genu_FA	Genu_MD
LTLE	SIS	BODY_FA	BODY_MD	BODY_AD	BODY_RD	Genu_FA
	SIRS	BODY_FA	BODY_MD	BODY_AD	BODY_RD	Genu_FA
	DC-SIS	BODY_FA	BODY_MD	BODY_AD	BODY_RD	Genu_FA

NOTE: WWS denotes the warped wavelet feature screening. Variables are ordered in decreasing order of importance.

presents boxplots of the ages of each group of individuals. This figure shows that the distribution of ages is similar in the three groups. We applied feature screening methods separately for each group and checked which features were selected the most. Later, we analyze the estimated ages for two specific features.

We applied the generalized correlation with cubic splines, SIS, DC-SIS, and SIRS. Our Algorithm 1 is applied using Daubechies' extremal phase wavelets with filter number six.[3] Table 3 presents the five features selected the most by each screening method. The methods have consistent results and identify similar sets of features. The features related to BODY and Genu are the

---

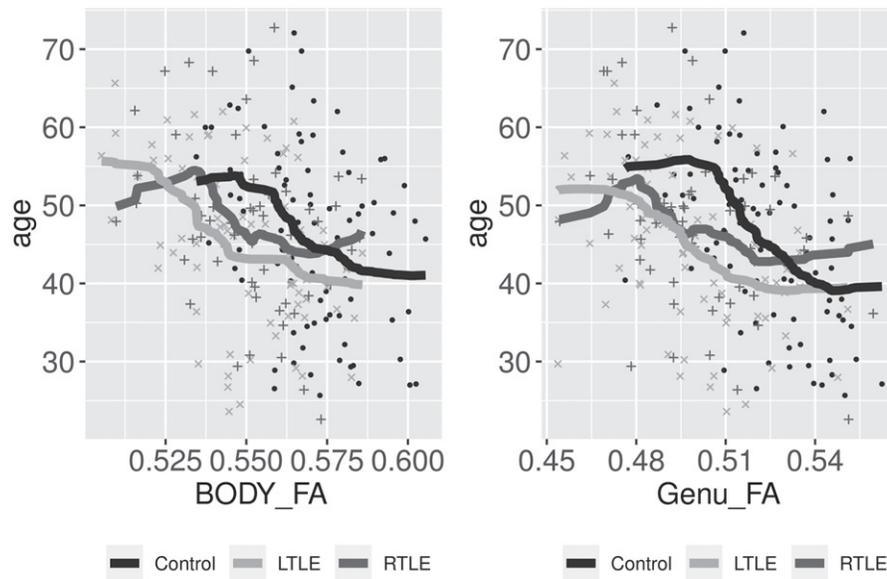[3] See the `wavethresh` manual for details (Nason et al. 2022).

Figure 5. Scatterplot of observed ages versus the covariates BODY_FA (right) and Genu_FA (left). The symbols ·, ×, and + correspond to control, LTLE, and RTLE groups, respectively. The solid curves were estimated via warped wavelet regression for each group of patients.

most selected. These brain structures were also highlighted in previous studies (Yasuda et al. 2023, Tables 2 and 5).

Table 3 shows that the two features selected the most for the three groups are BODY_FA and Genu_FA. Figure 5 presents scatterplots of these features versus the observed ages. The warped wavelet regression was applied to fit a smooth curve for each group of patients and control. Both plots in Figure 5 show that the curves of control individuals tend to be above the curves of RTLE and LTLE patients. This characteristic indicates that measurements of TLE patients are observed in controls with older ages, illustrating the degeneracy caused by the TLE disorder. Yasuda et al. (2023) reached similar conclusions using linear regression. Hence, the feature screening and wavelet analysis corroborate their results.

7. Conclusion

Feature screening is a common practice in various scientific applications. Scientists are frequently looking for relevant relationships between variables of interest and candidate features, and the correlation is probably the most used statistic to screen such relevant relations. However, high dimensionality and nonlinearity are common characteristics in practice and pose challenges to classical approaches in data analysis. High-dimensional data increases the chance of observing spurious correlations, and many important features may have a nonlinear relation with the variable of interest. In this context, many screening metrics have been proposed in the literature to cope with large dimensions and allow for flexible relations between variables. We propose a wavelet approach for performing feature screening. Wavelet methods are well-known tools in the literature of sparse models and can estimate various types of curves. The proposed method is based on the warped wavelet regression, and we prove that it can achieve the sure screening property under suitable conditions. Simulation studies also show that the proposed method is competitive with other feature screening approaches in the literature. Additionally, the

wavelet approach tends to outperform other methods when the link function between a feature and a response variable have sharp oscillations and peaks. Such an outcome agrees with the well-known ability of wavelets to adapt to signals with local nonlinearities. Our work also provides two applications in real data from relevant scientific problems: the relation of ozone concentration and atmospheric variables and the relation between brain white matter measurements and the age of patients with temporal lobe epilepsy. In both studies, the wavelet feature screening method provides results that are in line with other screening methods. Another application example in the supplementary material shows that wavelet feature screening also performs well in a case where $p \gg n$. Overall, all the screening methods can select features that have relations reported in the literature of their respective fields. Therefore, we believe the warped wavelet screening method is a powerful tool for scientists dealing with many linear and nonlinear features.

Appendix A. Auxiliary Lemma

Lemma 1. Assume that C1–C3 hold. Let κ, C', and γ be defined in Proposition 1, with $\gamma > 1/2$. Therefore, it follows that

$$P\left(\max_{j \in \mathcal{M}_*} \left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| > \nu_n \right) \leq \frac{s2^{J+1}C'}{n^{2\gamma}} \tag{8}$$

and

$$\sum_{j \notin \mathcal{M}_*} P\left(\|\hat{\boldsymbol{\beta}}_j\| > \nu_n \right) \leq \frac{(p-s)2^{J+1}C'}{n^{2\gamma}}, \tag{9}$$

where $s = |\mathcal{M}_*|$ and $\nu_n = 2^{(J+1)/2}\sqrt{\kappa}((\log n)/n)^{1/4}$.

Proof. Let us prove (8) first. By the inequality $\left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| \leq \|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|$ and the union bound, it follows that

$$P\left(\max_{j \in \mathcal{M}_*} \left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| > \nu_n \right) \leq \sum_{j \in \mathcal{M}_*} P\left(\|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\| > \nu_n \right).$$

We shall use the following event inclusion

$$\left\{ \|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\| > \nu_n \right\} = \left\{ \sum_{l=0}^{J} \sum_{k=0}^{2^l-1} (\hat{\beta}_{jlk} - \beta_{jlk})^2 > \nu_n^2 \right\}$$

$$\subset \bigcup_{l=0}^{J} \bigcup_{k=0}^{2^l-1} \left\{ (\hat{\beta}_{jlk} - \beta_{jlk})^2 > \frac{\nu_n^2}{2^{J+1}} \right\}. \quad (10)$$

Notice that (10) holds for all $j = 1, \ldots, p$. Then, by the union bound and since $\nu_n^2 = 2^{J+1}\kappa\sqrt{(\log n)/n}$,

$$P\left(\max_{j \in \mathcal{M}_*} \left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| > \nu_n \right)$$

$$\leq \sum_{j \in \mathcal{M}_*} \sum_{l=0}^{J} \sum_{k=0}^{2^l-1} P\left(|\hat{\beta}_{jlk} - \beta_{jlk}|^2 > \nu_n^2/2^{J+1} \right)$$

$$= \sum_{j \in \mathcal{M}_*} \sum_{l=0}^{J} \sum_{k=0}^{2^l-1} P\left(|\hat{\beta}_{jlk} - \beta_{jlk}| > \kappa\sqrt{\frac{\log n}{n}} \right).$$

Recall that $\mathcal{M}_*$ has cardinality s and notice that there are 2^{J+1} pairs (l,k) in the sum above. Hence, (8) follows by applying Proposition 1 with $\alpha = 2$.

Inequality (9) can be proved similarly. Indeed, since $\boldsymbol{\beta}_j = \mathbf{0}$ for $j \notin \mathcal{M}_*$, then

$$\sum_{j \notin \mathcal{M}_*} P\left(\|\hat{\boldsymbol{\beta}}_j\| > \nu_n \right) \leq \sum_{j \notin \mathcal{M}_*} \sum_{l=0}^{J} \sum_{k=0}^{2^l-1} P\left(|\hat{\beta}_{jlk}| > \kappa\sqrt{\frac{\log n}{n}} \right)$$

$$\leq \frac{(p-s)2^{J+1}C'}{n^{2\gamma}},$$

where the first inequality is due to the union bound and (10), and the second inequality is due to Proposition 1. □

Appendix B. Proof of Proposition 2

Proof. Consider the event $A_n = \{\max_{j \in \mathcal{M}_*} \left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| \leq \nu_n\}$. Let us show that if $\min_{j \in \mathcal{M}_*} \|\boldsymbol{\beta}_j\| > 2\nu_n$, then $A_n \subset \{\hat{\mathcal{M}} \supset \mathcal{M}_*\}$. Consider that A_n holds. It implies that $\|\hat{\boldsymbol{\beta}}_j\| \geq \|\boldsymbol{\beta}_j\| - \nu_n$ for all $j \in \mathcal{M}_*$. Since $\|\boldsymbol{\beta}_j\| - \nu_n \geq \nu_n$ by assumption, then $\|\hat{\boldsymbol{\beta}}_j\| \geq \nu_n$, that is, all $j \in \mathcal{M}_*$ also belong to $\hat{\mathcal{M}}$. Hence, $P\left(\hat{\mathcal{M}} \supset \mathcal{M}_* \right) \geq 1 - P\left(A_n^c \right)$. By (8)

$$P\left(A_n^c \right) = P\left(\max_{j \in \mathcal{M}_*} \left| \|\hat{\boldsymbol{\beta}}_j\| - \|\boldsymbol{\beta}_j\| \right| > \nu_n \right) \leq s2^{J+1}C'n^{-2\gamma},$$

where s is the cardinality of $\mathcal{M}_*$. This concludes the proof of (5).

Recall from the comment after (3) that $2^J \leq C\sqrt{n/\log n}$. Thus, $P\left(A_n^c \right) \leq 2sCn^{1/2-2\gamma}/\sqrt{\log n}$, where C is an absolute constant. Since $\gamma > 1/4$ by assumption, then $P\left(A_n^c \right) \to 0$ as $n \to \infty$ with s and J fixed. This concludes the proof of Proposition 2. □

Appendix C. Proof of Proposition 3

Proof. Recall that $s = |\mathcal{M}_*|$. We can write the cardinality $|\hat{\mathcal{M}}|$ as a sum of indicator functions, which leads to the following inequality:

$$|\hat{\mathcal{M}}| = \sum_{j=1}^{p} I\{\|\hat{\boldsymbol{\beta}}_j\| > \nu_n\} \leq s + \sum_{j \notin \mathcal{M}_*} I\{\|\hat{\boldsymbol{\beta}}_j\| > \nu_n\}.$$

Notice that $\boldsymbol{\beta}_j \equiv 0$ for $j \notin \mathcal{M}_*$. Then, it follows from (9) that

$$E\left\{ \sum_{j \notin \mathcal{M}_*} I\{\|\hat{\boldsymbol{\beta}}_j\| > \nu_n\} \right\} = \sum_{j \notin \mathcal{M}_*} P\{\|\hat{\boldsymbol{\beta}}_j\| > \nu_n\} \leq \frac{(p-s)2^{J+1}C'}{n^{2\gamma}}.$$

□

Supplementary Materials

Full simulation results: The complete results of the simulation presented in Section 4.

Application in high-dimensional data: A feature screening application on the cardiomiopathy data analyzed by Hall and Miller (2009) and Li, Zhong, and Zhu (2012).

Summary of applications: R markdown files outlining the data analysis presented in Sections 5 and 6.

Codes: All R codes used for simulation studies and to analyze the datasets about ozone concentration and cardiomyopathy.

Acknowledgments

Disclosure Statement

The authors report that there are no competing interests to declare.

Funding

ORCID

Rodney Fonseca http://orcid.org/0000-0003-3948-3145
Pedro Morettin http://orcid.org/0000-0002-8611-867X
Aluísio Pinheiro http://orcid.org/0000-0002-6132-7731

References

Barut, E., Fan, J., and Verhasselt, A. (2016), "Conditional Sure Independence Screening," *Journal of the American Statistical Association*, 111, 1266–1277. [1310]

Camalier, L., Cox, W., and Dolwick, P. (2007), "The Effects of Meteorology on Ozone in Urban Areas and their Use in Assessing Ozone Trends," *Atmospheric Environment*, 41, 7127–7137. [1315]

Donoho, D. L., and Johnstone, J. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425–455. [1313,1314]

Fan, J., Feng, Y., and Song, R. (2011), "Nonparametric Independence Screening in Sparse Ultra-High-Dimensional Additive Models," *Journal of the American Statistical Association*, 106, 544–557. [1310,1311,1312]

[4] *https://www.brainn.org.br/*

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [1313]

Fan, J., Li, R., Zhang, C.-H., and Zou, H. (2020), *Statistical Foundations of Data Science*, Boca Raton: CRC Press. [1310]

Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultrahigh Dimensional Feature Space," *Journal of the Royal Statistical Society*, Series B, 70, 849–911. [1310,1311]

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Helen Cross, J., Elger, C. E., et al. (2014), "ILAE Official Report: A Practical Clinical Definition of Epilepsy," *Epilepsia*, 55, 475–482. [1315]

Franco, D. M. P., Andrade, M. d. F., Ynoue, R. Y., and Ching, J. (2019), "Effect of Local Climate Zone (LCZ) Classification on Ozone Chemical Transport Model Simulations in Sao Paulo, Brazil," *Urban Climate*, 27, 293–313. [1315]

Gavidia-Calderón, M., Schuch, D., and de Fatima Andrade, M. (2022), *qualR: An R Package to Download São Paulo and Rio de Janeiro Air Pollution Data*, R package version 0.9.7. [1314]

Gómez, L. M., Porto, R. F., and Morettin, P. A. (2021), "Nonparametric Regression with Warped Wavelets and Strong Mixing Processes," *Annals of the Institute of Statistical Mathematics*, 73, 1203–1228. [1312]

Hall, P., and Miller, H. (2009), "Using Generalized Correlation to Effect Variable Selection in Very High Dimensional Problems," *Journal of Computational and Graphical Statistics*, 18, 533–550. [1310,1311,1313,1314,1318]

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998), *Wavelets, Approximation, and Statistical Applications*, New York: Springer. [1312]

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2), New York: Springer. [1310,1315]

Hwang, G., Hermann, B., Nair, V. A., Conant, L. L., Dabbs, K., Mathis, J., et al. (2020), "Brain Aging in Temporal Lobe Epilepsy: Chronological, Structural, and Functional," *NeuroImage: Clinical*, 25, 102183. [1315]

Kerkyacharian, G., and Picard, D. (2004), "Regression in Random Design and Warped Wavelets," *Bernoulli*, 10, 1053–1105. [1311,1312,1313]

Kovac, A., and Silverman, B. W. (2000), "Extending the Scope of Wavelet Regression Methods by Coefficient-Dependent Thresholding," *Journal of the American Statistical Association*, 95, 172–183. [1313]

Li, R., Huang, L., and Dziak, J. (2022), *VariableScreening: High-Dimensional Screening for Semiparametric Longitudinal Regression*, R package version 0.2.1. [1314]

Li, R., Zhong, W., and Zhu, L. (2012), "Feature Screening via Distance Correlation Learning," *Journal of the American Statistical Association*, 107, 1129–1139. [1310,1314,1318]

Lin, X., Trainer, M., and Liu, S. (1988), "On the Nonlinearity of the Tropospheric Ozone Production," *Journal of Geophysical Research*, 93, 15879–15888. [1314]

Mai, Q., and Zou, H. (2013), "The Kolmogorov Filter for Variable Screening in High-Dimensional Binary Classification," *Biometrika*, 100, 229–234. [1310]

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., et al. (2015), "Tropospheric Ozone and its Precursors from the Urban to the Global Scale from Air Quality to Short-Lived Climate Forcer," *Atmospheric Chemistry and Physics*, 15, 8889–8973. [1314]

Morettin, P. A., Pinheiro, A., and Vidakovic, B. (2017), *Wavelets in Functional Data Analysis*, Cham: Springer. [1311]

Morettin, P. A., and Porto, R. F. (2022), "Estimation of Nonparametric Regression Models by Wavelets," *São Paulo Journal of Mathematical Sciences*, 16, 539–565. [1311]

Nason, G., Barber, S., Downie, T., Frylewicz, P., Kovac, A., Ogden, T., and Silverman, B. (2022), *Package 'wavethresh'*, R package version 4.7.2. [1313,1315,1316]

Ogden, R. T. (1997), *Essential Wavelets for Statistical Applications and Data Analysis*, Boston: Birkhäuser. [1312]

Oliveira, M., Drumond, A., and Rizzo, L. (2022), "Air Pollution Persistent Exceedance Events in the Brazilian Metropolis of Sao Paulo and Associated Surface Weather Patterns," *International Journal of Environmental Science and Technology*, 19, 9495–9506. [1315]

Porto, R., Morettin, P., Percival, D., and Aubin, E. (2016), "Wavelet Shrinkage for Regression Models with Random Design and Correlated Errors," *Brazilian Journal of Probability and Statistics*, 30, 614–652. [1312]

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017), "Methods for Scalar–on–Function Regression," *International Statistical Review*, 85, 228–249. [1311]

Sánchez-Ccoyllo, O. R., Ynoue, R. Y., Martins, L. D., and Andrade, M. d. F. (2006), "Impacts of Ozone Precursor Limitation and Meteorological Variables on Ozone Concentration in Sao Paulo, Brazil," *Atmospheric Environment*, 40, 552–562. [1315]

Santolaya, C., Oliveira, M. C. Q. D., Rizzo, L. V., and Miraglia, S. G. E. K. (2019), "Contribution of Chemical and Meteorological Factors to Tropospheric Ozone Formation in São Paulo, Brazil," *Brazilian Journal of Environmental Sciences*, 54, 90–104. [1314]

Segal, M. R., Dahlquist, K. D., and Conklin, B. R. (2003), "Regression Approaches for Microarray Data Analysis," *Journal of Computational Biology*, 10, 961–980. [1312]

Thompson, M. L., Reynolds, J., Cox, L. H., Guttorp, P., and Sampson, P. D. (2001), "A Review of Statistical Methods for the Meteorological Adjustment of Tropospheric Ozone," *Atmospheric Environment*, 35, 617–630. [1315]

Vidakovic, B. (1999), *Statistical Modeling by Wavelets*, New York: Wiley. [1311,1312,1313]

Xue, J., and Liang, F. (2017), "A Robust Model-Free Feature Screening Method for Ultrahigh-Dimensional Data," *Journal of Computational and Graphical Statistics*, 26, 803–813. [1310]

Yasuda, C. L., Valise, C., Saúde, A. V., Pereira, A. R., Pereira, F. R., Costa, A. L. F., et al. (2010), "Dynamic Changes in White and Gray Matter Volume are Associated with Outcome of Surgical Treatment in Temporal Lobe Epilepsy," *Neuroimage*, 49, 71–79. [1315]

Yasuda, C. L., Ramalho Pimentel-Silva, L., Beltramini, G. C., Liu, M., de Campos, B. M., Coan, A. C., Beaulieu, C., Cendes, F., and Gross, D. W. (2023), "Brain Volumes and White Matter Diffusion Across the Adult Lifespan in Temporal Lobe Epilepsy," *Annals of Clinical and Translational Neurology*, 10, 1106–1118. [1316,1317]

Yousuf, K. (2018), "Variable Screening for High Dimensional Time Series," *Electronical Journal of Statistics*, 12, 667–702. [1310]

Zhang, J., Wei, Y., and Fang, Z. (2019), "Ozone Pollution: A Major Health Hazard Worldwide," *Frontiers in Immunology*, 10, 2518. [1314]

Zhao, Y., Chen, H., and Ogden, R. T. (2015), "Wavelet-based Weighted LASSO and Screening Approaches in Functional Linear Regression," *Journal of Computational and Graphical Statistics*, 24, 655–675. [1311]

Zhu, L.-P., Li, L., Li, R., and Zhu, L.-X. (2011), "Model-Free Feature Screening for Ultrahigh-Dimensional Data," *Journal of the American Statistical Association*, 106, 1464–1475. [1310,1314]