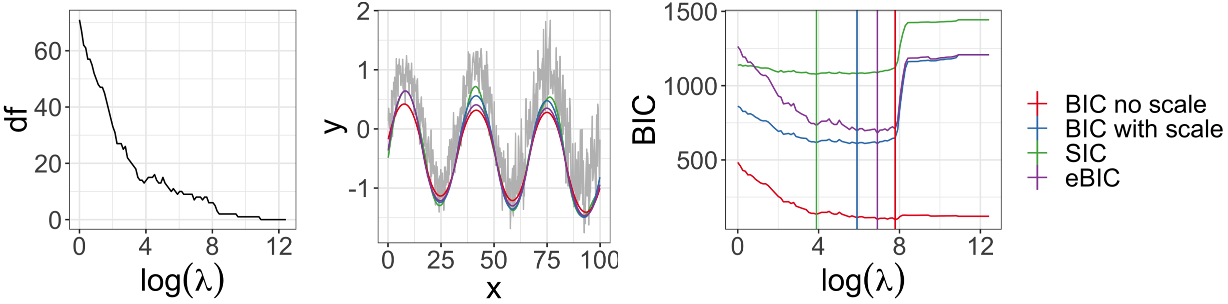
where P is the total number of possible parameters and ν is the number of non-zero parameters included in given model. We used this criteria with $\gamma = 1$, and $P = n - k - 1$. We use a single dataset from our simulation study to illustrate the difference between the scaled, unscaled ($\sigma = 1$)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

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.

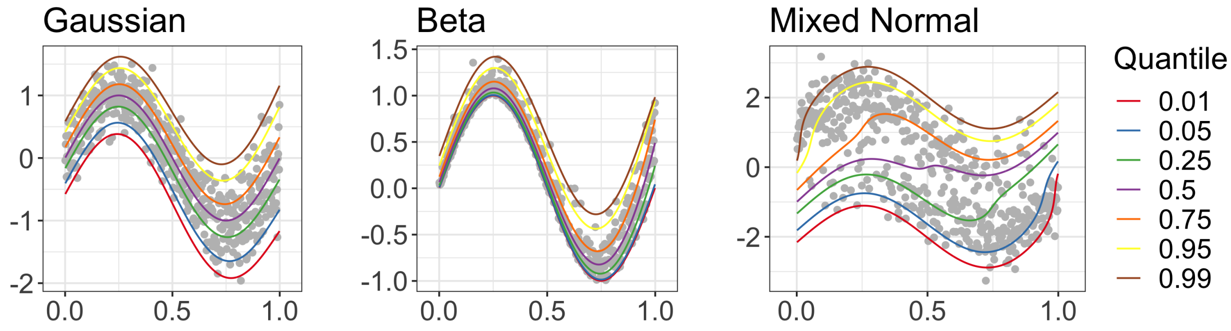


3 Simulation Studies

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).

maybe we should describe ours before theirs (i.e. reverse order of prev. 2 para)

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 ± 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 5th and 95_{th} quantiles. The `npqw` method performs particularly poorly in the mixed normal design due to the fact that it assumes the data come~~s~~ from a scale-location model which is violated in this case.

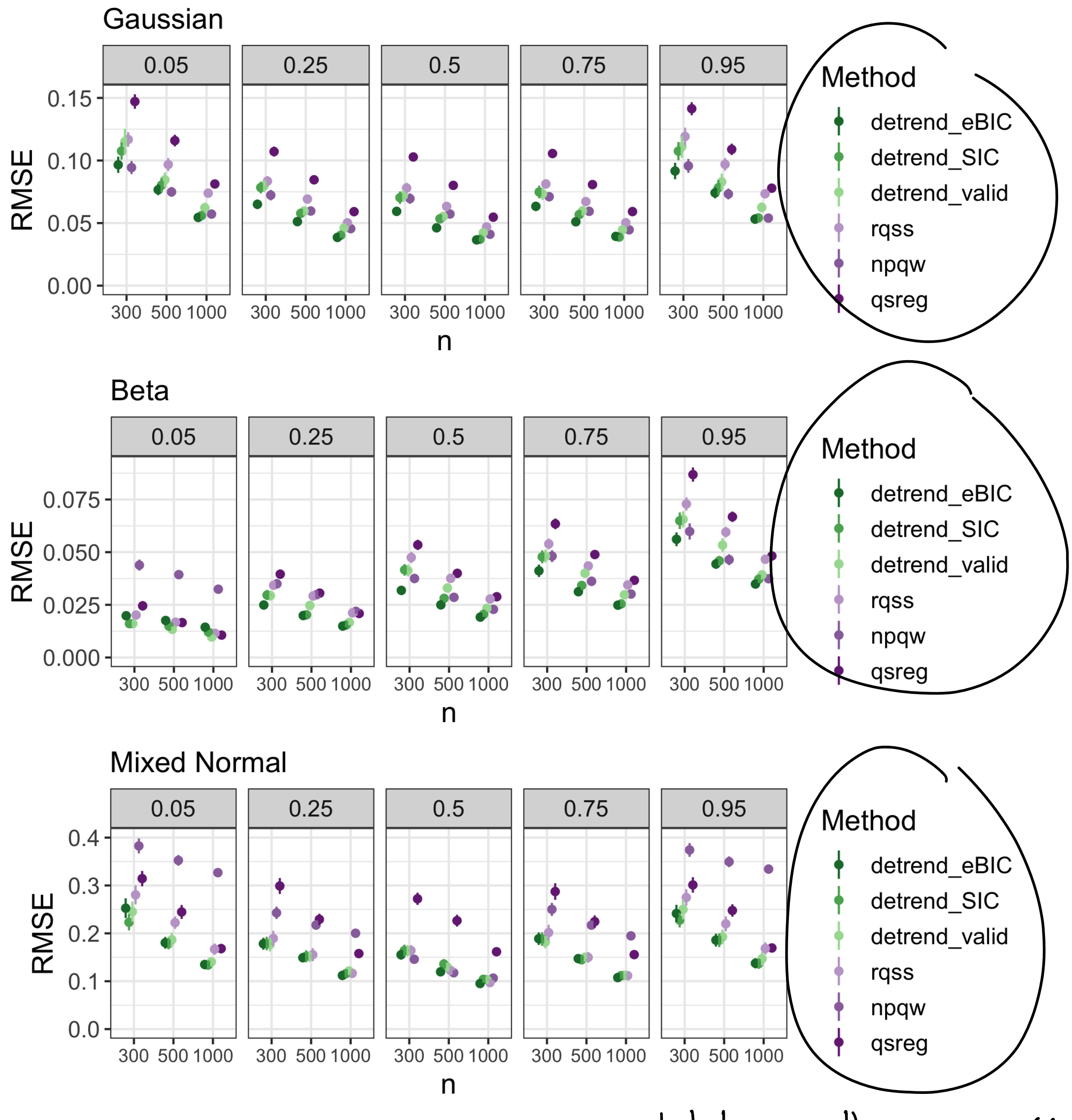
3.2 Peak Detection

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

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

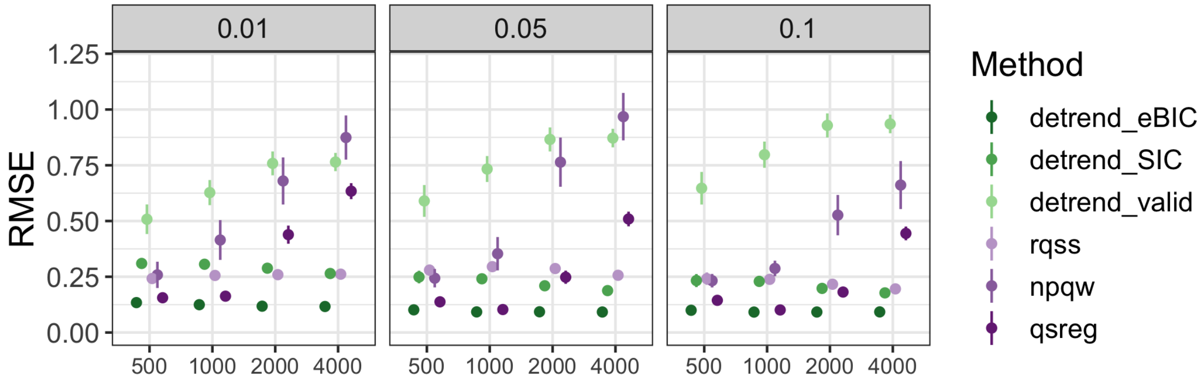
with $t = 1, ..., n$, where $s(t)$ is the true signal at time t, $b(t)$ is the drift component that varies smoothly over time and $\epsilon(t) \sim N(0, 0.25^2)$ is an error component. We generate $b(t)$ using cubic natural spline basis functions with degrees of freedom sampled from a Poisson distribution 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.



Since labels are the same, can you put them at bottom and widen the plots?

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.

is this true? Tighten up this statement.

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$^{\text{st}}$ or 5$^{\text{th}}$ 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.

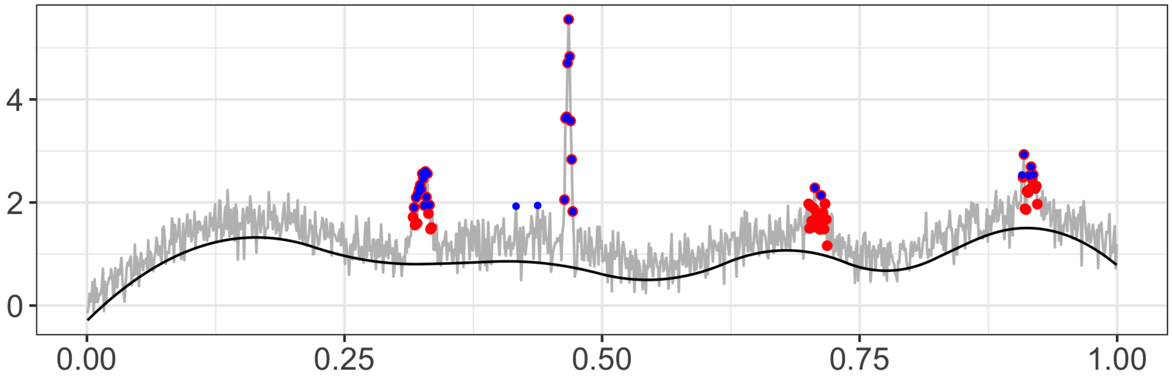
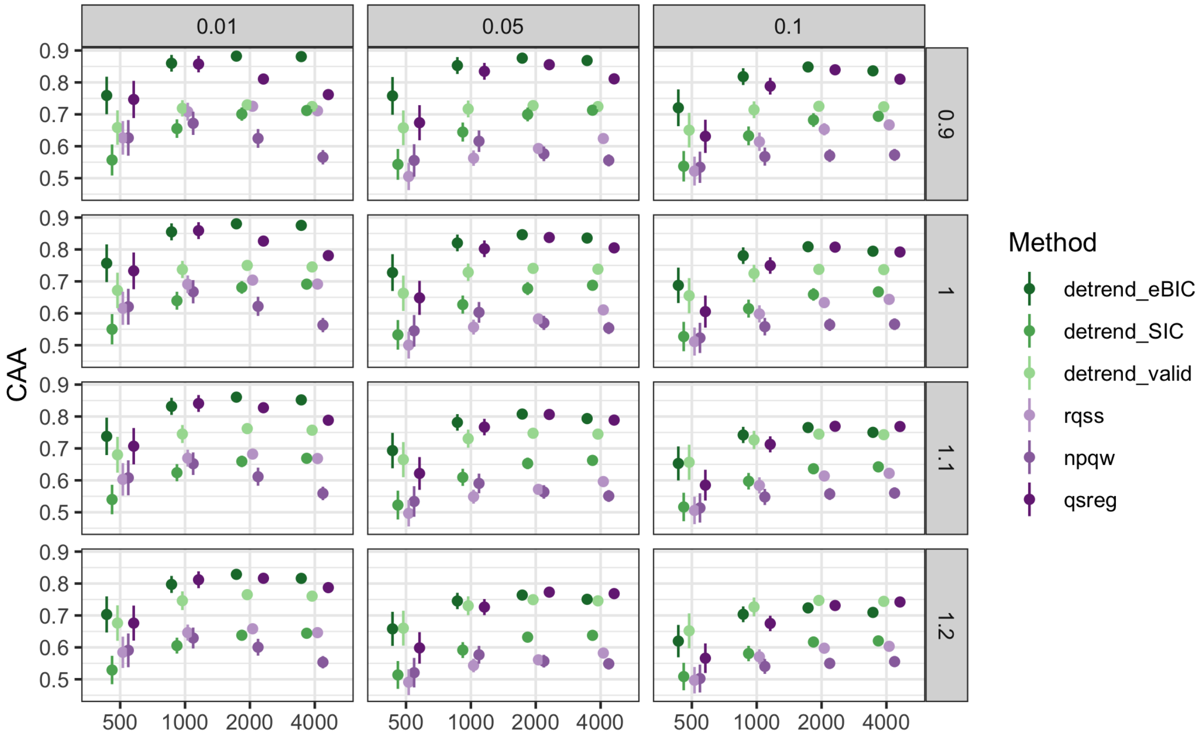


Figure 11: Class averaged accuracy by threshold, data size, and method (1 is best 0.5 is worst).

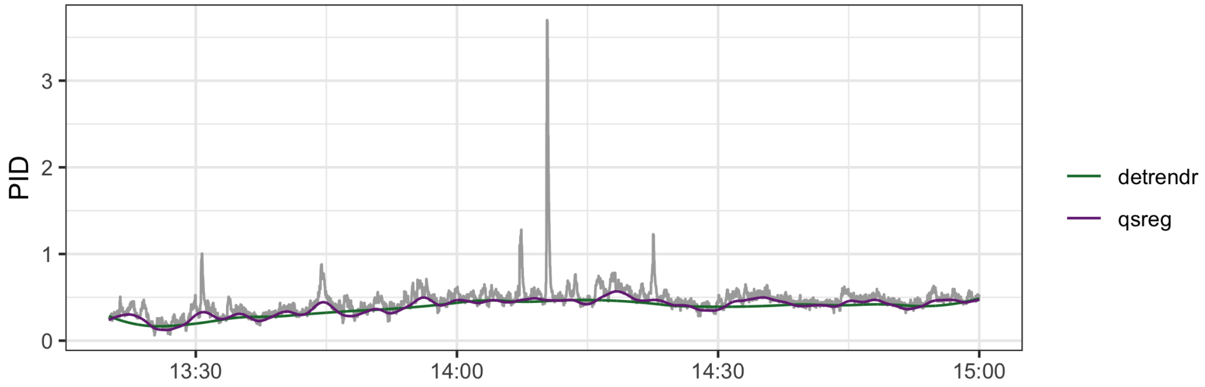


4 Application

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 10th and 15th quantiles and compare three thresholds for classifying signal. The thresholds are calculated using the median plus a multiple of

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



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

$$\text{MAD} = \text{median}(y - \tilde{y}), \qquad \text{median}\left(|y-\tilde{y}|\right)? \tag{18}$$

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 15th 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

inconsistent labels for the nodes.
Also "node" is not obviously clear.
Can we just call them sensors throughout?

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

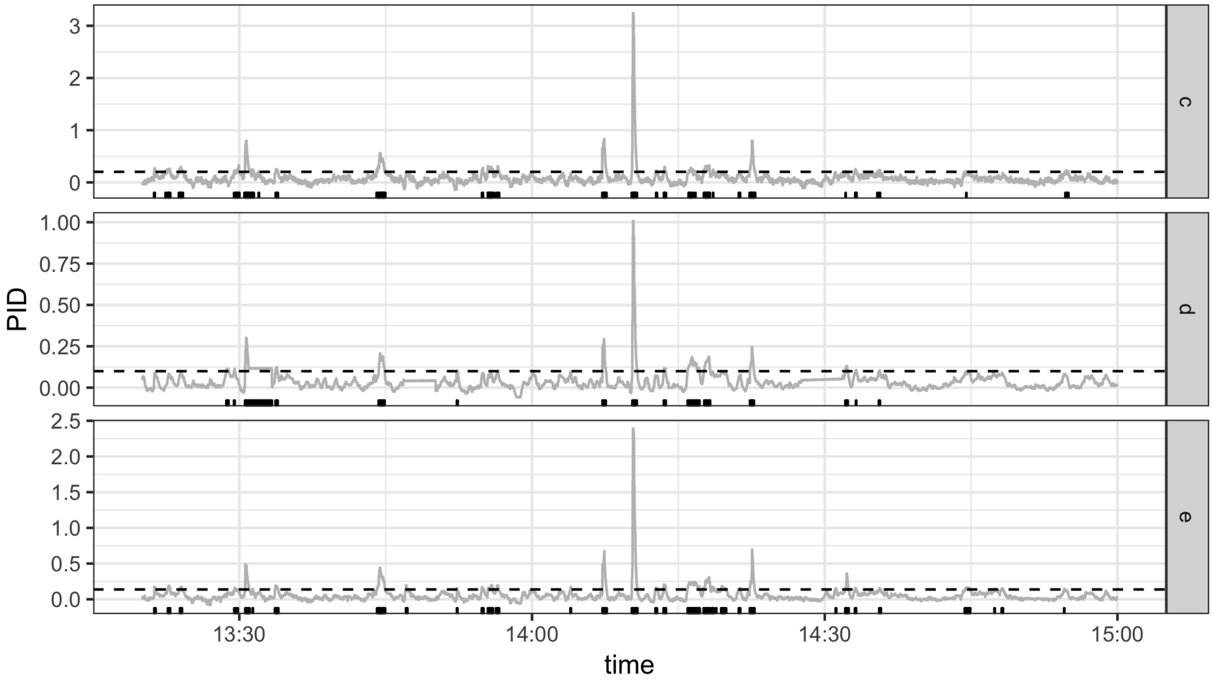
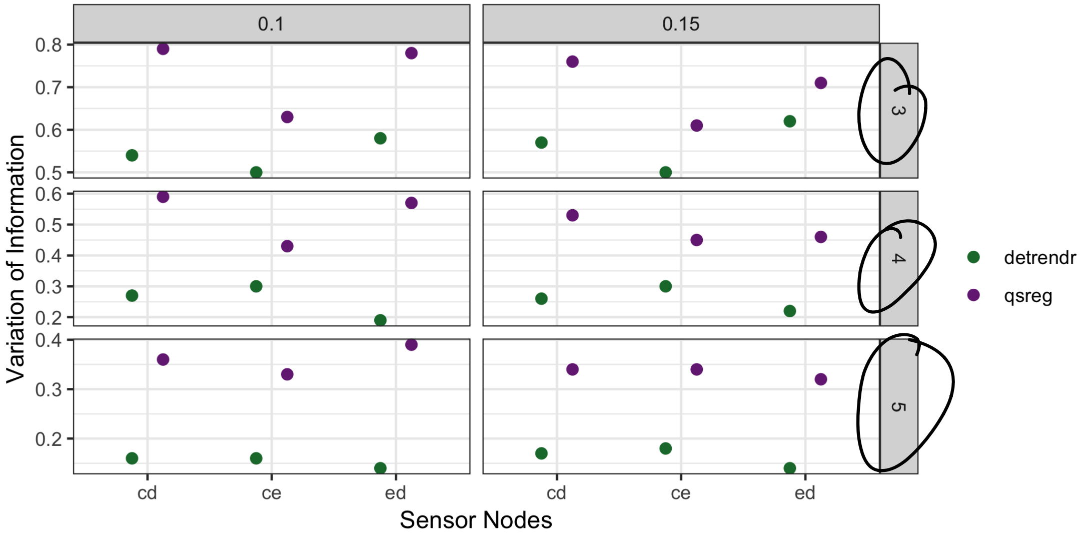


Figure 14: Variation of Information between sensor nodes after trend removal by quantile and method and thresholding by factor of MAD.



*

Table 1: Confusion matrices for 3 SPod nodes after baseline removal using 15th quantile 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

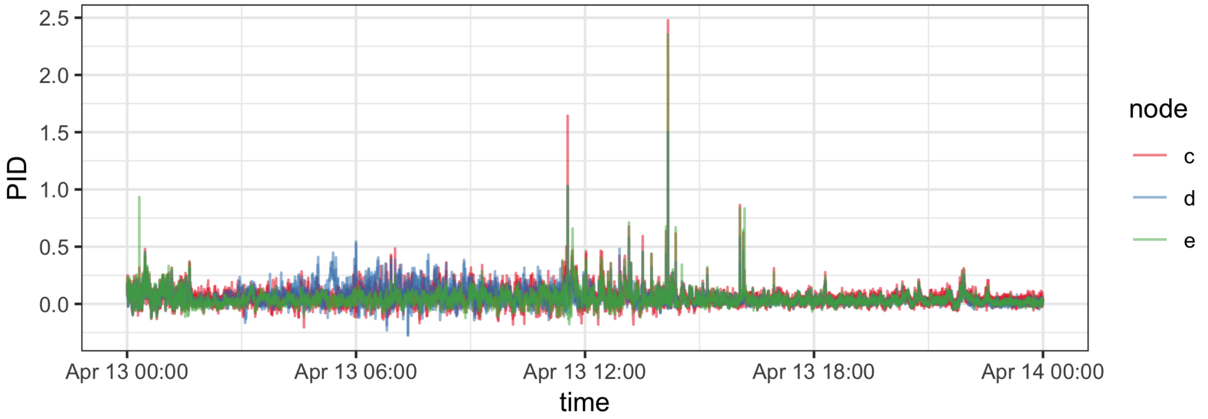
this tells me that d and e agree on signal more often than d+c or e+c. is that right?

can we use a,b,c or 1,2,3 for the node names? I imagine c,d,e will throw people off.

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.

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



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 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)