Drift Removal for Time Series Data Using Quantile Trend Filtering

Halley Brantley* Joseph Guinness† and Eric C. Chi‡

Abstract

The text of your abstract. 200 or fewer words.

Keywords: 3 to 6 keywords, that do not appear in the title

*Department of Statistics, North Carolina State University, Raleigh, NC 27695 (E-mail: hlbrantl@ncsu.edu)

†Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853 (E-mail: guinness@cornell.edu)

‡Department of Statistics, North Carolina State University, Raleigh, NC 27695 (E-mail: eric_chi@ncsu.edu).

1 Introduction

In many applications spanning the fields of chemistry (Ning et al., 2014), macroeconomics (Yamada, 2017), environmental science (Brantley et al., 2014), and medical sciences (Pettersson et al., 2013; Marandi and Sabzpoushan, 2015), scalar time series are observed and assumed to consist of a slowly varying trend other more rapidly varying components. Given observations $y(t)$, with $t = 1, ..., N$, Kim et al. (2009) proposed using ℓ_1 *trend filtering* to estimate trends that are smooth in the sense of being piecewise linear or piecewise polynomial. Tibshirani (2014) then discovered that empirically the trend filtering estimates adapt to the local level of smoothness much better than the more common smoothing splines. In the trend filtering problem (Kim et al., 2009; Tibshirani, 2014), the trend, $\theta \in \mathbb{R}^n$, is estimated by solving the following convex problem.

$$\arg\min_{\theta} \frac{1}{2}\|y - \theta\|_2^2 + \lambda\|\mathbf{D}^{(k+1)}\theta\|_1, \tag{1}$$

where $\lambda \geq 0$ is a regularization parameter and the matrix $\mathbf{D}^{(k+1)} \in \mathbb{R}^{(n-k-1)\times n}$ is the discrete difference operator of order $k+1$. To understand the purpose of penalizing $\mathbf{D}^{(k+1)}$ consider the difference operator when $k = 0$.

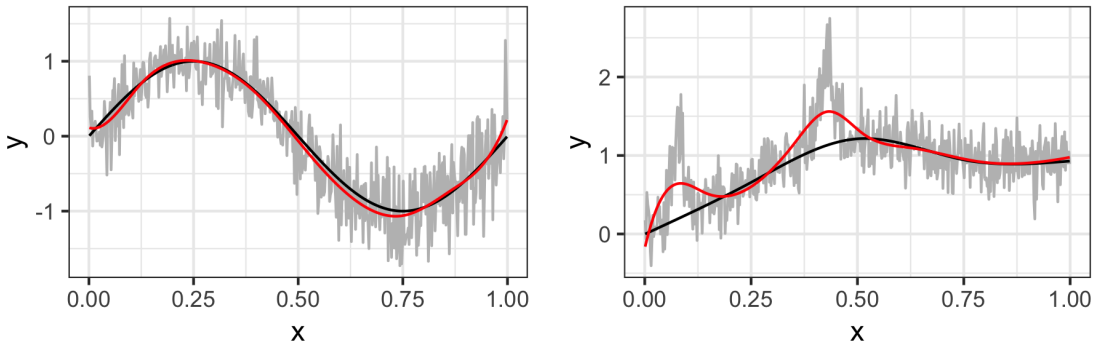
$$\mathbf{D}^{(1)} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \tag{2}$$

Thus, $\|\mathbf{D}^{(1)}\theta\|_1 = \sum_{i=1}^{n-1}|\theta_i - \theta_{i+1}|$ which is just total variation denoising in one dimension. The penalty incentivizes solutions which are piece-wise constant. For $k \geq 1$, the difference operator $\mathbf{D}^{(k+1)} \in \mathbb{R}^{(n-k-1)\times n}$ is defined recursively as follows

$$\mathbf{D}^{(k+1)} \;=\; \mathbf{D}^{(1)}\mathbf{D}^{(k)}. \tag{3}$$

By penalizing the $k+1$ fold composition of the discrete difference operator, we obtain solutions which are piecewise polynomials of order k. Trend filtering provides excellent estimates of trends when the only components of the time series are the trend and random noise (Fig. 1). In some cases, time series may also include a rapidly varying signal

Figure 1: Examples of trend filtering solutions. The true trend is shown in black, while the .estimated trend is shown in red.



component in addition to the trend and noise, in these cases the trend filtering estimate over-estimate the trend in the places where signal is present (Fig. 1).

One application in which the observed time series consists of a slowly varying trend, non-negative signal, and rapidly varying noise is the output of low cost air quality sensors. The use of low-cost, portable, air quality sensors has increased dramatically in the last decade. These sensors can provide an un-calibrated measure of a variety of pollutants in near real time, but deriving meaningful information from sensor data remains a challenge (Snyder et al., 2013). The "SPod" is a low-cost sensor currently being investigated by researchers at the U.S. Environmental Protection Agency to detect volatile organic compound (VOC) emissions from industrial facilities (Thoma et al., 2016). To reduce cost and power consumption of the SPod, the relative humidity of the air presented to the photoionization detectors (PIDs) is not controlled and as a result the output signal exhibits a slowly varying baseline drift on the order of minutes to hours (Fig. 2).

In this application, as well as those described in Ning et al. (2014), Marandi and Sabzpoushan (2015), and Pettersson et al. (2013), it is not the mean trend that is desired but rather the baseline trend, which we can think of as the trend in a low quantile of the data. Koenker and Bassett (1978) were the first to propose substituting the squared loss used in regression with the check loss function (Eq. 4) to estimate a conditional quantile instead of the conditional mean.

$$\rho_\tau(z) = \sum_{i=1}^{n} z_i(1 - \mathbf{I}(z_i < 0)). \tag{4}$$

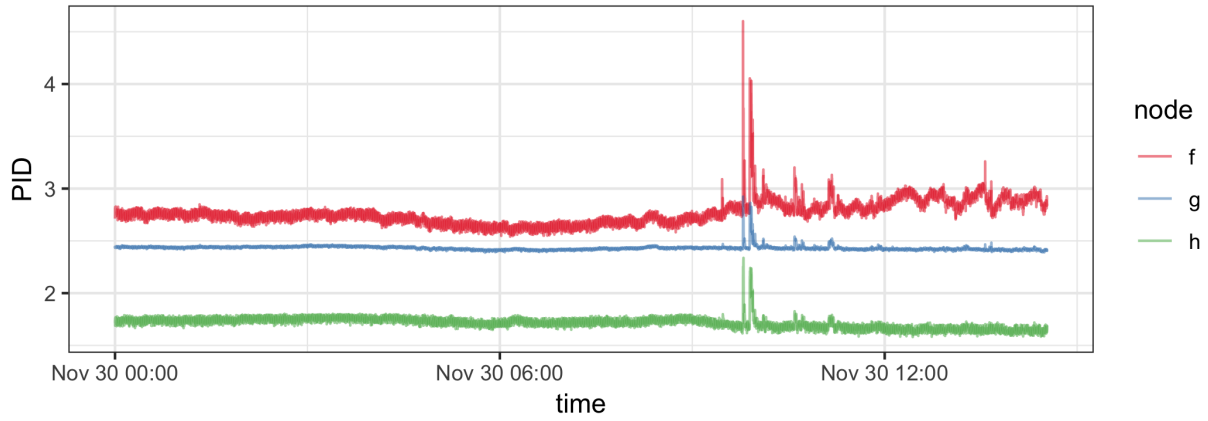
A variety of approaches have been proposed for estimating quantile trends. The regression vector θ is estimated by finding the minimizer of $\rho_\tau(y - X\theta)$ where is the check-loss function and $\mathbf{I}$ is the indicator function. Nychka et al. (1995) estimated quantile splines by combining the smoothing spline penalty commonly used in non-parametric regression with the check-loss function used in qunatile regression:

$$\rho_\tau(Y - f) + \lambda \int (f''(t))^2 dt,$$

where $Y(t)$ is a response variable observed at time t, $f(t)$ is a smooth function of time and λ is a tuning parameter that controls the degrees of smoothing.

We propose to use the trend filtering penalty with the check loss function to produce a non-parametric quantile regression estimate for removing trends in time series. The formulation was proposed by Kim et al. (2009) as a possible extension of ℓ_1-trend filtering but not studied. Moreover we extend the basic framework to ensure non-crossing while modeling multiple quantiles. We also implement a parallel ADMM algorithm for series that are too large to be computed simultaneously and proposed a modified criteria for choosing the smoothing parameter. We demonstrate through simulation studies that our

Figure 2: Example of 3 co-located SPod PID sensor readings.



proposed model provides better or comparable estimates of non-parametric quantile trends than existing methods and is a more effective method of drift removal for low-cost air quality sensors.

2 Methods

2.1 Quantile Trend Filtering

We combine the ideas of quantile regression and trend filtering, namely consider the case where the design $\mathbf{X}$ is the identity matrix. For a single quantile level τ the estimation of the quantile trend filtering model can be posed as the following optimization problem.

$$\min_{\theta} \rho_\tau(y-\theta) + \lambda \|\mathbf{D}^{(k+1)}\theta\|_1, \tag{5}$$

where λ is a non-negative regularization parameter. We address the problem of choosing λ in Section 2.3. As with the classic quantile regression, the quantile trend filtering problem is a linear program which can be solved by a number of free or commercial solvers. In many cases, including ours, we are interested in estimating multiple quantiles simultaneously. We also want to ensure that our quantile estimates are valid by enforcing the constraint that if $\tau_2 > \tau_1$ then $Q(\tau_2) \geq Q(\tau_1)$. Given quantiles $\{\tau_1, ..., \tau_J\}$ such that $\tau_1 < \tau_2 < ... < \tau_J$, the optimization problem becomes

$$\min_{\theta_1,...,\theta_J} \sum_{j=1}^{J} \left[\rho_{\tau_j}(y-\theta_j) + \lambda_j \|\mathbf{D}^{(k+1)}\theta_j\|_1\right] \tag{6}$$

$$\text{subject to: } \theta_{1i} \leq \theta_{2i} \leq ... \leq \theta_{Ji} \text{ for all } i, \tag{7}$$

where $\theta_j \in \mathcal{R}^n$. The additional constraints are linear in the parameters so the non-crossing quantile trends can still be estimated by a number of available solvers. In the rest of this paper we rely on the commercial solver Gurobi (Gurobi Optimization, 2018).

The number of parameters to be estimated is equal to the number of observations multiplied by the number of quantiles of interest. As the size of the data and the number of quantiles grows, all solvers will eventually break.

2.2 ADMM for Big Data

To our knowledge, no one has addressed the problem of finding smooth quantile trends of series that are too large to be processed simultaneously. We propose an alternating direction method of multipliers (ADMM) algorithm for solving large problems in a piecewise fashion. The ADMM algorithm is fully described by Boyd et al. (2011); Gabay and Mercier (1975); Glowinski and Marroco (1975). We apply the consensus ADMM algorithm to the the quantile regression trend filtering problem given in Eq. 5, by dividing our observed series $y(t)$ with $t = \{1, ..., N\}$ into overlapping windows

$$\begin{cases} y_1(t) = y(t) & \text{if } \ 1 \leq t \leq u_1 \\ y_2(t) = y(t) & \text{if } \ l_2 \leq t \leq u_2 \\ y_3(t) = y(t) & \text{if } \ l_3 \leq t \leq u_3 \\ \dots & \\ y_M(t) = y(t) & \text{if } \ l_M \leq t \leq N \end{cases}$$

with boundaries $1 < l_2 < u_1 < l_3 < u_2 < l_4 < u_3 < ... < N$. Given quantiles $\tau_1 < ... < \tau_J$ to be estimated we define $\theta_{j,m}(t)$ as the value of the $\tau_j{}^{\text{th}}$ quantile trend in window M at time point t. In order to write out the constraint the that overlapping sections must be equal we define a consensus variable

$$\bar{\theta}_{j,m} = g(\theta_{j,m-1}, \theta_{j,m}, \theta_{j,m+1}) \tag{8}$$

$$= \begin{cases} \frac{\theta_{j,m-1}(t)+\theta_{j,m}(t)}{2} & \text{if } \ l_m \leq t \leq u_{m-1} \\ \theta_{j,m}(t) & \text{if } \ u_{m-1} \leq t \leq l_{m+1} \\ \frac{\theta_{j,m}(t)+\theta_{j,m+1}(t)}{2} & \text{if } \ l_{m+1} \leq t \leq u_m \end{cases}, \tag{9}$$

defining $\theta_{j,M+1} = \theta_{j,M}$ and $\theta_{j,0} = \theta_{j,1}$. Our windowed quantile trend optimization problem

can then be written as

$$\sum_{m=1}^{M} \min_{\theta_{1,m},\ldots,\theta_{J,m}} \sum_{j=1}^{J} \left[\rho_{\tau_j}(y_m - \theta_{j,m}) + \lambda_j \|\mathbf{D}^{(k+1)}\theta_{j,m}\|_1\right] \tag{10}$$

$$\text{subject to: } \theta_{1,m}(t) \leq \theta_{2,m}(t) \leq \ldots \leq \theta_{J,m}(t) \text{ for all } m, t \tag{11}$$

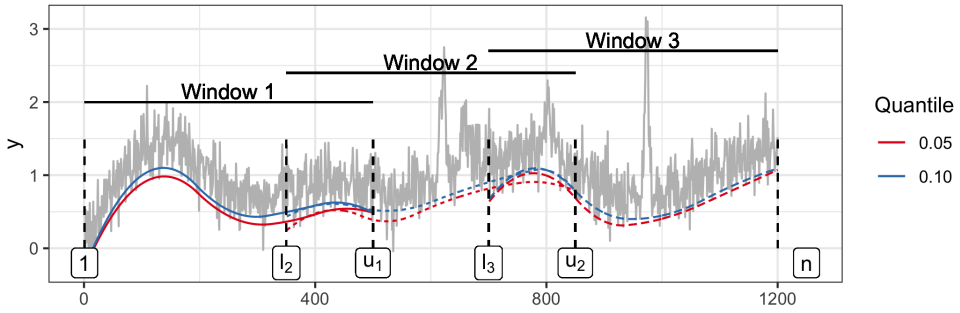
$$\text{subject to: } \theta_{j,m}(t) = \bar{\theta}_{j,m}(t) \text{ for all } j, m, t \tag{12}$$

Given values for the Lagrange multiplier $\omega_{j,m}$ and the consensus variable $\bar{\theta}_{j,m}$ we can write the augmented Lagrangian for finding the trends in window m:

$$\mathcal{L}(\theta_{j,m}, \bar{\theta}_{j,m}, \omega_{j,m}) = \sum_{j=1}^{J} \rho_{\tau_j}(y_m - \theta_{j,m}) + \lambda \|\mathbf{D}^{(k+1)}\theta_{j,m}\|_1 + \omega_{j,m}^T(\theta_{j,m} - \bar{\theta}_{j,m}) + \frac{\gamma}{2}||\theta_{j,m} - \bar{\theta}_{j,m}||_2^2 \tag{13}$$

We then estimate the trend separately in each window while constraining the overlapping pieces of the trends to be equal according the algorithm 1.

Figure 3: Window boundaries and trends fit separately in each window.



We used the stopping criteria described by Boyd et al. (2011). The criteria are based on the primal and dual residuals which represent the residuals for the primal and dual feasibility, respectively. The primal and dual residuals are defined as

$$r_p^{(q)} = \sqrt{\sum_{m=1}^{M}\sum_{j=1}^{J} \|\theta_{j,m}^{(q)} - \bar{\theta}_{j,m}^{(q)}\|_2^2} \tag{14}$$

$$r_d^{(q)} = \gamma\sqrt{\sum_{m=1}^{M}\sum_{j=1}^{J} \|\bar{\theta}_{j,m}^{(q)} - \bar{\theta}_{j,m}^{(q-1)}\|_2^2}. \tag{15}$$

Algorithm 1 ADMM algorithm for quantile trend filtering with windows

Define $D = D^{(k+1)}$.

initialize:

$\theta_{j,m}^{(0)} = \arg\min \sum_{j=1}^{J} \rho_{\tau_j}(y_m - \theta_{j,m}) + \lambda \|D\theta_{j,m}\|_1$ subject to $\theta_{1,m}(t) < ... < \theta_{J,m}(t)$ for all t.

$\omega_{j,m}^{(0)} = \mathbf{0}$

repeat

$\bar{\theta}_{j,m}^{(q)} = g(\theta_{j,m-1}^{(q-1)}, \theta_{j,m}^{(q-1)}, \theta_{j,m+1}^{(q-1)})$

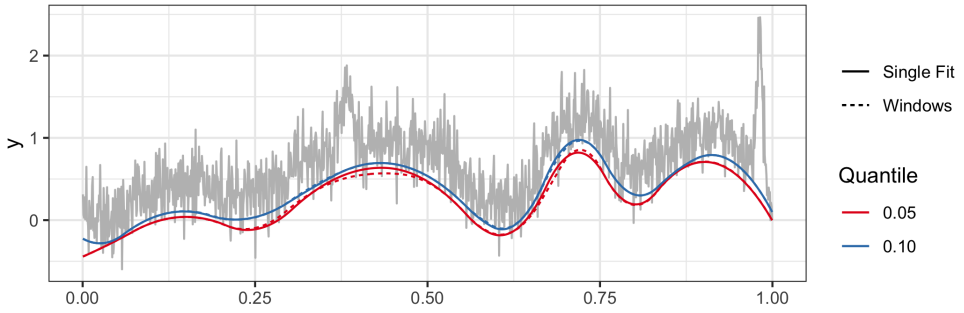
$\omega_{j,m}^{(q)} = \omega_{j,m}^{(q-1)} + \gamma(\theta_{j,m}^{(q-1)} - \bar{\theta}_{j,m}^{(q)})$

$\theta_{j,m}^{(q)} = \arg\min \mathcal{L}(\theta_{j,m}, \bar{\theta}_{j,m}^{(q-1)}, \omega_{j,m}^{(q-1)})$ subject to $\theta_{1,m}(t) < ... < \theta_{J,m}(t)$ for all t.

until convergence

return Non-overlapping sequence of $\bar{\theta}_{j,m}^{(q)}$ for all j, m.

Figure 4: Trend fit with our ADMM algorithm with 3 windows which converged in 7 iterations compared to trend from simultaneous fit.



The primal residual, r_p, represents the difference between the trend values in the windows and the consensus trend value while the dual residual, r_ds represents the change in the consensus variable from one iterate to the next. The algorithm is stopped when

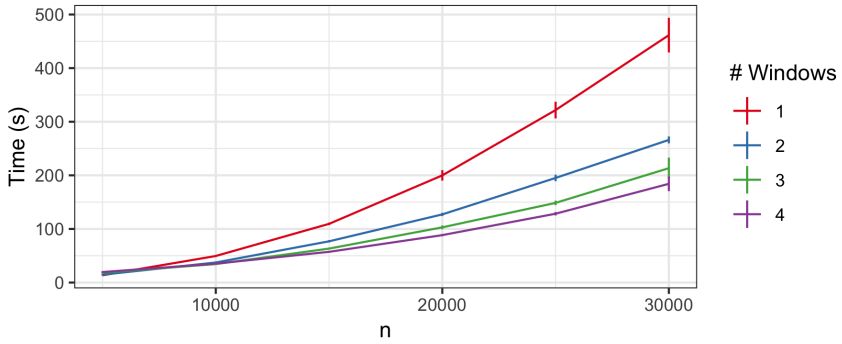
$$r_p^{(q)} < \epsilon_{abs}\sqrt{NJ} + \epsilon_{rel} \max_m \left[\max \left(\sqrt{\sum_{j=1}^{J} \|\theta_{j,m}^{(q)}\|_2^2}, \sqrt{\sum_{j=1}^{J} \|\bar{\theta}_{j,m}^{(q)}\|_2^2} \right) \right] \tag{16}$$

$$r_d^{(q)} < \epsilon_{abs}\sqrt{NJ} + \epsilon_{rel} \sqrt{\sum_{m=1}^{M} \sum_{j=1}^{J} \|\omega_{j,m}^{(q)}\|_2^2} \tag{17}$$

Timing experiments illustrate the advantages of using our ADMM algorithm even on

datasets where solving the problem simultaneously is possible. We use from one to four windows for each data size with an overlap of 500. The windows algorithm was run until the stopping criteria were met using $\epsilon_{abs} = 0.01$ and $\epsilon_{rel} = 0.001$. For each data size, n, 25 datasets were simulated using the peaks simulation design described below and trends for three quantiles were fit simultaneously: 0.05, 0.1, and 0.15 using a $\lambda = n/5$.

Figure 5: Timing experiments comparing quantile trend filtering with varying numbers of windows by data size.



2.3 Regularization Parameter Choice

Our method can easily handle missing data by changing the check loss function to output 0 for missing values. This allows us to leave out validation observations that can be used to select the tuning parameter λ and to compare method performance on real data. A number of methods have been proposed for selecting the quantile regression smoothing spline tuning parameter Yuan (2006). Koenker et al. (1994) relate λ to the number of interpolated points $p_\lambda = \sum I(y_i = \widehat{g}_i(x_i))$, which can be thought of as active knots, they propose the Schwarz criterion for the selection of λ

$$\text{SIC}(p_\lambda) = \log[\frac{1}{n}\rho_\tau(y - \widehat{g}(x))] + \frac{1}{2n}p_\lambda \log n \tag{18}$$

The traditional Bayesian Information Criterion (BIC) is given by

$$\text{BIC}(s) = -2\log(L\{\hat{\theta}(s)\}) + \nu(s)\log n \tag{19}$$

where $\theta(s)$ is the parameter θ with those components outside s being set to 0, and $\nu(s)$ is the number of components in s. If we assume an asymmetric Laplace likelihood $L(y|\theta) =$

$\left(\frac{\tau^n(1-\tau)}{\sigma}\right)^n \exp\left\{-\sum_i \rho_\tau(\frac{y_i-\theta_i}{\sigma})\right\}$ and the number of non-zero elements of $D\theta$ as df

$$\text{BIC}(df) = 2\frac{1}{\sigma}\rho_\tau(y-\theta) + \text{df}\log n \tag{20}$$

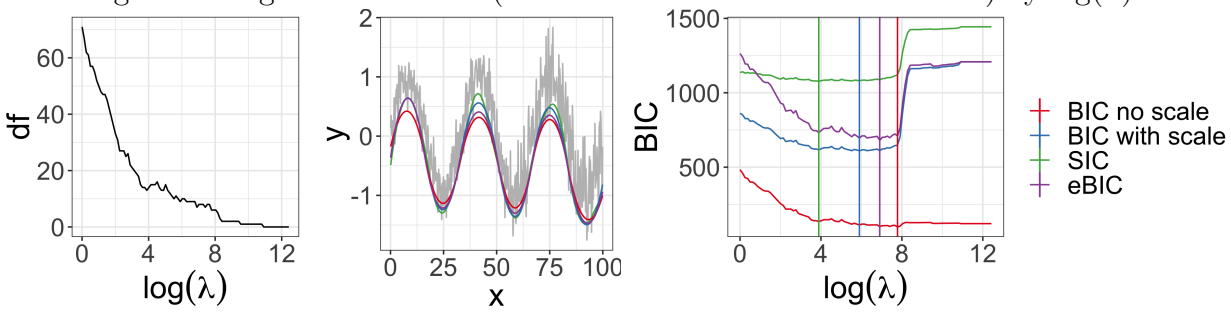
We can choose any $\sigma > 0$ and have found empirically that $\sigma = \frac{1-|1-2\tau|}{2}$ produces stable estimates. Chen and Chen (2008) proposed the extended BIC for large parameter spaces

$$\text{BIC}_\gamma(s) = -2\log(L\{\hat{\theta}(s)\}) + \nu(s)\log n + 2\gamma\log\binom{P}{j,} \quad \gamma \in [0,1] \tag{21}$$

where P is the total number of possible parameters and j is the number of parameters included in given model. We used this criteria with $\gamma = 1$, $P = n - k$ where k is the order of the differencing matrix and $j = \nu(s)$ is the number of non-zero entries in $D^{(k)}\theta$.

We used a single dataset to illustrate the difference between the scaled, unscaled and extended BIC criteria.

Figure 6: Degrees of freedom (number of non-zero elements of $D\theta$) by $\log(\lambda)$.

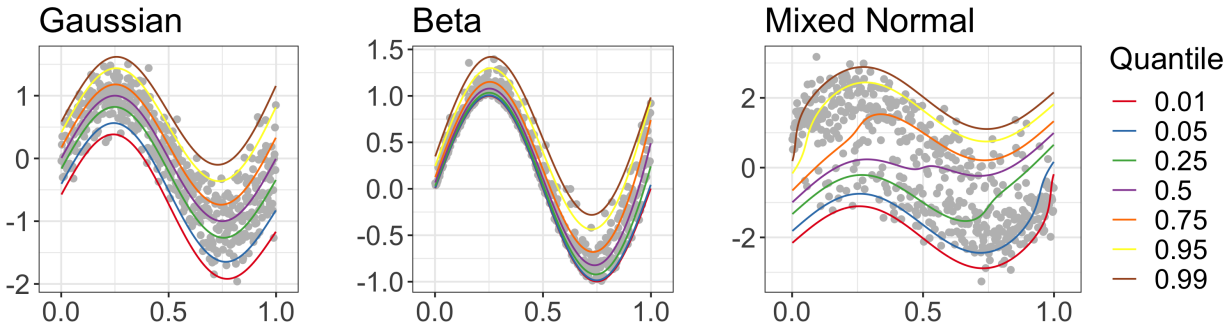


3 Simulation Studies

3.1 Estimating Quantiles

We compare the performance of our quantile trend filtering method with the three previously published methods using designs proposed by Racine and Li (2017). The methods compared are: `npqw` which is the quantile-ll method described in Racine and Li (2017), code was obtained from the author; `qsreg` in the `fields` R package and described in Oh et al. (2011); `rqss` available in the `quantreg` package and described in Koenker et al. (1994). The

Figure 7: Simulated data with true quantiles $\tau \in \{0.01, 0.05, 0.25, 0.5, .75, 0.95, 0.99\}$



smoothing parameter λ for the `rqss` method is chosen using a grid search and minimizing the SIC criteria as described in Koenker et al. (1994). We further compare three criteria for choosing the smoothing parameter for our detrend method: `detrendr_SIC`: Our method where we minimize $\sum_i \rho_\tau(y_i - \theta_i) + \lambda||D\theta||_1$ and λ is chosen using SIC (Koenker et al., 1994). `detrendr_valid`: Our method where lambda is chosen by leaving out every 5th observation as a validation data set and minimizing the evaluating the check loss function evaluated at the validation data. `detrendr_eBIC`: the new criteria we have proposed based on the extended BIC proposed by Chen and Chen (2008).

Three simulation designs from Racine and Li (2017) were considered. For all designs X_i was generated as a uniformly spaced sequence in $[0, 1]$ and the response Y was generated as

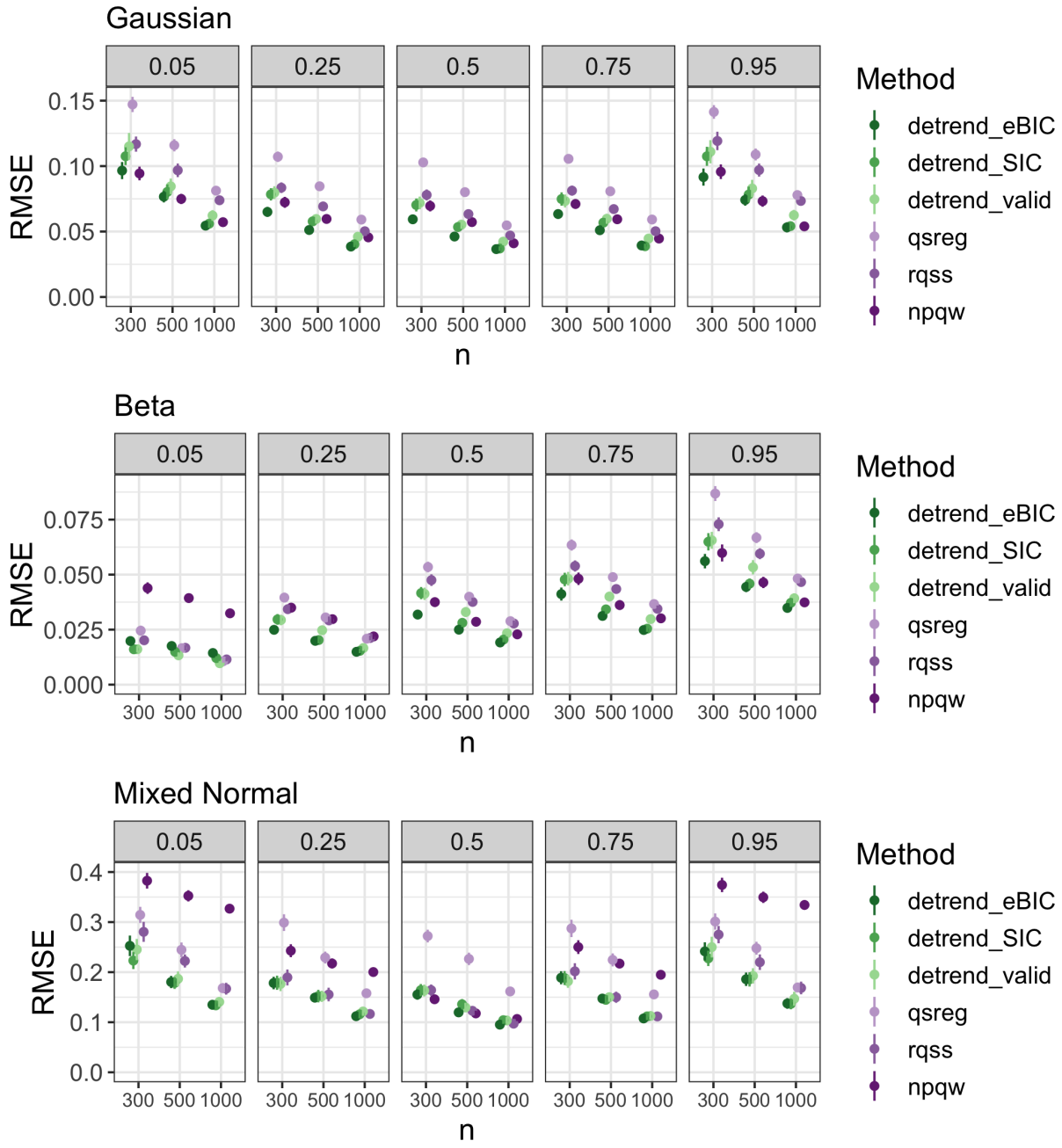
$$Y_i = sin(2\pi x_i) + \epsilon_i(x_i)$$

The three error distributions considered were

- Gaussian: $\epsilon_i(x_i) \sim N\left(0, \left(\frac{1+x_i^2}{4}\right)^2\right)$
- Beta: $\epsilon_i \sim Beta(1, 11 - 10x_i)$
- Mixed normal: ϵ_i is simulated from a mixture of $N(-1, 1)$ and $N(1, 1)$ with mixing probability x_i.

100 datasets were generated of sizes 300, 500 and 1000. The MSE was calculated as $\frac{1}{n}\sum_i(\hat{q}_\tau(x_i) - q_\tau(x_i))^2$. The plots below show the mean MSE $\pm$ twice the standard error by method, quantile level and sample size.

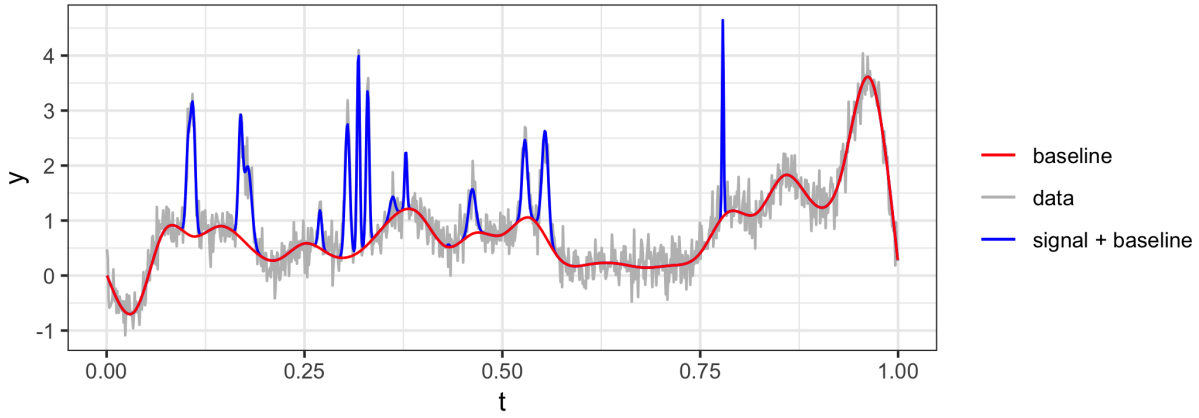
Figure 8: RMSE by design, method, quantile and data size.



In all of the designs the proposed detrend methods are either better than or comparable to existing methods. The npqw method performs particularly poorly in the mixed normal design, due to the fact that it assumes the data comes from a scale-location model which is violated in this case.

3.2 Peak Detection

Figure 9: Example of simulated peaks, baseline, and observed measurements.

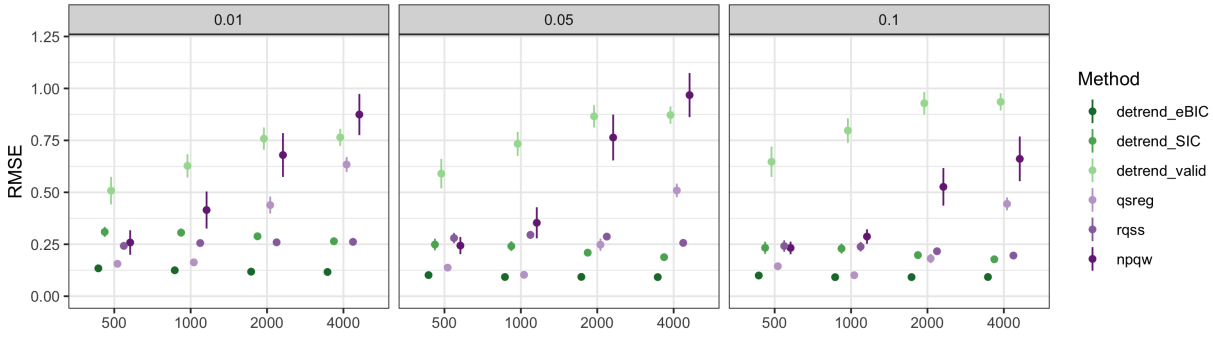


We use another simulation design based on the applied problem we aim to solve. We assume that the measured data can be represented by

$$Y(t) = s(t) + b(t) + \epsilon(t) \tag{22}$$

with $t = 1, ..., n$, where $s(t)$ is the true signal at time t, $b(t)$ is the drift component that varies smoothly over time and $\epsilon(t) \sim N(0, \sigma^2)$ is an error component. We generate $b(t)$ using a cubic natural spline basis function with degrees of freedom sampled from a Poisson distribution with mean parameter equal to $n/100$, and coefficients drawn from an exponential distribution with rate 1. The true signal function is assumed to be zero with peaks generated using the Gaussian density function. The number of peaks is sampled from a binomial distribution with size equal to n and probability equal to 0.005 with location parameters uniformly distributed between 1 and $n - 1$ and bandwidths uniformly distributed between 2 and 12. The peaks were then multiplied by a factor that was randomly drawn from a normal distribution with mean 20 and standard deviation of 4. One hundred datasets

Figure 10: RMSE by method, quantile and data size for peaks design.



were generated for $n = \{500, 1000, 2000, 4000\}$. We compare the ability of the methods to estimate the true quantiles of $b(t) + \epsilon$ for $\tau \in \{0.01, 0.05, 0.1\}$ and calculate the RMSE (Fig. 10).

In our application, we want to accurately classify the observations into signal and no signal using a threshold. To evaluate the accuracy of our method compared to other methods we define true signal as any time when the simulated peak value is greater than 0.5. We compared three different trends for the baseline estimation and four different thresholds for separating the estimated signal after subtracting the estimated baseline from the observations. An illustration of the observations classified as signal after detrending compared to the "true signal" is shown in Fig. 11. To compare the resulting signal classifications we calculate the class averaged accuracy (CAA). Defining $s(t) \in \{0, 1\}$ as the true vector of signal classification and $\hat{s}(t) \in \{0, 1\}$ as the estimated signal classification, the CAA is defined as

$$\text{CAA} = \frac{1}{2}\left(\frac{\sum_{t=1}^{n} \mathbf{I}[s(t) = 1 \cap \hat{s}(t) = 1]}{\sum_{t=1}^{n} \mathbf{I}[s(t) = 1]} + \frac{\sum_{t=1}^{n} \mathbf{I}[s(t) = 0 \cap \hat{s}(t) = 0]}{\sum_{t=1}^{n} \mathbf{I}[s(t) = 0]}\right). \tag{23}$$

We use this metric because our classes tend to be very un-balanced in this case with many more 0s than 1s. The CAA metric will always give a score of 0.5 for random guessing and also for trivial classifiers such as $\hat{s}(t) = 0$ for all t.

Our detrend_BIC method performs the best overall in terms of both RMSE and miss-classification rate. The lowest miss-classification rates were obtained using the detrend_eBIC method and a threshold of 1 for all data sizes. While qsreg was competitive with our method

in some cases, both the RMSE and miss-classification rate increased substantially with the size of the dataset.

Figure 11: Example signal classification using threshold. Red indicates true signal > 0.5, blue indicates observations classified as signal after baseline removal using eBIC detrendr and a threshold of 1.2.

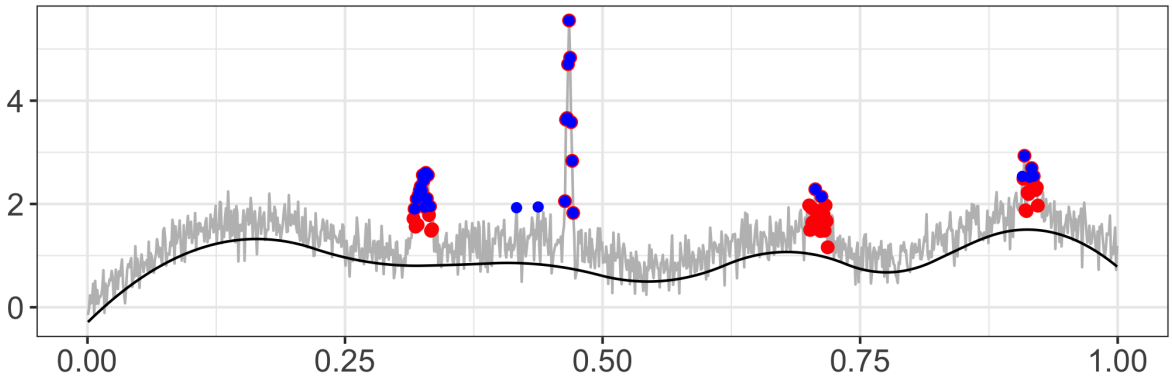
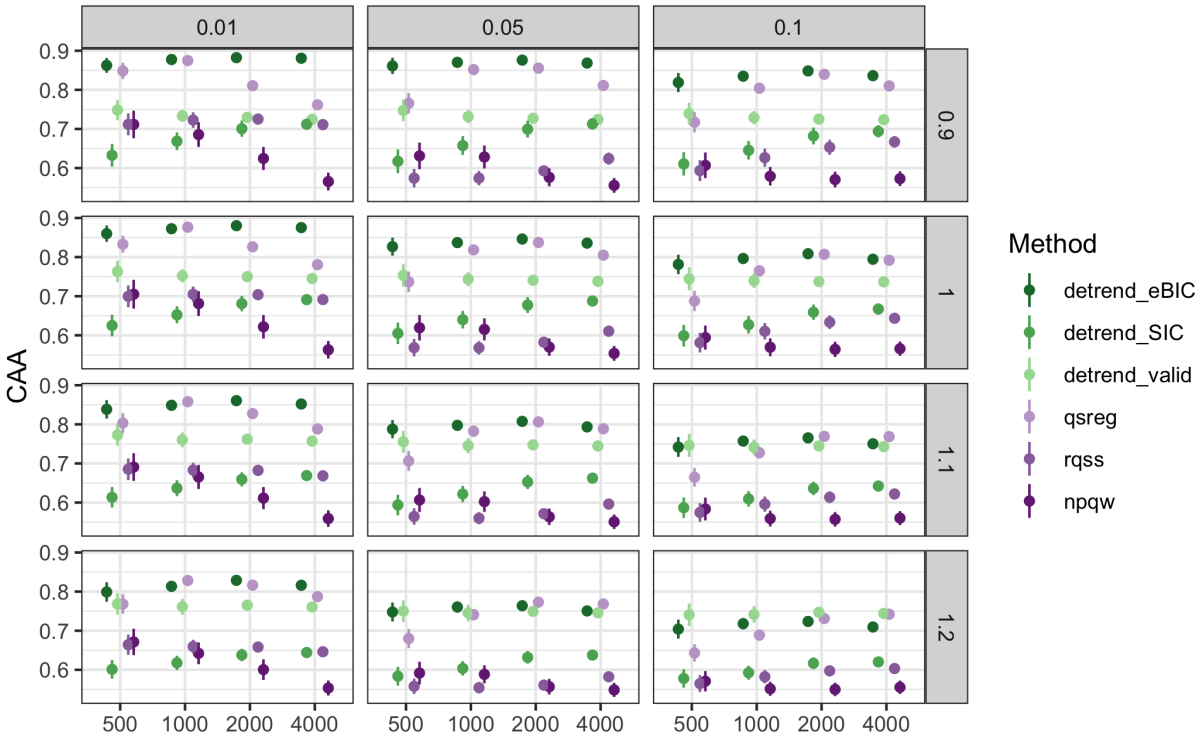


Figure 12: Class averaged accuracy by threshold, data size, and method (1 is best 0.5 is worst).



4 Application

We compared our detrend method with the eBIC criteria with the qsreg method on a subset of the real data since the qsreg method cannot handle all 15 hours simultaneously. We estimate the baseline trend using both the 5^{th}, 10^{th}, and 15^{th} quantiles. We compare three thresholds for separating signal from not signal, the thresholds are calculated using the median and a factor multiple of the median absolute deviation of the detrended series. We report both the confusion matrices and the variation of information (VI) distances.

Our windowed detrend method was used to removed the baseline drift from low cost air quality sensors so that the signal could be categorized using a simple threshold. The measurements were first standardized to have mean zero and variance 1. Three quantile levels for estimating the baseline trend were compared. The total dataset consisted of 52,322 observations per node. The signal thresholds were set using the first 15,000 observations

Figure 13: Variation of Information between sensor nodes after trend removal by quantile and method and thresholding by factor of MAD.

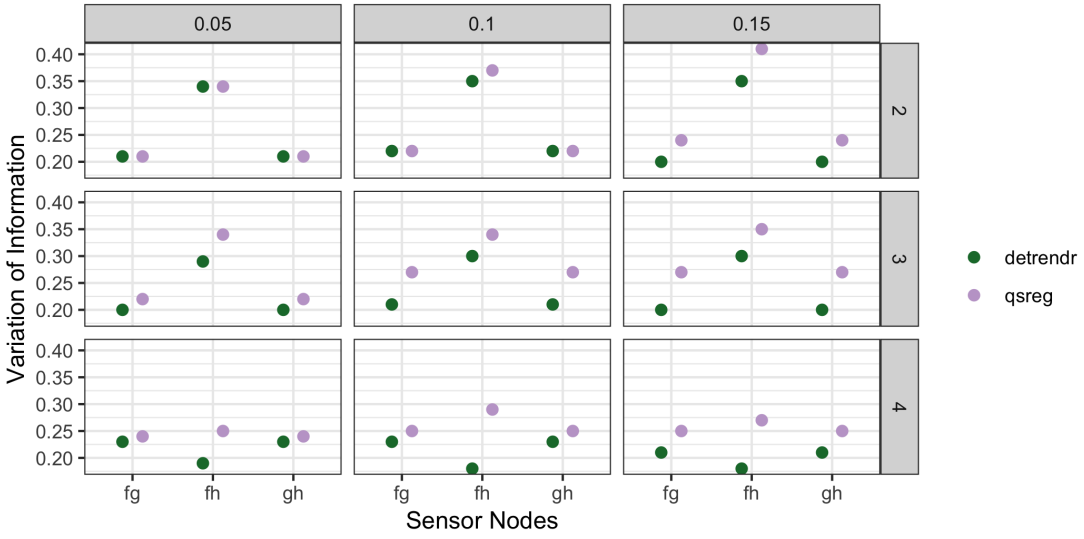


Figure 14: Rugplot showing locations of signal after baseline removal using detrendr estimate of 10th quantile.

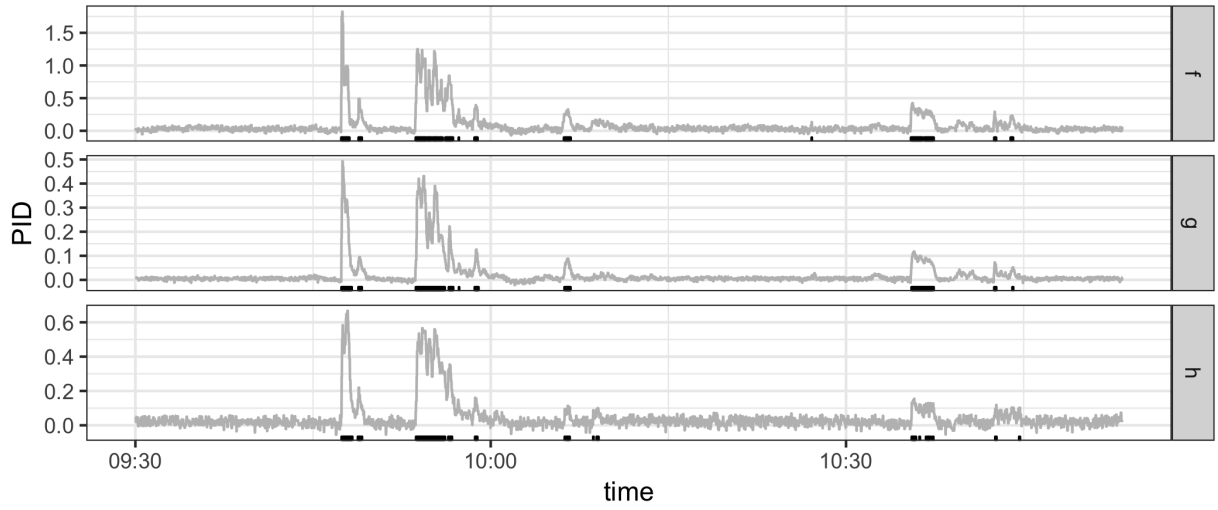


Figure 15: Low cost sensor data after drift removal using windowed detrend with eBIC.

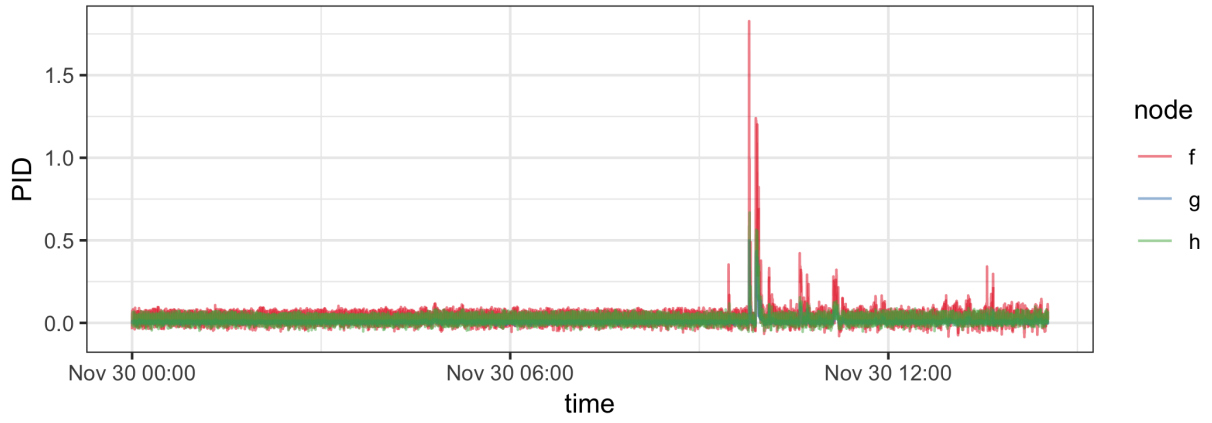


Table 1: Confusion matrices for 3 SPod nodes after baseline removal using 15th quantil and threshold of 3*MAD (n=52322).

	f = 0		**f = 1**	
	h = 0	h = 1	h = 0	h = 1
g = 0	50779	72	186	14
g = 1	336	67	314	554

where it was known no signal was present. The thresholds were set as 3 times the standard deviation plus the mean of observations in this time period. The total number of seconds of signal for each node as well as the number of seconds where multiple nodes both reported signal is shown in Table 1. Table 2 shows the fraction of observations with different signal classifications by node combination. The BVI scores for the full dataset were 0.14, 0.13, and 0.14 for nodes f and g, f and h, and g and h, respectively.

5 Conclusion

SUPPLEMENTARY MATERIAL

R-package for detrend routine: R-package detrendr containing code to perform the

diagnostic methods described in the article. The package also contains all datasets used as examples in the article. (GNU zipped tar file)

6 References

References

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011), "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, 3, 1–122.

Brantley, H., Hagler, G., Kimbrough, E., Williams, R., Mukerjee, S., and Neas, L. (2014), "Mobile air monitoring data-processing strategies and effects on spatial air pollution trends," *Atmospheric measurement techniques*, 7, 2169–2183.

Chen, J. and Chen, Z. (2008), "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, 95, 759–771.

Gabay, D. and Mercier, B. (1975), *A dual algorithm for the solution of non linear variational problems via finite element approximation*, Institut de recherche d'informatique et d'automatique.

Glowinski, R. and Marroco, A. (1975), "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires," *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9, 41–76.

Gurobi Optimization, L. (2018), "Gurobi Optimizer Reference Manual," .

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009), "ℓ_1 Trend Filtering," *SIAM Review*, 51, 339–360.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., Ng, P., and Portnoy, S. (1994), "Quantile smoothing splines," *Biometrika*, 81, 673–680.

Marandi, R. Z. and Sabzpoushan, S. (2015), "Qualitative modeling of the decision-making process using electrooculography," *Behavior research methods*, 47, 1404–1412.

Ning, X., Selesnick, I. W., and Duval, L. (2014), "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems*, 139, 156 – 167.

Nychka, D., Gray, G., Haaland, P., Martin, D., and O'connell, M. (1995), "A nonparametric regression approach to syringe grading for quality improvement," *Journal of the American Statistical Association*, 90, 1171–1178.

Oh, H.-S., Lee, T. C. M., and Nychka, D. W. (2011), "Fast Nonparametric Quantile Regression With Arbitrary Smoothing Methods," *Journal of Computational and Graphical Statistics*, 20, 510–526.

Pettersson, K., Jagadeesan, S., Lukander, K., Henelius, A., Hæggström, E., and Müller, K. (2013), "Algorithm for automatic analysis of electro-oculographic data," *Biomedical engineering online*, 12, 110.

Racine, J. S. and Li, K. (2017), "Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach," *Journal of Econometrics*, 201, 72–94.

Snyder, E., Watkins, T., Solomon, P., Thoma, E., Williams, R., Hagler, G., Shelow, D., Hindin, D., Kilaru, V., and Preuss, P. (2013), "The changing paradigm of air pollution monitoring." *Environmental science & technology*, 47, 11369.

Thoma, E. D., Brantley, H. L., Oliver, K. D., Whitaker, D. A., Mukerjee, S., Mitchell, B., Wu, T., Squier, B., Escobar, E., Cousett, T. A., et al. (2016), "South Philadelphia passive sampler and sensor study," *Journal of the Air & Waste Management Association*, 66, 959–970.

Tibshirani, R. J. (2014), "Adaptive piecewise polynomial estimation via trend filtering," *The Annals of Statistics*, 42, 285–323.

Yamada, H. (2017), "Estimating the trend in US real GDP using the 1 trend filtering," *Applied Economics Letters*, 24, 713–716.

Yuan, M. (2006), "GACV for quantile smoothing splines," *Computational statistics & data analysis*, 50, 813–829.