

Response to Referee # 2

April 23, 2016

This paper presents a method to correct for variability in PTFE filter baseline IR absorbance brought about by non-uniform filters and by the stretching incurred during long sampling periods. This tool will help researchers access data from networks, such as IMPROVE, in order to increase our understanding of organic aerosol transformation. The paper is generally well-written, with only minor grammatical or typographical errors. There are some sections that could use more descriptive, perhaps less mathematical, text in order to reach a wider audience. I recommend this paper for publication with very minor revisions.

The authors thank the reviewer for the constructive comments and suggestions that help improve the quality of the paper. We have revised the manuscript to address the reviewers comments and made response to each comment in this file. The responses are in red.

1. Beginning in section 2.1, the authors introduce a large number of variables represented by symbols. Given the large number, a list should be provided to help the reader follow along or be able to easily look up any variable in one place.

As also suggested by Reviewer 1, we agree that providing an explicit glossary upfront will help with keeping track of variables and their variants introduced throughout Section 2. In this light, we start the Methods section with a summary of notation for most commonly used variables, together with their description. We also break down variables into 3 distinct categories with respect to their roles they play in our smoothing splines model implementation. We include pointers to the sections where the variables are formalized to allow for quick and selective reading.

2. Section 3.1 is hard to read. Granted, this kind of mathematical treatment is new to me, but the authors might consider using more descriptive text, in addition to the variables, to help a wider range of researchers access this tool. This applies broadly to the paper to some extent, but especially to this section.

We agree in some parts of Section 3.1 would benefit from a more qualitative discussion. We have rewritten the section by describing the general result and interpretation from figures. For example, the revised paragraph 1 in Section 3.1 now reads:

“Qualitatively, in Figure 3 we compare the behavior of PM (left panel) and blank sample spectra (right panel) for various values of EDF_T (4, 5, 7, and 200: from top to bottom). In this analysis, we used all 54 blank samples and randomly sampled 54 out of 794 PM samples to keep the counts equal and allow for representative cross-comparison. The trend from top to bottom shows both PM and blank samples exhibit increasing sensitivity to the amount of smoothing applied. With increasing EDF_T , baseline corrected ambient spectra begin to exhibit negative analyte absorbance (left column). Simultaneously, baseline corrected blanks in the region at 3700 – 2500 and 1820 – 1600 cm^{-1} begin to depart from our target, zero absorbance (right column).”

We also included descriptions of Figure 4 layout to help readers orientate and provide more effective transitions between paragraphs in Section 3.1 and Figure 4. For example,

“Quantitatively, in Figure 4 we evaluate the impact of EDF_T on negative absorbance fraction metric, NAF , (top panel) and total normalized absolute blank absorbance metric, $\|\vec{a}_B\|_1^*$, (bottom panel) in segments 1 and 2 (left and right panel). Horizontal panels share the same x-axis and

vertical panels share the same y-axis to allow for representative cross-comparison. Therefore, each plot in the matrix in Figure 4 corresponds to a unique condition in terms of a metric and segment.”

Section 3.1 includes additional pointers and structure modifications interspersed throughout the text. We hope these modifications will lend the paper greater clarity and be understood by a variety of researchers.

3. I wondered when reading the introduction and conclusion exactly how much time, or computational time, was saved by doing the baseline fitting with spline fits rather than polynomial. It also was not entirely clear how the spline technique alleviated the necessary “expert” interpretation relative to the polynomial fit. If the authors could elaborate on these two points, it would strengthen the paper.

This is a significant question. Computationally, the polynomial and spline fitting method are linear operations and are therefore comparable. The reduction in user time is difficult to generalize due to high variability across different users. Qualitatively, we can reason that if N values are considered for each free parameter, the time for expert examination of each solution scales with N^4 for the polynomial method (for explicitly defining the background regions), and N for the spline method (for defining the EDF). Additionally, the quantitative criteria (NAF and $\|\vec{a}_B\|_1^*$) that we introduce in this manuscript have been shown to sufficiently simplify the parameter selection process for users of any level of experience. We have included the following statement in the Conclusion section:

“Although the exact reduction in user time may be difficult to generalize due to high variability across different users, we reason the following approximation applies. Qualitatively, if N values are considered for each free parameter in each method, then the amount of time for expert examination of each model solution scales up with N^4 for the polynomial method (due to 4 boundary points as free parameters) and N for the smoothing splines method (due to 1 EDF parameter). Also importantly, the evaluation metrics, which we established in this manuscript, have been shown to sufficiently simplify the parameter selection process for users of any level of experience.”

4. Abstract line 6: I suggest moving the word “remains” directly after “the question”

We moved “remains” after “of” since the preposition expresses the relationship between a part and a whole. The sentence now reads:

“Therefore, the question of how to develop an automated method for baseline correcting hundreds to thousands of ambient aerosol spectra remains given the variability in both environmental mixture composition and PTFE baselines.”

5. Pg 2 line 8: “relatively” Pg 2 line 26: “A growing number of papers have been published in recent years” Pg 2 line 27: “One of the applications includes” Pg 4 line 8: “a priori. Therefore”, Pg 14 line 13: omit comma

These have been fixed.

6. Figure 3: x-axis tick mark labels on left and right panels overlap. Figure 7 may be unnecessary (especially given the large number of high-information figures that are certain to be large).

X-axis tick marks in Figure 3 have been adjusted and caption in Figure 3 has been modified to include:

“X-axis range from 4000 to 1500 cm^{-1} in both left and right panels.”

Figure 7 has been moved to supplemental information section.