Ye Yuan et al., Adaptive Baseline Finder, a statistical data selection strategy to identify atmospheric CO_2 baseline levels and its application to European elevated mountain stations

Answers to J. Kim (Referee)

The referee's comments are in black, answers are in blue.

Short notice:

According to the suggestion from J. Kim (Referee #2), we changed the name of our method "Adaptive Baseline Finder (ABF)" into "Adaptive Diurnal Minimum Variation (ADMV)". All the names and abbreviations of this method have been adjusted throughout the answer.

This work presents a new statistical algorithm, named ABF, for identifying "baseline" levels from CO2 measurements. The title of the work refers to elevated mountain sites as its application focus, but the work also includes some analysis of non-mountain sites as well. While there are some issues that I would like to see the authors address, overall I do feel the authors have done a good job of presenting a unique algorithm and comparing it to other frequently used methods in the measurement community, and as such I suggest that the manuscript be published with some revisions.

Before I proceed with my comments on the paper, I would like to comment on the term "baseline" itself. My concern is that the definition of "baseline" is very subjective open to interpretation. For example the authors mention that ABF in this study was used specifically to identify periods of free troposphere concentrations in the high elevation sites, and that is certainly one valid definition of "baseline". With this definition, however, sites that may have statistically stable concentrations at certain times of the day but do not necessarily measure the free troposphere will by definition have no "baseline". If the definition of "baseline" was "typical concentrations you would probably measure at a certain location at a certain time" with the goal of creating a global spatial map of average concentrations, I suppose you would end up with something close to the trend and seasonal components in the STL analysis, which you may (or may not) be able to find through statistical methods such as ABF. On the other end of the spectrum, for a regional modeler, the useful definition of "baseline" would be whatever concentrations enter the modeling domain and not necessarily any clean/stable condition, and if the air was polluted coming into the grid box then the model needs to know about it. I've seen attempts to distinguish between "baseline" and "background" to try to navigate through the subtle (and sometimes not-so-subtle) differences in definitions, but in my view all attempts at defining "baseline" is inherently subjective and the best practice is to be specific about what the particular definition for the study is, and that definition should encompass the specific intended use of this definition. All this to say, I feel the name Adaptive BASELINE Finder, while sounding nice, can be misleading. I would suggest that the authors consider another name, but will leave the decision to the authors.

Thank you very much for your concerns and elaboration. We fully agree with the elaborate remark of the reviewer. We would like to recall that our study largely focuses on elevated stations, but we define the term "baseline" in the beginning of the introduction (see Page 2 line 8-10) in a broader sense referring to "well-mixed air masses with minimized short-term external influences." We also agree that the name "Baseline Finder" can be misleading, thus after discussion with all co-authors, we decided to change the name into Adaptive Diurnal Minimum Variation (abbreviated: ADMV). Besides, we also added more arguments mainly in the Results and

discussion, as well as the Conclusion section.

[General comments, questions]

- P5, ln 15: Why the window of 6 hours? I suppose this assumes that baseline conditions occur for longer than 6 hours? Have you tried shorter windows and found you come to the same conclusion? I almost wonder whether it would be more beneficial as a general algorithm to have as short a window as possible, such that the window never exceeds the actual window of a baseline occurrence?

The window of 6 hours and the standard deviation threshold were choices based on empirical visual inspection of the available datasets and on literature review. For this study, we specify that the 6 hours in the *start time window* have to meet two constraints: the standard deviation of measured values less than 0.3 ppm and the missing data in the 6 hours less than 50%. If the requirements are fulfilled, then the data selection will start with the *start time window* for that day. Otherwise, all values in that day with the *start time window* are labelled as "unselected."

The length of 6 hours was considered as a reasonable time length to determine whether the measured air masses are well-mixed and thus most representative, largely following the approach of Pales and Keeling (1965) (as mentioned in the introduction, page 3 line 15). More detailed information can also be found in Levin et al. (1995) and Brailsford et al. (2012). Shorter time windows will lead to less robust statistics and thus more variable standard deviations. Thus the selection procedure might rely on less representative data and risk of accepting the wrong *start time window* increases.

However it is worth noting that 6 hours were only chosen for this study. It can be variably adjusted by users according to their measurement sites.

- P10, ln 15: The increase in the mean annual growth rates is within the noise, I'm not sure that much can be made of this.

We agree that, the tendency is not statistically significant based on the confidence intervals.

We rephrased as, "Moreover, the following fact is observed for all sites except for *SSL*."

- Figure 2: I have a hard time understanding this figure. First off, the figure seems to represent data from the full data set (spanning years), and yet the method describes that the baseline "window" is adaptive, potentially changing each day and by season. What criteria was used to derive a representative window for the whole period?

We apologize for the misleading wording. There are different terms regarding time window in our ADMV method: *start time window* and *selected time window*.

The *start time window* is different based on different running frequency of ADMV. It is the result from the first part of ADMV data selection – *starting selection* (see Section 2.2.1). For this study we applied overall frequency, indicating the *start time window* for the full data set (spanning years) is the same. Figure 2 shows the *start time windows* at each measurement site. Theoretically we can also apply ADMV by yearly, seasonally or monthly depending on the requirements.

And for calculation, the *start time window* is derived from the diurnal cycles which are the mean over the detrended data. The criterion for selecting such window is the least variable time period (6 hours) during the night time (6 p. m. to 5 a. m. LT), due to the focus on mountain

stations for this study. More details can be found in Section 2.2.1 (Page 5 line 6 onwards).

On the other hand, the *selected time window* represents the selected data sets from ADMV data selection. It is the result from the second part of ADMV data selection – *adaptive selection* (see Section 2.2.2). After both *forward* and *backward adaptive selection*, the *selected time window* result is different for each individual day.

- P10, ln 27: Regarding "active vegetation", wouldn't signals from respiration also explain these results, and wouldn't that also be one form of active vegetation? I think this possibility can't be ignored since the authors suggest that the lower VAL values in summer are likely due to vegetation. Are the anthropogenic emission activities in this region such that you would expect emissions only in winter, or are they small enough to be masked by the summer drawdown? I do think that the authors' interpretations on the findings are likely to be correct, however I do think that a much deeper analysis of the data (perhaps beyond the scope of this paper) may be needed to conclusively determine the source of these discrepancies.

Based on our results, it is very likely that the lower free troposphere will respond in a delayed manner to CO_2 concentration changes by effective sources and sinks on the ground. The lower free troposphere acts like an atmospheric "memory" with delayed reaction.

Regarding anthropogenic emissions in summer, we agree that they are certainly small enough to be masked in the drawdown. One example can be found in Oney et al. (2017) for a comparison of biospheric and anthropogenic contribution from CO_2 observations at a tall tower station on the Swiss Plateau, which is the most populated and most industrialized region in Switzerland. Both Fig. 2 and Fig 3 in Oney et al. (2017) show the difference in anthropogenic and biospheric signals, especially in summer time. The magnitudes of anthropogenic contributions are much smaller than the biospheric ones.

- One discussion I think is missing is regarding the "adaptiveness" of the algorithm, in other words do the results show baseline windows changing with season. The authors state this as a strength of the ABF (P4 ln 29), so I had expected this to be one of the early points of discussion.

The adaptiveness of the algorithm is indeed the ability to select values in different time windows for every individual day. It is the ability to adapt the selection on a daily basis in order to receive a maximum amount of representative data (For more details please see Section 2.2.1 and 2.2.2). One of the results is shown in Section 3.1 for the different *start time* windows among stations.

Moreover, the ADMV data selection can also be processed for each season individually (with individual settings manually). A comparison of the resulting *start time windows* between overall and seasonal running frequencies can be found in Supplement S1.1.

[Minor comments]

- Page 4, ln 10: "At last", change to "Finally"?

This was corrected.

- P4, ln 27: "No upwind air masses with depleted CO2 levels by photosynthesis of vegetation like in summer are recorded." -> "Unlike summer, no upwind air masses with depleted CO2 levels by

photosynthesis of vegetation are recorded."

This was corrected.

- P5, ln 12: "but preserves of the diurnal pattern." -> "while preserving the diurnal pattern."

This was corrected.

- P6, ln 10: "Step 3" is not actually a step, but a general description of Step 5 and 6. Perhaps it makes more logical sense to include it in "Step 2", presenting it as an "If/Else" step.

We combined Step 3 into Step 2 with the "If/Else" step, and changed the following step numbers accordingly.

- P9, ln 5, Table 2: Can the authors clarify whether the percentages are based on just the time windows considered in the algorithm or the complete dataset?

The percentages refer to the ratio of the selected data values in the data values of the complete data sets. The selection percentage regarding the selected time windows and the selected days can be found in Supplement S3.1 in detail.

Thus for clarification, we rephrased "data in all data for all stations" as, "data values in all values of the complete data sets".

Reference

Brailsford, G. W.; Stephens, B. B.; Gomez, A. J.; Riedel, K.; Mikaloff Fletcher, S. E.; Nichol, S. E.; Manning, M. R. (2012): Long-term continuous atmospheric CO_2 measurements at Baring Head, New Zealand. In Atmospheric Measurement Techniques 5 (12), pp. 3109–3117. DOI: 10.5194/amt-5-3109-2012.

Levin, I.; Graul, R.; Trivett, N. B. A. (1995): Long-term observations of atmospheric CO_2 and carbon isotopes at continental sites in Germany. In Tellus 47B, pp. 23–34.

Oney, B., Gruber, N., Henne, S., Leuenberger, M., & Brunner, D. (2017). A CO-based method to determine the regional biospheric signal in atmospheric CO_2. Tellus B: Chemical and Physical Meteorology, 69(1), 1353388.

Pales, J. C. and Keeling, C. D.: The Concentration of Atmospheric Carbon Dioxide in Hawaii, J. Geophys. Res., 70, 6053–6076, doi:10.1029/JZ070i024p06053, 1965.