Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Interactive comment on "Adaptive Baseline Finder, a statistical data selection strategy to identify atmospheric CO$_2$ baseline levels and its application to European elevated mountain stations" *by* Ye Yuan et al.

Anonymous Referee #1

Received and published: 7 October 2017

Yuan et al. present a data selection method for records of atmospheric CO2 mole fraction observations from mountain locations. Their method, the adaptive baseline finder (ABF), is an interesting one and in that sense worth publishing. However, unfortunately the manuscript in its current form remains very descriptive and does not include clear conclusions on how the community would benefit from using this method in comparison to the other methods to which ABF is compared.

Another main point is that the English should be checked by a language editor, as in several places the manuscript is not written in correct English (e.g. articles are often

omitted and commas are used incorrectly).

All in all, I think the authors have done a substantial amount of interesting work, and could be worth publishing after taking into account the specific comments below and especially focus on placing their work in larger context and making more explicit what the use of ABF could contribute to the field.

Specific comments:

Page 1 line 21: 'measuring sites' should be replaced by 'measurement sites', throughout the manuscript.

Page 1 line 22: Why would this lead to a bias when comparing different stations? Only when the data of these different stations has been selected with different methods.

Page 1 line 23: pattern -> patterns

Page 1 line 24: 'measuring records' -> records of atmospheric CO2 observations

Page 1 line 27: implemented -> included/applied

Page 1 line 27: Among the studied methods, our ABF method . . .

Page 1 line 27: This is very descriptive: lower percentage of selected data; is this 'better'? What does it imply to have less or more data selected?

Page 1 line 30: STL is not explained

Page 2 line 13: what do you mean by correction for interference from other GHG?

Page 2 line 24: here it would be good to elaborate on the work of Uglietti et al. 2011 (ACP), which is referred to on the same page.

Page 2 line 29: explain why afternoon values should be excluded.

Page 3 line 3: MHD flasks are only sampled during restricted base line conditions, so no filtering is applied.

Page 3 line 6: Hawaii, USA. Also add Switzerland for JFJ.

Page 3 line 18: what is REBS?

Page 3 line 23: why do the authors choose to focus on mountain sites only? This should be made more clear in the manuscript.

Page 4 line 7-9: what do these classifications mean? E.g. "weakly influenced, constant deposition" is not very clear.

Page 4 line 13: did you use hourly data or higher time resolution? This is not clear from this section.

Page 4 line 15-16: specify which reference is for which station.

Page 4 line 19-20: This sentence is very vague, make more clear what the motivation of this research is.

Page 4 line 21-22: This sentence is not clear: what traffic activities are relevant to the mountain sites? And why is vegetation active in the afternoon only? How about respiration?

Page 4 line 22-24: this sentence is not clear. What do you mean by 'which in turn in an effective tool'? What tool?

Page 4 line 25: explain PBL and explain the changing degree of entrainment.

Page 4 line 27-31: The level of English needs to be assessed particularly in these sentences.

Page 5 line 7: What is the time resolution of the data sets?

Page 5 line 11-20, and page 6 line 6-22: Revise English especially here, including use of complete sentences including articles ('the') and correct use of commas.

Page 6 line 26: This is a vague sentence, data only exists on a single day, so why talking about selecting it in 'any day'?

C3

Page 7 line 15: photosynthesis starts long before 11 a.m.

Page 8 line 10: Why hourly? How did you define hourly values? As the average of the whole hour? Or just last part? Is the hour defined at the beginning of the averaging interval or at the end? This is important information and should be included in methods.

Page 8 line 15: Does it make sense to have different windows at the different levels?

Page 8 line 19: The results ARE not fully comparable. Does it even make sense to analyze such a short record which does not even give a complete annual cycle?

Page 9 line 2: It would make sense to look at the differences by season, as the diurnal cycle is not the same throughout the year. Also, the data sets all cover different time periods, so it is difficult to compare.

Page 9 line 4-10: Revise English.

Page 9 line 6-9: But what do these percentages actually mean? This is too descriptive and needs more analysis and perspective.

Page 9 line 10: what is 'major step' and what do the percentages by each step mean?

Page 10 line 3: This previous section remains very descriptive. What do the differences between all methods mean, and what is more useful for what type of analysis? This needs more work.

Page 10 line 10: What is the use of comparing growth rates for different time periods? Growth rates are very variable from year to year, so choosing a different period gives different growth rates.

Page 10 line 11: A positive trend in what? In the CO_2 concentrations in general?

Page 10 line 12: Explain VAL

Page 10 line 12: what differences?

Page 10 line 15: What do you mean by tendency?

C4

Page 10 line 18: 2015 had a much higher growth rate compared to the years before, so that also influences the results at SSL. Why not including 2015? It is publically available through ObsPack.

Page 10 section 3.3.1: I do not understand the added value of this paragraph. It should include more details on what was exactly studied and more conclusive remarks instead of only descriptive statements.

Page 10 line 20: this is not clearly described (the difference between val and selected data).

Page 10 line 24: this percentage is given too much precision.

Page 10 line 27: if VAL is all validated data it can never over- or underestimate CO_2 levels, as they are the actual observations!

Page 11 line 1-7: Very descriptive, add more details and analyses and perspective.

Page 11 line 6: explain in more detail 'thickness of the mixing layer'.

Page 11 line 14: what does 'least standard deviations' mean?

Page 11 line 14: we already knew that IZO is least influenced.

Page 11 line 15: what are intermediate results?

Page 11 line 19: could, but why not done?

Page 11 line 25, figure 6: why is red included in the color scale as those values do not occur? Also the caption of figure 6 contains a lot of information that should be included in the main text as well (pearson corr. matrix etc).

Page 12 line 1: what does this mean/imply?

Page 12 conclusions: should be especially checked for level of English.

Page 12 line 7: not all 6 cover this period.

Page 12 line 7: rewrite, ABF does not select..

Page 12 line 8: growing elevation?

Page 12 line 10: but what does it mean/imply that is the most restrictive? When would you recommend the ABF method?

Page 12 line 14: what do reduced and delayed mean here?

Page 12 line 18-19: what do you mean?

Page 12 line 21: how applicable is the method to other stations?

Figure 1: add larger map to know which region of the world this is.