Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2021-429-RC1, 2022

Comment on amt-2021-429

Anonymous Referee #1

Referee comment on "Automated identification of local contamination in remote atmospheric composition time series" by Ivo Beck et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-429-RC1, 2022

This work describes an algorithm able to distinguish between the local component and the anthropogenic contamination of environmental related datasets. The algorithm is shared not only as a flow diagram but also as a functional code, which makes it even more important. Data cleaning is a cumbersome procedure that anyone working with environmental monitoring has faced. Therefore, this work can contribute towards the automation of these labour intensive procedures.

Most articles related to data cleaning deal with a specific dataset only, and the reader is always left wondering what are the limitations of the proposed algorithm and whether it would be worthwhile to apply it to other datasets. In this work, the same algorithm is applied to several different instruments to prove its wide applicability. I must note though that practically two different methods are presented in Section 2.4.1 that have been incorporated under one software.

The only weakness of the proposed method is that the user should decide on up to 7 parameters to make the code operate optimally, even though it is discussed in the manuscript that 3 are necessary and the remaining optional. This makes the proposed algorithm quite subjective. Can the authors comment on that?

The manuscript is very descriptive and answers most questions that may arise. However, there are a few clarifications required, mostly related to the applicability of the algorithm (named PDA in the manuscript).

In specific

The manuscript would benefit if the underlying assumptions of applying the PDA are clearer. PDA is used only in datasets obtained in pristine conditions, where the concentration difference between anthropogenic and local components can be an order of magnitude. What is the smallest difference that can be detected? Of course this relates to the parameters selected by the user. This point should be discussed further.

How does the PDA respond to data gaps? Is there any restriction if gradient filter method A is applied?

How does PDA respond to the edges of the dataset? Please discuss.

How big dt in Eq 1 should be and how that is determined? How is dt related to the time resolution of the dataset and the expected duration of the anthropogenic events.

In the IQR method how is the duration of the moving window related to the dataset with respect to time resolution and expected duration of the anthropogenic events.

In the IQR method a moving window is mentioned and hence a data point can be evaluated multiple times. It is not clear when it is flagged though. If it exceeds the IQR threshold once or multiple times. If it is the latter case how many exceedances should occur? Please clarify.

To further investigate the limitations of the PDA I am taking advantage that this is an open discussion and share two datasets to discuss how the algorithm behaves towards them. These are two case studies not discussed in the manuscript.

Dataset 1: It is frequent, eg due to A/C influence, that the standard deviation of the measurements changes abruptly even though the mean remains the same. How this case should be treated?

Dataset 2: It is assumed that any contamination would add to the local component. How does the algorithm treat data below the local component that are scarcely met but still exist?

There are of course limitations not related to the algorithm but to the processes themselves. A major assumption is that the time resolution of the dataset should be higher than the duration of the anthropogenic influence.

A flow chart with the with respect to the two data gradient methods should be added, to make clearer the algorithm process. Please use the standard schematics.

As discussed in the manuscript, there are difference of the PDA and the manual method, which relate to false positive (measurements not identified as polluted even though they are) and false negative (non polluted measurements identified as polluted).

Please include in either Table 1 or 2 how many false negatives and positive, compared to the visual method, the PDA leaves behind in each step.

How are the false negatives and positive distributed? Is there a pattern or are they random?

A discussion on how varying each of the 7 parameters affects the amount of false negatives in a case study presented in the manuscript would be beneficial.

A method quite similar to this work has already been published (Gallo et al., 2020). Also additional methods have been applied, such as smoothing, to mask short term local events (Liu et al., 2018). This is a subject the community has spent some time to investigate and there is some literature out there. Most notably Giostra et al., 2011; McNabola et al., 2011; Brantley et al., 2014.

Brantley, H. L., Hagler, G. S. W., Kimbrough, E. S., Williams, R. W., Mukerjee, S., and Neas, L. M.: Mobile air monitoring dataprocessing strategies and effects on spatial air pollution trends, Atmos. Meas. Tech., 7, 2169–2183,

https://doi.org/10.5194/amt-7-2169-2014, 2014.

Gallo, F., Uin, J., Springston, S., Wang, J., Zheng, G., Kuang, C., Wood, R., Azevedo, E. B., McComiskey, A., Mei, F., Theisen, A., Kyrouac, J., and Aiken, A. C.: Identifying a regional aerosol baseline in the eastern North Atlantic using collocated measurements and a mathematical algorithm to mask high-submicron-number-concentration aerosol events, Atmos. Chem. Phys., 20, 7553–7573, https://doi.org/10.5194/acp-20-7553-2020, 2020.

Giostra, U., Furlani, F., Arduini, J., Cava, D., Manning, A. J., O'Doherty, S. J., Reimann, S., and Maione, M.: The determination of a "regional" atmospheric background mixing ratio for anthropogenic greenhouse gases: A comparison of two independent methods, Atmos. Environ., 45, 7396–7405, https://doi.org/10.1016/j.atmosenv.2011.06.076, 2011

Liu, J., Dedrick, J., Russell, L. M., Senum, G. I., Uin, J., Kuang, C., Springston, S. R., Leaitch, W. R., Aiken, A. C., and Lubin, D.: High summertime aerosol organic functional group concentrations from marine and seabird sources at Ross Island, Antarctica, during AWARE, Atmos. Chem. Phys., 18, 8571– 8587, https://doi.org/10.5194/acp-18-8571-2018, 2018

McNabola, A., McCreddin, A., Gill, L. W., and Broderick, B. M.: Analysis of the relationship between urban background air pollution concentrations and the personal exposure of office workers in Dublin, Ireland, using baseline separation techniques, Atmos. Pollut. Res., 2, 80–88, https://doi.org/10.5094/APR.2011.010, 2011

Please also note the supplement to this comment:
https://amt.copernicus.org/preprints/amt-2021-429/amt-2021-429-RC1-supplement.zip