

Baseline Drift Estimation for Air Quality Data Using Quantile Trend Filtering

General Remarks

We thank the referees for their useful feedback and believe the manuscript is improved as a result.

In our revision we have denoted in **red** substantial newly added text. For the sake of readability we did not highlight minor edits in the revision. Finally, we give point-by-point responses to referee comments below.

Reviewer 1

1. The approach implies different tuning parameters: the penalty λ , the number of windows W , the bounds of the windows (l_w, u_w) and the degree of the polynomials (k) . The selection of λ is well described in Section 3. However, how the number of windows and the bounds of the windows are selected?

The referee raises a good question. Our motivation for casting the problem as a linked set of estimations over overlapping windows is computational. Ideally there would be just one window to estimate the trend over, rather than $W > 1$ but as the trends in the timing experiments in Figure 5 indicate, using a single window will become computationally impractical if one wishes to fit a model over a very large set of time points. Our numerical experiments show a little less than a linear speed up in the number of windows used. Thus, if one wished to cut the overall running time to fit a family quantile trends by a roughly factor of 10, one would choose in the ballpark of 11 or 12 windows. The bounds of the windows and amount of overlap between windows depends on the application and should be larger than the expected time duration of transient events. We have added additional sentences addressing these points in the revised manuscript at the end of Section 2.

2. In the paper, the degree of polynomials is $k = 2$. Can the information criteria introduced in Section 3 be used to estimate k ?

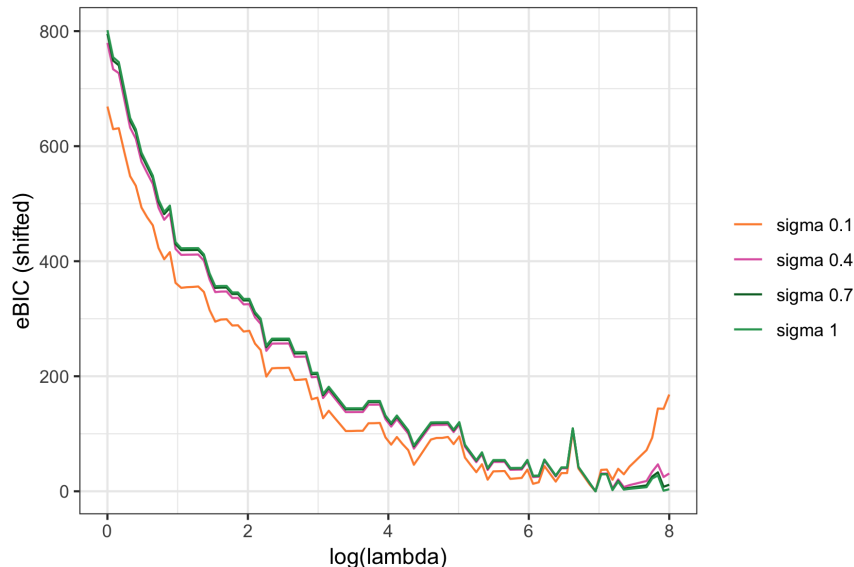
The expected degrees of freedom in the standard trend-filtering problem is given by the number of knots in θ plus $k + 1$ [Tibshirani, 2014]. Consequently, a natural strategy to estimate k is to use the eBIC criterion introduced in Section 3 but with ν equal to the number of knots in θ plus $k + 1$. We have added additional sentences addressing these points in the revised manuscript at the end of Section 3.

3. In Section 3.1, authors said that the method can easily handle missing data. What are the assumption on the process of randomness (MAR, MCAR,...)?

This is a good question. It turns out the penalized negative log-likelihood in Section 3.1 corresponds to making a MCAR assumption. We've added a short discussion of this in Section 3.1.

4. In Section 3.2, the BIC is introduced. I'm agree that for Bayesian quantile regression, the asymmetric Laplace likelihood is generally used. Does the choice of σ impact the results?

We did not find that the choice of σ affected the results. The following figure plots the eBIC criteria against σ for an example dataset. The value of λ chosen is the same for all values of σ compared, but the σ used in the analysis (0.1) results in a eBIC value that rises more rapidly after achieving the minimum.



5. Moreover, in a Bayesian framework, a penalized regression can often be defined with a specific prior distribution on the parameters (L1-type regression can be defined by using a Laplace distribution as prior). Is it possible to define such prior for the problem defined by Equation (10)?

Yes it is possible to define such a prior for the problem defined by Equation (10). Faulkner and Minin [2018] proposed a few fully Bayesian variations of the trend filter utilizing shrinkage priors on the k th-order differences in values of the unknown target function. They used a Gaussian likelihood for the data corresponding to the squared loss used by Kim et al. [2009]. A fully Bayesian variation of the quantile trend filter could be implemented by substituting the Laplace likelihood for the Gaussian likelihood. We have added these points in the revised discussion.

Minor comments:

6. At the end of Section 3.2, maybe more details should be given about the properties of the eBIC for the large parameter spaces (is it more efficient than the BIC)?

Thank you for the suggestion. The eBIC has been proven to be model selection consistent in cases where the BIC can be model selection inconsistent. This difference between the eBIC and BIC manifests itself practically in the trend filtering problem in that the BIC selected models tend to be undersmoothed in comparison to the eBIC selected models. We have added sentence highlighting these advantageous properties of the eBIC for large parameter spaces in Section 3.2.

7. Figure 7,8 and 10 are not really clear. May be they can be replaced by tables.

We have added tables with the results to the supplemental materials but feel that they are too long to include in the manuscript itself.

Reviewer 2

The paper address the issue of estimating multiple quantile trends simultaneously in time series data. This problem is common in several fields from economics to medicine. The authors focused on the application to low cost air quality sensor. Thus the paper fits with AOAS policy.

8. However, to improve the suitability, I suggest the authors to add more on how to employ their model and algorithm to data beyond to the application presented and which specific question this algorithm can address in different disciplines. While the problem to be addressed is clearly explained, the introduction miss a broader presentation of the topic to attract Statistician readers from differ fields. The authors claim that demixing problems is common to different applications with some references, thus they should comment more on this and reinforce the interest of the paper.

We have expanded our discussion of baseline trend filtering applications to include electrocardiogram data, electrooculographic data, mass spectrometry, chromatography, macroeconomics, environmental science, laser-induced breakdown spectroscopy, and astronomy. We have also included one or more references for each example that requires baseline removal. These additions have been made on page 2 of the revision. We have also added a short paragraph at the end of the Section 6 on how our proposed method could potentially be used more generally for robust smoothing of time series data.

9. Part of the technical material on the ADMM algorithm should be moved to appendices or online supplements.

We have moved part of the technical details of the ADMM algorithm to the supplementary materials.

10. Authors should check some typos spread in the manuscript.

Thank you. We have checked and corrected for typos.

References

- James R. Faulkner and Vladimir N. Minin. Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian Analysis*, 13(1):225, 2018.
- Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dmitry Gorinevsky. ℓ_1 trend filtering. *SIAM Review*, 51(2):339–360, 2009.
- Ryan J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323, 02 2014.