Baseline Drift Estimation for Time Series Data Using Quantile Trend Filtering

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

Abstract

We address the problem of estimating smoothly varying baseline trends in time series data using quantile trend filtering. We first extend the basic framework to ensure non-crossing while estimating multiple quantile trends simultaneously. We also implement a parallelizable alternating direction method of moments (ADMM) algorithm for estimating trends in longer series. The ADMM algorthim enables the estimation of trends in a piecewise manner, both reducing the computation time and extending the limits of the method to larger data sizes. We also address smoothing parameter selection and propose a modified criterion based on the extended Bayesian Information Criterion. Through simulation studies and an application to low cost air quality sensor data, we demonstrate that our model provides better quantile trend estimates than existing methods and improves signal classification of low-cost air quality sensor output.

Keywords: quantile regression, non-parametric, trend estimation, smoothing splines

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

†Department of Statistics and Data Science, 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 applications spanning 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 and other more rapidly varying components. Kim et al. (2009) proposed ℓ_1 *trend filtering* to estimate trends that are piece-wise linear or polynomial, assuming that the series consists of a trend plus uncorrelated noise. Tibshirani (2014) later demonstrated empirically that the trend filtering estimates adapt to the local level of smoothness better than smoothing splines. In the trend filtering problem, given observations $y(t)$ with $t = 1, ..., N$, 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 $y \in \mathbb{R}^n$ is the vector of observations, $\lambda \geq 0$ 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, $\|\mathbf{D}^{(1)}\theta\|_1 = \sum_{i=1}^{n-1}|\theta_{i+1} - \theta_i|$ which is just total variation in one dimension. The penalty incentivizes solutions that 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 piece-wise polynomials of order k.

Tibshirani (2014) showed that trend filtering is both fast and locally adaptive when the time series consists of only the trend and random noise. However, in some applications,

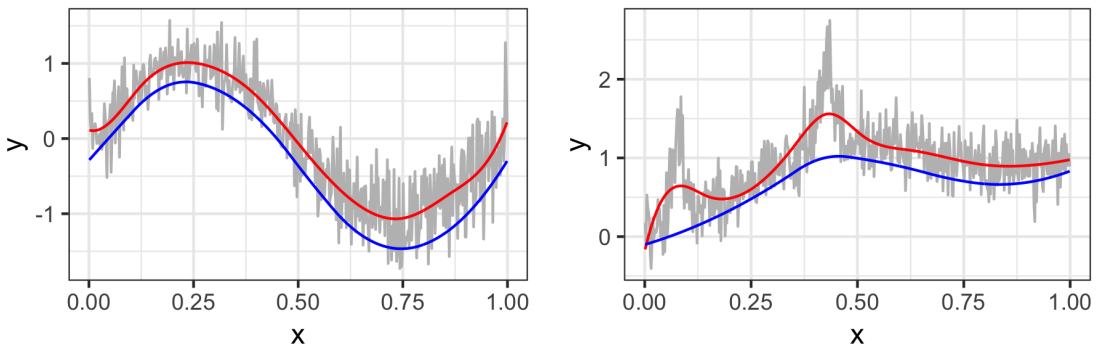

Figure 1: Examples of trend filtering solutions (red) and 15^{th} quantile trending filtering solution (blue).

such as the air quality monitoring problem considered in this paper, the data contain a rapidly varying signal in addition to the slowly varying trend and noise. Standard trend filtering struggles to distinguish between the slowly varying trend and the rapidly-varying signal. We propose quantile trend filtering as a more robust method for detrending in these situations (Fig. 1). To estimate the trend in the τ^{th} quantile, we minimize the criterion

$$\min_{\theta} \rho_\tau(y-\theta) + \lambda \|\mathbf{D}^{(k+1)}\theta\|_1, \tag{4}$$

where $\rho_\tau(z)$ is the check-loss function:

$$\rho_\tau(z) = \sum_{i=1}^{n} z_i(\tau - \mathbf{I}(z_i < 0)), \tag{5}$$

and $\mathbf{I}(\cdot)$ is the indicator function. This formulation was proposed by Kim et al. (2009) as a possible extension of ℓ_1-trend filtering but not studied. Here, we extend the basic framework to model multiple quantiles simultaneously. We also implement non-crossing constraints to ensure valid trend estimates. To reduce computation time and extend the method to larger series, we develop and a parallelizable ADMM algorithm. Finally, we propose a modified criterion for choosing the smoothing parameter.

Our research is motivated by the analysis of low cost and portable air quality sensors, whose use 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

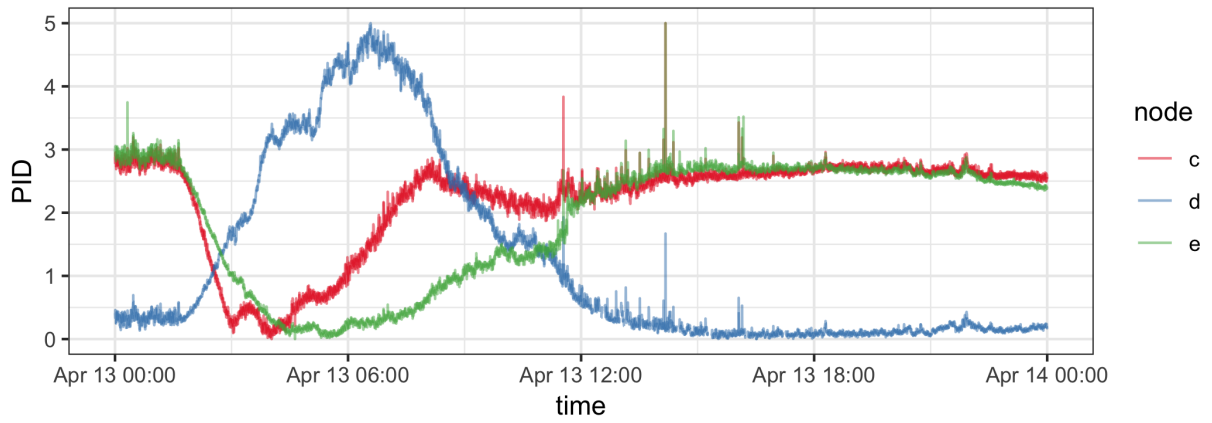

Figure 2: Example of 3 co-located SPod PID sensor readings.

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 temperature and 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. Figure 2 provides an example of measurements from three SPod sensor nodes co-located at the border of an industrial facility. All of the sensors respond to the pollutant signal, but the amount of baseline drift varies from one node to another.

In this application, as well as those described in Ning et al. (2014), Marandi and Sabzpoushan (2015), and Pettersson et al. (2013), the goal is to estimate the trend in the baseline not the mean. We can define the trend in the baseline as the trend in a low quantile of the data. A variety of methods for estimating quantile trends have already been proposed. Koenker and Bassett (1978) were the first to propose substituting the squared-loss penalty with the check-loss penalty (Eq. 5) to estimate a conditional quantile instead of the conditional mean. Later, Koenker et al. (1994) proposed quantile trend filtering with $k = 2$ producing quantile trends that are piecewise linear, but did not extent it to higher orders. Rather than using the ℓ_1-norm to penalize the discrete differences, Nychka et al. (1995) used the smoothing spline penalty based on the ℓ_2-norm:

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

where $f(t)$ is a smooth function of time and λ is a tuning parameter that controls the degree of smoothing. Oh et al. (2011) proposed an algorithm for solving the quantile smoothing spline problem by approximating the check-loss function with a similar differentiable function. Racine and Li (2017) take a different approach and constrain the response to follow a smooth location scale model of the form $y(t) = a(t)+b(t)\epsilon_i$ and estimate the τ^{th} conditional quantile using a kernel estimator and local linear approach.

In the following sections we describe in detail our quantile trend filtering algorithms and regularization parameter criterion (Section 2). We demonstrate through simulation studies that our proposed model provides better or comparable estimates of non-parametric quantile trends than existing methods (Section 3). We further show that quantile trend filtering is a more effective method of drift removal for low-cost air quality sensors and results in improved signal classification compared to quantile smoothing splines (Section 4). Finally, we discuss potential extensions of quantile trend filtering (Section 5).

2 Methods

2.1 Quantile Trend Filtering

We combine the ideas of quantile regression and trend filtering. For a single quantile level τ, the quantile trend filtering problem is given in Eq. 4. As with classic quantile regression, the quantile trend filtering problem is a linear program which can be solved by a number of methods. 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

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

$$\text{subject to: } \theta_1(t) \leq \theta_2(t) \leq \ldots \leq \theta_J(t) \text{ for all } t, \tag{7}$$

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) and its R package implementation. However, we could easily substitute a free solver such as the Rglpk package by Theussl and Hornik (2017).

2.2 ADMM for Big Data

The number of parameters to be estimated in this problem 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. 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 ADMM algorithm for solving large problems in a piece-wise fashion. The ADMM algorithm (Gabay and Mercier, 1975; Glowinski and Marroco, 1975) is fully described by Boyd et al. (2011). We apply the consensus ADMM algorithm to the quantile regression trend filtering problem given in Eq. (4), by dividing our observed series $y(t)$ with $t = \{1, ..., n\}$ into overlapping windows

$$\begin{aligned} y_1(t) &= y(t) \text{ if } 1 \leq t \leq u_1 \\ y_2(t) &= y(t) \text{ if } l_2 \leq t \leq u_2 \\ &\cdots \\ y_M(t) &= y(t) \text{ if } l_M \leq t \leq n \end{aligned}$$

with boundaries $1 < l_2 < u_1 < l_3 < u_2 < l_4 < u_3 < ... < n$. An illustration is given in Fig. 3. Given quantiles $\tau_1 < ... < \tau_J$, we define $\theta_{j,m}(t)$ as the value of the $\tau_j{}^{\text{th}}$ quantile trend in

window m at time t. In order to constrain the overlapping sections to be equal we define a consensus variable

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

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

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

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

subject to:

$$\theta_{1,m}(t) \leq \theta_{2,m}(t) \leq \ldots \leq \theta_{J,m}(t) \text{ for all } m, t$$

$$\theta_{j,m}(t) = \bar{\theta}_{j,m}(t) \text{ for all } j, m, t$$

Defining Lagrange multipliers $\omega_{j,m}$, and penalty parameter $\gamma > 0$, the augmented Lagrangian for finding the trends in window m is

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

We then estimate the trend separately in each window, which can be done in parallel, while constraining the overlapping pieces of the trends to be equal as outlined in Algorithm 1. The use of the Augmented Lagrangian converts the problem from a linear program into a quadratic program. The `gurobi` R package (Gurobi Optimization, 2018) can solve quadratic programs in addition to linear programs, but we can also use the free R package `quadprog` (Weingessel and Turlach, 2013).

We measure convergence using the stopping criteria described by Boyd et al. (2011). The criteria are based on the primal and dual residuals, which represent the residuals for

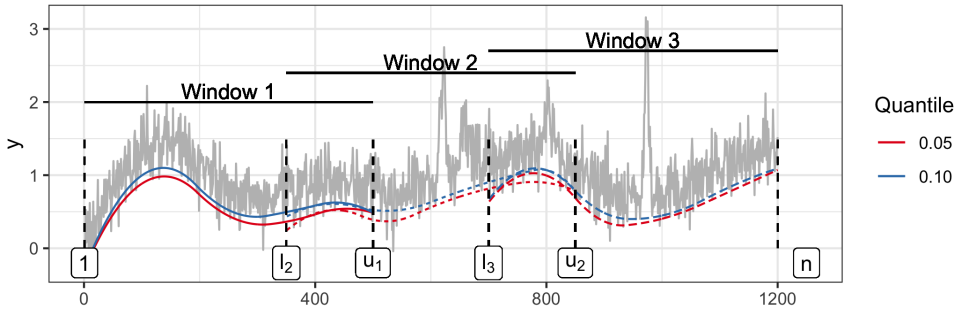

Figure 3: Window boundaries and trends fit separately in each window.

Algorithm 1 ADMM algorithm for quantile trend filtering with windows

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

initialize:

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

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

repeat

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

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

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

until convergence

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

primal and dual feasibility, respectively. The primal residual,

$$r_p^{(q)} = \sqrt{\sum_{m=1}^{M}\sum_{j=1}^{J} \|\theta^{(q)}_{j,m} - \bar{\theta}^{(q)}_{j,m}\|_2^2}, \tag{8}$$

represents the difference between the trend values in the windows and the consensus trend value. The dual residual,

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

represents the change in the consensus variable from one iterate to the next. The algorithm

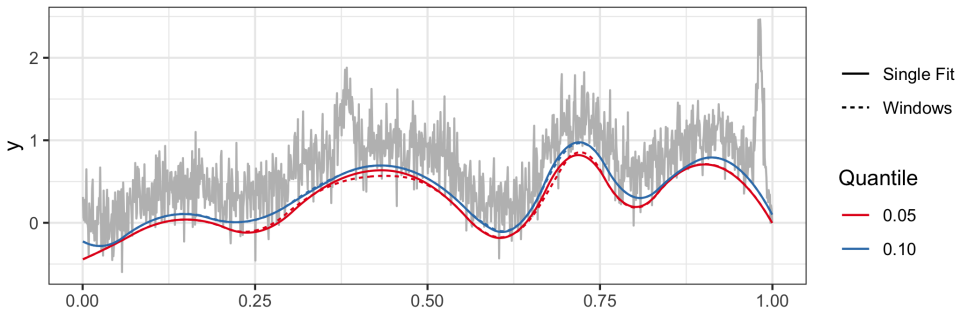

Figure 4: Trend fit with our ADMM algorithm with 3 windows which converged in 7 iterations compared to trend from simultaneous fit.

is stopped when

$$r_p^{(q)} < \epsilon_{abs}\sqrt{nJ} + \epsilon_{rel}\max_m \left[\max\left(\sqrt{\sum_{j=1}^{J} \|\theta_{j,m}^{(q)}\|_2^2}, \sqrt{\sum_{j=1}^{J} \|\bar{\theta}_{j,m}^{(q)}\|_2^2} \right) \right] \tag{9}$$

$$r_d^{(q)} < \epsilon_{abs}\sqrt{nJ} + \epsilon_{rel}\sqrt{\sum_{m=1}^{M}\sum_{j=1}^{J} \|\omega_{j,m}^{(q)}\|_2^2} \tag{10}$$

Timing experiments illustrate the advantages of using our ADMM algorithm when the trend can be estimated with a single window. For each data size, n, 25 datasets were simulated using the peaks simulation design described below. Trends for three quantiles were fit simultaneously: 0.05, 0.1, and 0.15 using $\lambda = n/5$. 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 $\epsilon_{rel} = 0.001$. As is shown in Fig. 5, using 4 windows instead of one on data sizes of 55000 provides a factor of 3 decrease in computation time. The timing experiments were conducted on an Intel Xeon based Linux cluster using two processor cores.

2.3 Regularization Parameter Choice

An important problem in trend estimation is the choice of regularization parameter, or degree of smoothness. Our method can easily handle missing data by defining the check loss function to output 0 for missing values, allowing us to implement cross-validation.

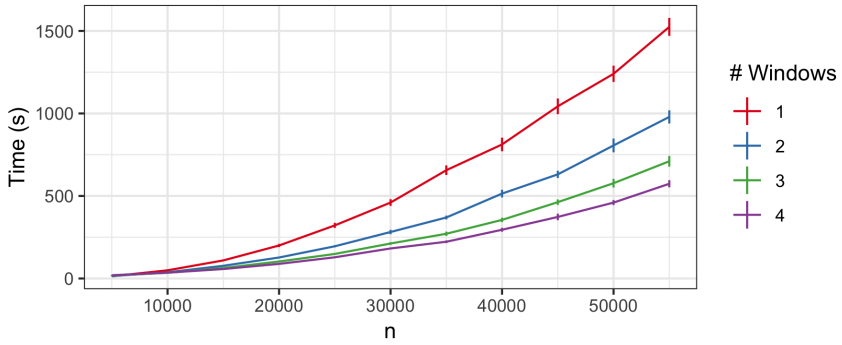

Figure 5: Timing experiments comparing quantile trend filtering with varying numbers of windows by data size.

We also explore three information criteria. Koenker et al. (1994) addressed the choice of regularization parameter by proposing the Schwarz criterion for the selection of λ

$$\text{SIC}(p_\lambda) = \log\left[\frac{1}{n}\rho_\tau(y-\theta)\right] + \frac{1}{2n}p_\lambda \log n. \tag{11}$$

where $p_\lambda = \sum_t I(y(t) = \widehat{\theta}(t))$ is the number of interpolated points, which can be thought of as active knots. The SIC is based on the traditional Bayesian Information Criterion,

$$\text{BIC}(\nu) = -2\log(L\{\hat{\theta}\}) + \nu \log n \tag{12}$$

where L is the likelihood function and ν is the number of non-zero components in $\hat{\theta}$. If we take the approach used in Bayesian quantile regression (Yu and Moyeed, 2001), and view minimizing the checkloss function as maximizing the asymmetric Laplace likelihood,

$$L(y|\theta) = \left(\frac{\tau^n(1-\tau)}{\sigma}\right)^n \exp\left\{-\sum_t \rho_\tau\left(\frac{y(t)-\theta(t)}{\sigma}\right)\right\}, \tag{13}$$

we can compute the BIC as

$$\text{BIC}(\nu) = 2\frac{1}{\sigma}\rho_\tau(y-\hat{\theta}) + \nu \log n \tag{14}$$

where ν is the number of non-zero elements of $\mathbf{D}^{(k+1)}\hat{\theta}$. We can choose any $\sigma > 0$ and have found empirically that $\sigma = \frac{1-|1-2\tau|}{2}$ produces stable estimates.

Another criteria, the extended Bayesian Information Criteria (eBIC), specifically designed for large parameter spaces was proposed by Chen and Chen (2008).

$$\text{eBIC}_\gamma(s) = -2\log(L\{\hat{\theta}\}) + \nu \log n + 2\gamma \log\binom{P}{\nu}, \quad \gamma \in [0,1] \tag{15}$$

where P is the total number of possible parameters and ν is the number of non-zero parameters included in given model. We used this criteria with $\gamma = 1$, and $P = n - k - 1$. We use a single dataset from our simulation study to illustrate the difference between the scaled, unscaled ($\sigma = 1$) and scaled extended BIC criteria in Fig. 6. In the simulation study, we compare the performance of the SIC, scaled eBIC, and validation methods.

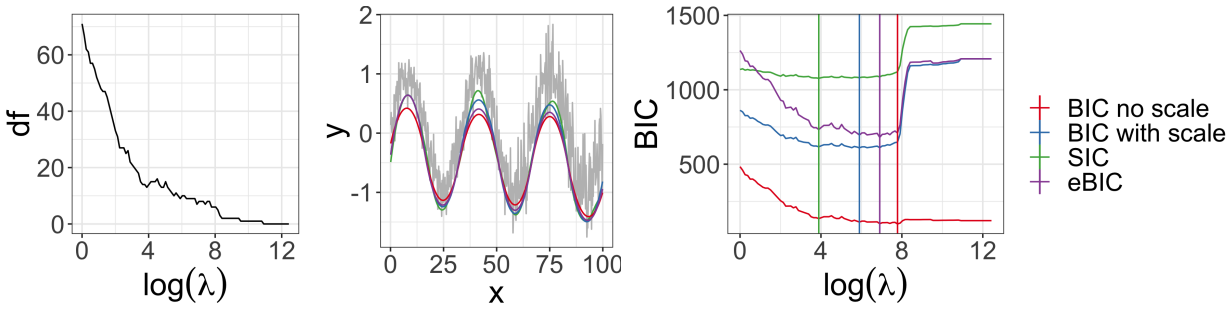

Figure 6: (Left) Degrees of freedom (number of non-zero elements of $D\theta$) by $\log(\lambda)$. (Middle) Estimated 10th quantile trend with regularization parameter chosen using various criterion. (Right) Criterion values by $\log(\lambda)$ vertical lines indicate locations of minima.

3 Simulation Studies

We conduct two simulation studies to compare the performance of our quantile trend filtering method and regularization parameter selection criteria with previously published methods. The first study compares the method's ability to estimate quantiles when the only components of the observed series are a smooth trend and a random component. The second study is based on our application and compares the method's ability to estimate baseline trends and enable peak detection when the time series contains a non-negative signal component in addition to the trend and random component.

We compare the performance of our quantile trend filtering method with three previously published methods: `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); Nychka et al. (1995); `rqss` available in the `quantreg` package and described in Koenker et al. (1994). The regularization parameter λ for the `rqss` method is

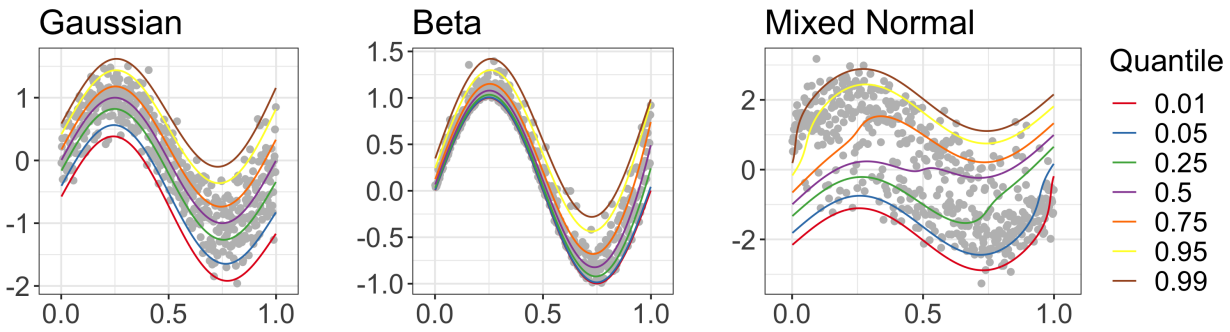

Figure 7: Simulated data with true quantile trends.

chosen using a grid search and minimizing the SIC criteria as described in Koenker et al. (1994), the regularization parameter for `qsreg` was chosen using generalized cross-validation based on the quantile criterion Oh et al. (2011).

We also compare three criteria for choosing the smoothing parameter with our quantile trend filtering method with a single window: `detrendr_SIC` λ chosen using SIC (Eq. 11) (Koenker et al., 1994); `detrendr_valid`: λ is chosen by leaving out every 5th observation as a validation data set and minimizing the check loss function evaluated at the validation data; `detrendr_eBIC`: the proposed scaled eBIC criteria (Eq. 15).

3.1 Estimating Quantiles

To compare performance in estimating quantile trends, three simulation designs from Racine and Li (2017) were considered. For all designs $t = 1, ..., n$, $x(t) = t/n$, and the response y was generated as

$$y(t) = sin(2\pi x(t)) + \epsilon(x(t))$$

The three error distributions considered were

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

One hundred datasets were generated of sizes 300, 500 and 1000. For each method quantile trends were estimated for $\tau = \{0.05, 0.25, 0.5, 0.75, 0.95\}$. Only our detrend methods guarantee non-crossing quantiles. For each quantile trend and method the root mean squared error was calculated as RMSE $= \sqrt{\frac{1}{n}\sum_t (\hat{q}_\tau(t) - q_\tau(t))^2}$. The mean RMSE $\pm$ twice the standard error for each method, quantile level and sample size is shown in Figure 8. In all three designs the proposed detrend methods are either better than or comparable to existing methods. Overall the `detrend_eBIC` performs best, and especially in the mixed normal design our methods have lower RMSEs for the 5^{th} and 95_{th} quantiles. 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

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

$$Y(t) = s(t) + b(t) + \epsilon(t) \tag{16}$$

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 $\epsilon(t) \sim N(0, 0.25^2)$ is an error component. We generate $b(t)$ using cubic natural spline basis functions 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 simulated peaks were multiplied by a factor that was randomly drawn from a normal distribution with mean 20 and standard deviation of 4. One hundred datasets were generated for each $n = \{500, 1000, 2000, 4000\}$. We compare the ability of the methods to estimate the true quantiles of $Y(t) - s(t)$ for $\tau \in \{0.01, 0.05, 0.1\}$ and calculate the RMSE (Fig. 9). In this simulation study our `detrend_eBIC` method outperforms the others substantially. The `qsreg` method is comparable to the `detrend_eBIC` method on

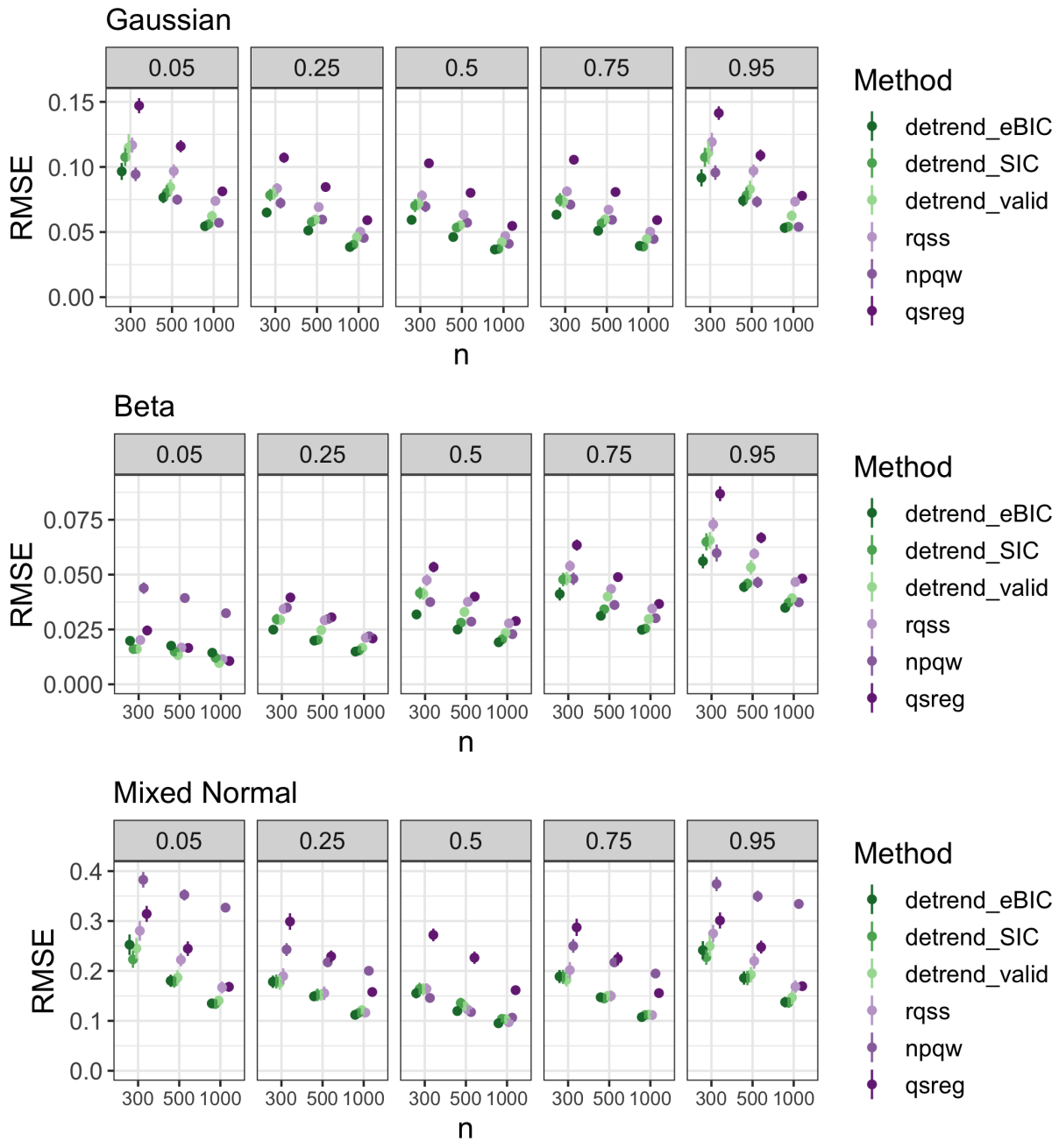

Figure 8: RMSE by design, method, quantile and data size. Points and error bars represent mean RMSE ± twice the standard error.

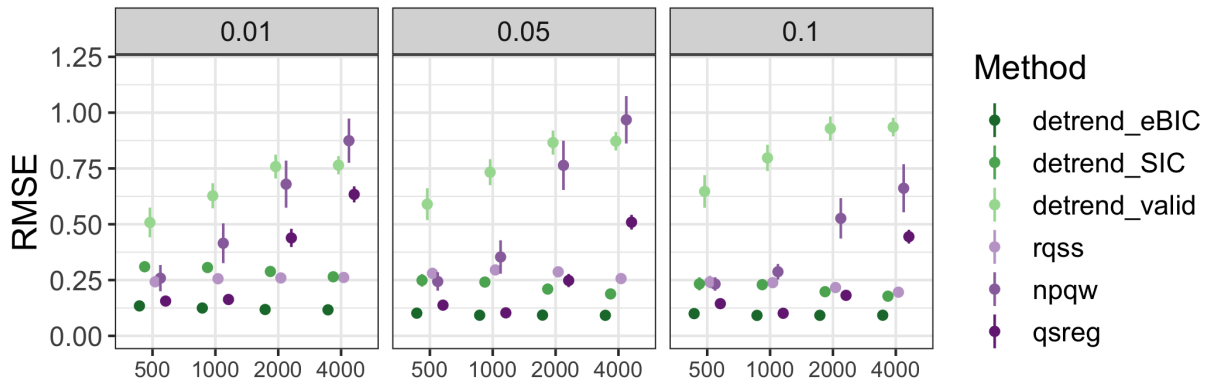

Figure 9: RMSE by method, quantile and data size for peaks design.

the smaller datasets but its performance deteriorates as the data size grows. The `npqw` and `detrend_valid` methods both perform poorly on this design.

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 point when the simulated peak value is greater than 0.5. We compare three different quantiles for the baseline estimation and four different thresholds for classifying the signal after subtracting the estimated baseline from the observations. An illustration of the observations classified as signal after subtracting the baseline trend compared to the "true signal" is shown in Fig. 10. To compare the resulting signal classifications we calculate the class averaged accuracy (CAA). Defining $s(t) \in \{0, 1\}$ as the vector of true signal classification and $\hat{s}(t) \in \{0, 1\}$ as the estimated signal classification, the CAA is defined as

$$\text{CAA} = \frac{1}{2}\left(\frac{\sum_{t=1}^{n} \mathbf{I}[s(t) = 1 \cap \hat{s}(t) = 1]}{\sum_{t=1}^{n} \mathbf{I}[s(t) = 1]} + \frac{\sum_{t=1}^{n} \mathbf{I}[s(t) = 0 \cap \hat{s}(t) = 0]}{\sum_{t=1}^{n} \mathbf{I}[s(t) = 0]}\right). \tag{17}$$

We use this metric because our classes tend to be very un-balanced 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 CAA. While `qsreg` was competitive with our method in some cases, in the majority of cases the largest CAA values for each threshold were produced using the `detrend_eBIC` method with

the 1^{st} or 5^{th} quantiles.

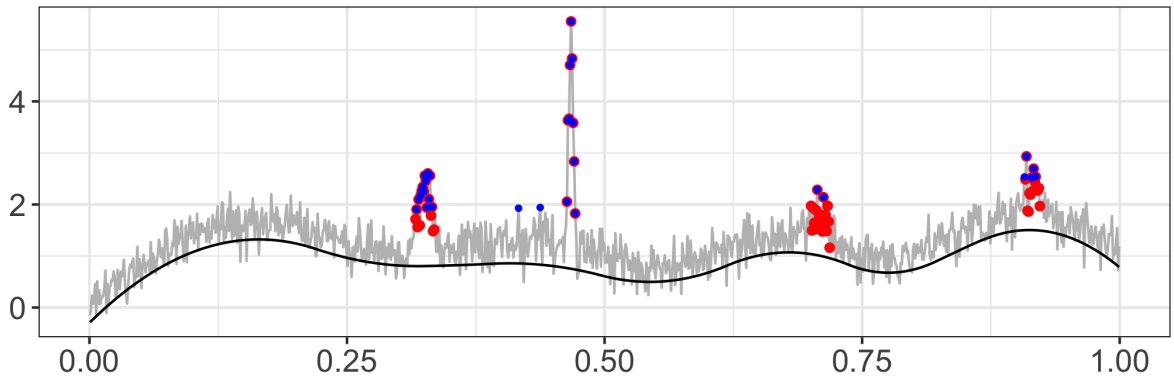

Figure 10: Example signal classification using threshold. Red indicates true signal > 0.5, blue indicates observations classified as signal after baseline removal using `detrend_eBIC` and a threshold of 1.2.

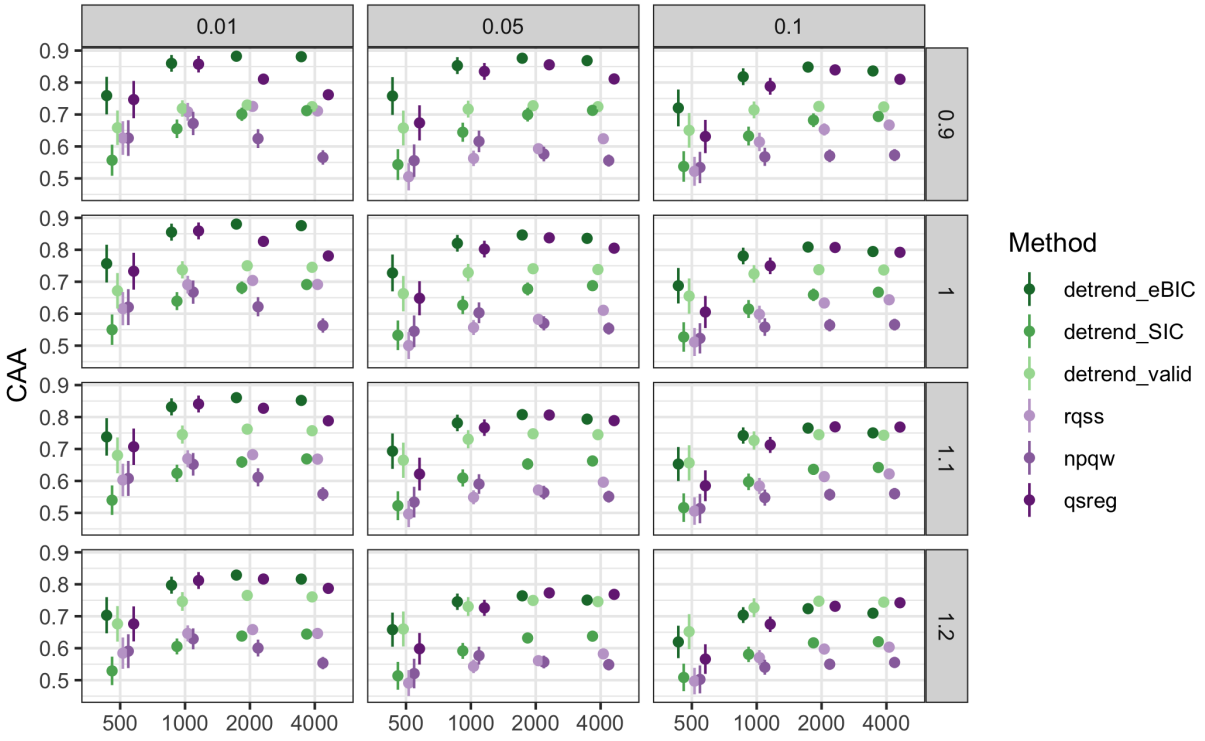

Figure 11: Class averaged accuracy by threshold, data size, and method (1 is best 0.5 is worst).

4 Application

The low-cost "SPod" air quality sensors output a time series that includes a slowly varying baseline, high frequency random noise, and the sensor response to pollutants. A potential use for these sensors is to monitor pollutant concentrations at the fence lines of industrial facilities and detect time points when high concentrations are present. Ideally, three co-located and time aligned sensors (as shown in Fig. 2) responding to a pollutant plume would result in the same signal classification after baseline trend removal and proper threshold choice.

We compare our `detrend_eBIC` method with the `qsreg` method on a subset of the SPod data (n=6000) since the `qsreg` method cannot handle all 24 hours simultaneously. We estimate the baseline trend using the 10^{th} and 15^{th} quantiles and compare three thresholds for classifying signal. The thresholds are calculated using the median plus a multiple of

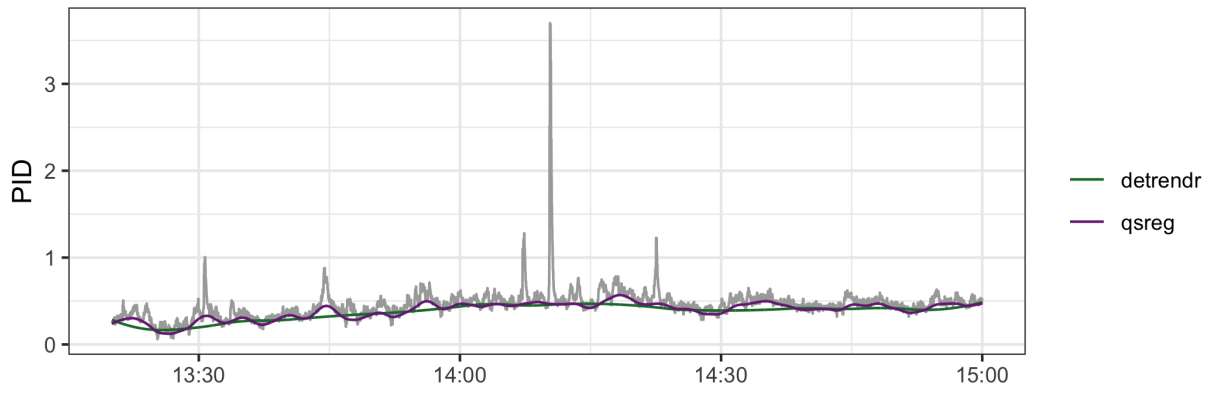

Figure 12: Estimated 15^{th} quantile trends on subset of the data using `qsreg` and `detrend_eBIC`.

the median absolute deviation (Eq. 18) of the detrended series.

$$\text{MAD} = \text{median}(y - \tilde{y}), \tag{18}$$

where $\tilde{y}$ is the median of y. Given a method, quantile level, and MAD multiple, we estimate the quantile trend for each of the sensor nodes and subtract it from the observations. We then calculate the threshold using the median plus the MAD multiple of the corrected series and classify the corrected series based on the threshold. An example of the estimated baseline fit for each method is shown in Fig. 12, while Fig. 13 shows the series after subtracting the `detrend_eBIC` estimate of the 15^{th} quantile and classifying the signal using a MAD multiple of 3.

Given the signal classifications for node 1, $s_1(t) \in \{0, 1\}$ and node 2, $s_2(t) \in \{0, 1\}$ we want to compare the similarity between the two classifications. One metric for evaluating the distance between two classifications is the variation of information (VI):

$$\begin{aligned} r_{ij} =& \frac{1}{n} \sum_t \mathbf{I}(s_1(t) = i \cap s_2(t) = j) \\ VI(s_1, s_2) =& -\sum_{i,j} r_{ij} \left[\log \left(\frac{r_{ij}}{\frac{1}{n}\sum_t \mathbf{I}(s_1(t) = i)} \right) + \log \left(\frac{r_{ij}}{\frac{1}{n}\sum_t \mathbf{I}(s_2(t) = j)} \right) \right] \end{aligned}$$

The VI is a distance metric for measuring similarity of classifications and will be 0 if the classifications are identical and increase as the classifications become more different. The

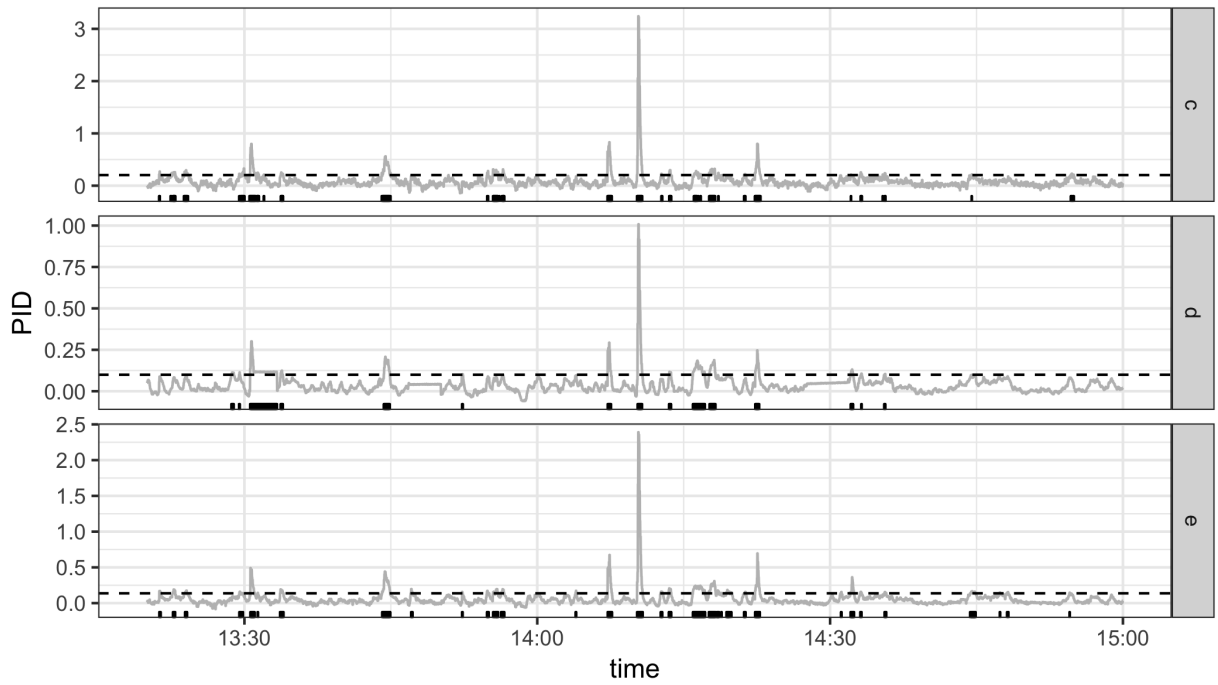

Figure 13: Rugplot showing locations of signal after baseline removal using detrendr estimate of 15th quantile. Horizontal dashed lines represent the thresholds.

VIs by method, quantile and trend are shown in Fig. 14. In all cases our `detrend` method results in classifications that are more similar than those from the `qsreg` method. As is illustrated in Fig. 12 and Fig. 13, our method results in a smoother baseline estimate which improves signal classification.

Our windowed `detrend_eBIC` method was used to removed the baseline drift from the total dataset consisting of 86,401 observations per node. The VI scores for the full dataset were 0.36, 0.24, and 0.41 for nodes c and d, c and e, and d and e, respectively. The complete confusion matrices for the nodes using a MAD multiple of 5 for the threshold is given in Table 1.

5 Conclusion and Discussion

We have expanded the quantile trend filtering method by implementing a non-crossing constraint and a new algorithm for processing large series, and proposing a modified criteria for

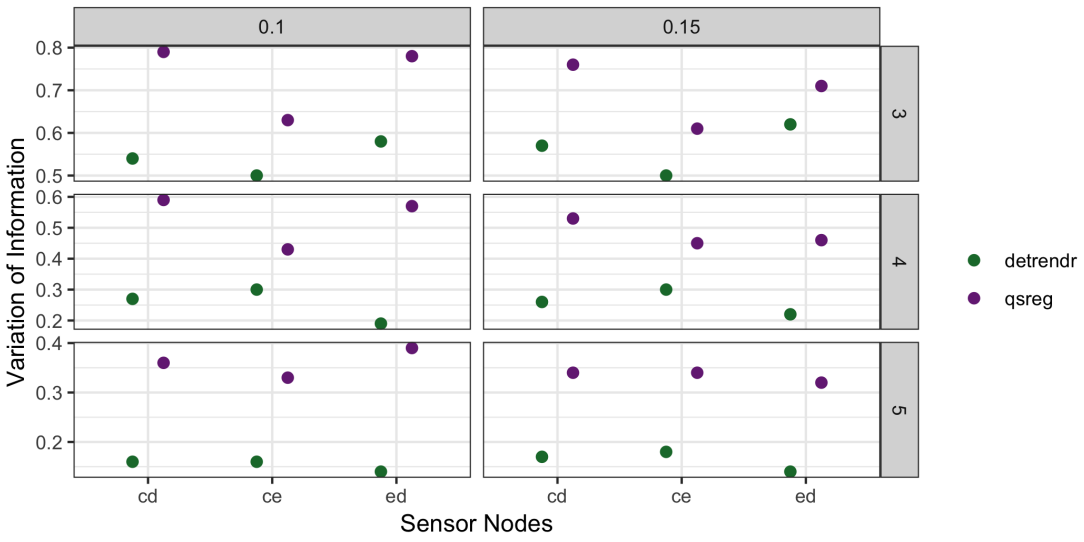

Figure 14: Variation of Information between sensor nodes after trend removal by quantile and method and thresholding by factor of MAD.

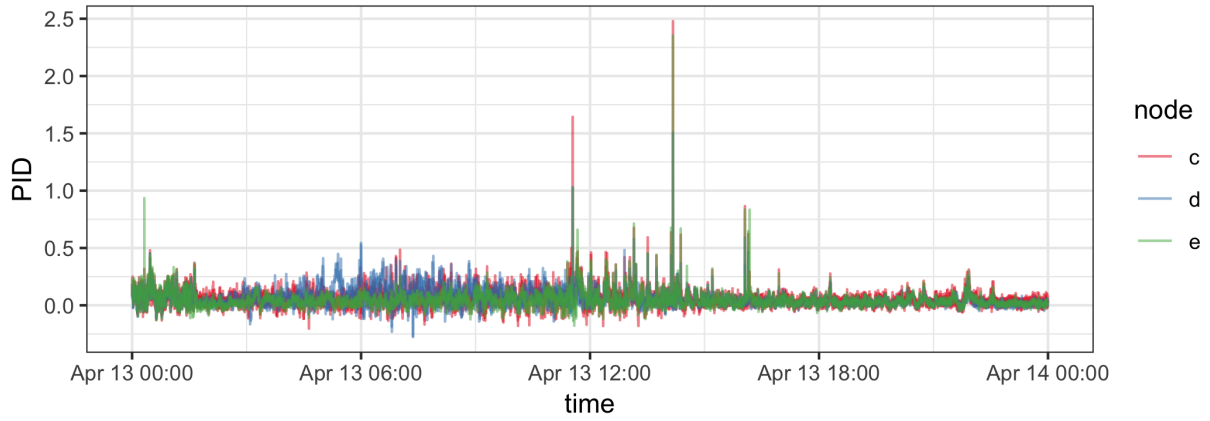

Figure 15: Low cost sensor data after drift removal using windowed detrend with eBIC.

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

	c = 0		**c = 1**	
	e = 0	e = 1	e = 0	e = 1
d = 0	80383	578	2551	145
d = 1	1298	737	111	598

smoothing parameter selection. Furthermore we have demonstrated the utility of quantile trend filtering in both simulations and applied settings. Our ADMM algorithm for large series both reduces the computing time and allows trends to be estimated on series that cannot be estimated simultaneously while our scaled extended BIC criteria was shown to provided better estimated of quantile trends in series with and without a signal component. We have also shown that the baseline drift in low cost air quality sensors can be removed through estimating quantile trends.

In the future, quantile trend filtering could be extended to observations measured at non-uniform spacing by incorporating the distance in covariate spacing into the differencing matrix. It could also be extended to estimate smooth spatial trends by a similar adjustment to the differencing matrix based on spatial distances between observations.

SUPPLEMENTARY MATERIAL

R-package for detrend routine: R-package detrendr containing code to perform the diagnostic methods described in the article. (GNU zipped tar file)

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.

Theussl, S. and Hornik, K. (2017), *Rglpk: R/GNU Linear Programming Kit Interface*, r package version 0.6-3.

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.

Weingessel, A. and Turlach, B. A. (2013), "quadprog: Functions to solve Quadratic Programming Problems." R package version 1.5-5.

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

Yu, K. and Moyeed, R. A. (2001), "Bayesian quantile regression," *Statistics & Probability Letters*, 54, 437–447.