Answers to

Review of revised manuscript by Yuan et al.

"Adaptive selection of Diurnal Minimum Variation: a statistical strategy to obtain representative atmospheric CO_2 data and its application to European elevated mountain stations"

We would like to thank the referee again for reading our comments carefully and giving comments. Answers are given under each new comment, and the corresponding changes are made in the main text. Moreover, slight changes regarding affiliations, acknowledgement, and level of English were made.

We keep the color code as it is in the review. Some comments from the first review are repeated here in gray. Author replies are included in blue, and new comments are given in black.

The new replies are shown in red.

The manuscript has improved in this revision, the authors have especially clarified parts of the text. The replies to my comments are also helpful, and clarify the method and some issues. However, there are a number of questions that have been answered only in the reply, and the clarifications were not included in the main text. I highlight these questions/answers below and think this should still be included in the manuscript. Besides these suggestions, I also give some comments below on the revised version.

Thank you for your comment. We have carefully checked these questions/answers listed below and included changes accordingly in the main text. Details can be seen later.

The authors write that the manuscript was proofread for English, however, the level of English in this new version still does not seem adequate and would need to be improved.

Thank you for your suggestion. We have gone through the main text again carefully and tried to improve the English. Certain changes can be seen in the main text.

All in all, as I wrote in the review of the first version of the manuscript, I think the work done by the authors is interesting enough to be published, after taking into account the following comments.

Thank you for your kind comment.

Some comments on the revised manuscript or replies to the first review:

I agree with the second reviewer that replacing the terminology "baseline" in the name of the presented method is a good idea. However, I think that the new name "Adaptive Diurnal Minimum Variation (ADMV)" is not clear and possibly not correct English, and specifically I do not understand what the authors mean by "variation". The previous choice for "finder" gave an idea of the goal of the new technique, this is now missing. In the title it is more clear, as it includes 'selection'. Please reconsider the name again, possibly by including 'selection technique' or a similar term in the full name.

Thank you very much for your suggestion. After discussion, we changed the name to "Adaptive Diurnal least Variation Selection (ADVS)". We added "selection" to the name and adjusted the order of phrasing. Also, "variation" in the name indicates that it is a statistical measure of variation in the diurnal cycle of the time series.

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.

Hourly values are used because of the availability of hourly averages as the highest time resolution in the World Data Center for Greenhouse Gases (WDCGG). Therefore in order to keep the format of input data constant for ADMV method, we calculated the average of the whole hour for all data sets. The time stamp for the hourly average was defined as the beginning of the averaging interval.

Moreover, originally ADMV was developed based on 30-min time resolution at the station ZSF. Therefore ADMV method can also handle data with higher time resolution than one hour.

We added in Section 2.1, "In addition, the time stamp was defined at the beginning of the averaging interval."

The argument to use hourly data because that is what is available on WDCGG does not make sense, as for each measurement site there is a co-author included in the manuscript, suggesting that they contributed to the research. Since continuous data is available, it would have been more solid if the authors had analyzed the influence of using hourly averages, versus original data. The reply that the method works on half hourly averages in that sense does not add much information, it would be more informative to compare the outcomes, and conclude what would be the optimal time resolution of the observations. Can the authors add some information on that?

Thank you very much for your comment. The purpose of using hourly averages based on the availability on WDCGG is that, it is easily applicable for most of the users with open data (most frequently as hourly data) to apply our data selection method. A central purpose is to design an applicable methodology for Global Atmosphere Watch. And the worldwide GAW Database has no finer time resolution than 60 minutes.

Moreover, we agree that analyzing the influence of hourly averages versus original data would make a difference. However, the original time resolutions of validated data sets we got from different stations are different (from seconds, 30-min, to hourly). Therefore it is difficult to perform an overall analysis for all stations. We have included such comparisons at 2 sites in the supplement (S1.3 Time resolution *tr*). We compared 30-min time resolution with hourly averages at ZSF, showing a significant difference in the resulted selection percentages. Then we compared the data sets at JFJ by using time resolution of 10-min, 20-min, 30-min, and hourly. As a result, it follows the same pattern that, the higher the time resolution is, the lower the selection percentage is. This can be explained by the statistical property of our data selection method focusing on the variation of data under a certain time period. Therefore, with more data (higher time resolution) in the same time period, more variation could be derived so that less data can be selected based on the same threshold criteria.

As a result, the use of higher time resolution requires more strict control of the measurement process and better measurement data quality, in order to derive of the similar level of selection percentages. As a balance between selection percentage and time resolution regarding applicability, we would still recommend the use of hourly averages.

Therefore, we changed the following sentence on Page 4 line 20 as, "For this study unless otherwise indicated, hourly data were used consistently for the purpose of evaluating the data selection method as practical as possible."

Page 11 line 10: is this referring to the global growth rate? Why would you compare to an average over the last 10 years, and not to the mean growth rate during the same time period as your data sets? You could use e.g. the annual global growth rates from https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html.

Thank you very much for your suggestion. We recalculated the mean annual global CO_2 growth rates from 2010 to 2015 based on the data from https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html for comparison in the main text now. Then we changed the sentence as, "the mean annual growth rate of *IZO* agrees well with the mean annual global CO_2 growth rates (2.31 ppm) during the same time period (2010-2015) based on data from https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html."

Page 13 line 16 'increasing percentage of data': explain that this is the amount of data left after selection and that it is supposed to represent background conditions. In that way it is more clear to read the conclusions without the rest of the text.

We rewrote the sentence as, "The ADMV method resulted in an increasing number of selection percentages representing the background conditions with growing altitude of measurement sites, which is reasonable due to the underlying atmospheric dynamics."

Page 13 line 19 and conclusions: 'ADMV is the most restrictive'. This should be put in perspective? Is it good or bad to have a restrictive filter? Would ADMV result in better representative conditions than the other methods?

When we express, that ADMV is the most restrictive in accepting data as being representative compared with the other methods in this comparison, it cannot be argued with good or bad a priori. Rather, we face a problem in two respects.

The comparatively low yield results from the physical property of a good mixing of the lower free troposphere. As such, however, it is not good or bad, but only shows the variation of a measured physical property.

Of course, in the sense of scientific practice, it is always the aim of an investigation to generate as much representative data as possible. But what would be the point if a larger amount of data could not be used to explain the conditions in the lower free troposphere? Therefore, it is better to obtain only a comparatively low yield of representative data if it is better suited to draw conclusions about the

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!

We apologize for the misleading wording.

VAL data are validated correct measurements, adjusted to the international standard reference scales and following the Global Atmosphere Watch quality objectives. Nevertheless, due to the different time scales of transportation effects VAL data may contain values from a time period where the well mixing assumption is violated (short time events). Since we referred VAL as validated unselected measurements, the CO_2 levels mentioned here refer to the background level of CO_2 which are supposed to take place at the measurement sites.

If all validated data are used, this would result in an overestimation of the atmospheric CO_2, due to the dominance of anthropogenic activities and no active vegetation in winter. Thus, it indicates that the VAL data are not representative.

Part of the answer here shows that my question was misunderstood: 'If all validated data are used, this would result in an overestimation of the atmospheric CO_2'. This is not true, as validated data are the measurements of atmospheric CO_2 concentrations themselves. Yes, this includes local effects etc. and are not background levels, but it is the actual CO_2 concentration at that location. Change on page 11 line 21-22 to e.g.: 'indicating that the CO_2 concentrations estimated by VAL are above the background levels'.

Sorry for the misunderstanding, and thank you very much for your clarification.

This was corrected in main text as, "When taking a closer look at the monthly averages, lower CO_2 values are found in the selected datasets in the winter months from October to April, indicating that the CO_2 concentrations estimated by VAL are above the background levels because of more dominant anthropogenic activities and no active vegetation."

In a number of places, I had raised some questions with the intention to also clarify these issues in the manuscript itself. In some cases the authors have just answered these questions in the reply, but have not updated the text accordingly. I would recommend to revise the text especially to include parts of the following replies in the text:

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

The different start time windows at the different levels result automatically from the ADMV method. It always searches for the optimal start time window based on specific data sets. In our opinion, these are very interesting and valuable results, which reflect to some extent the different characteristics of

different measurement sites and also different levels. In this respect, the different time windows at the different sampling levels are results of differences in the dynamics of atmospheric transport.

We rewrote the corresponding sentences in the main text on page 8 line 28 as, "The observed differences in the *start time windows*, as well as in the widths of the confidence intervals (gray shades), reflect the characteristics of differently situated measurement sites and different sampling levels. The first subplot column (*HPB50*, *HPB93*, and *HPB131*), representing the three sampling heights at *HPB*, shows similar detrended diurnal patterns with similar *start time windows*. The slightly different *start time window* at *HPB131* potentially indicates different dynamics of the atmospheric transport at higher elevation."

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?

We agree that the data were not fully comparable because the time period was too short in contrast to the other stations. However, the results showed that for time periods shorter than a full year, the ADMV method was still applicable to the data from the tower measurements, which highlights the flexibility of the approach.

We rewrote the corresponding sentences in the main text on page 9 line 5 as, "The results may not be fully comparable, but instead it shows that the data selection method is also applicable to data with time periods shorter than one year."

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.

We agree that there are differences in diurnal patterns among seasons. We also applied the ADMV method separated by season, i.e., data sets were processed and selected by the ADMV method only during a specific season over the whole time period. However, we found that the start time windows didn't differ significantly (see Supplement S1.1).

Regarding different time periods of the sites we also included data of 2015 for SSL. Now except for HPB, all the measurement sites cover the same time period.

We understand this comment. And we have made such comparisons of *start time windows* by seasons in Supplement S1.1. And in order to clarify this problem as early as possible, we made such an argument in the method section on page 6 line 10 as, "Being aware that calculating the *start time window* from overall data could differ from the *start time windows* calculated by season, the overall generated *start time windows* have been compared with seasonally generated *start time windows* for high elevated mountain stations (see Supplement S1.1). Because these differences were mostly minimal to moderate and this work aims at a methodical comparison under identical conditions, the *start time windows* are always derived from overall data."

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

This would be one of our next research questions and would be tested in the near future.

We rewrote the corresponding sentences in the main text on page 14 line 4 as, "Finally, it would be extremely interesting to test as the next step, whether this presented method is applicable to stations in other regions and on other continents."

Adaptive selection of diurnal least variation Diurnal Minimum Variation: a statistical strategy to obtain representative atmospheric CO₂ data and its application to European elevated mountain stations

Ye Yuan[1], Ludwig Ries[2], Hannes Petermeier[3], Martin Steinbacher[4], Angel J. Gómez-Peláez[5,11], Markus C. Leuenberger[6], Marcus Schumacher[7], Thomas Trickl[8], Cedric Couret[2], Frank Meinhardt[9], Annette Menzel[1,10]

[1]Department of Ecology and Ecosystem Management, Technische Universität München, Freising, Germany
[2]German Environment Agency (UBA), Zugspitze, Germany
[3]Department of Mathematics, Technische Universität München, Freising, Germany
[4]Empa, Laboratory for Air Pollution/Environmental Technology, Dübendorf, Switzerland
[5]Izaña Atmospheric Research Center, Meteorological State Agency of Spain (AEMET), Santa Cruz de Tenerife, Spain
[6]Climate and Environmental Physics Division, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[7]Meteorological Observatory Hohenpeissenberg, Deutscher Wetterdienst (DWD), Hohenpeissenberg, Germany
[8]Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany
[9]German Environment Agency (UBA), Schauinsland, Germany
[10]Institute for Advanced Study, Technische Universität München, Garching, Germany
[11]At present: Meteorological State Agency of Spain (AEMET), Delegation in Asturias, Oviedo, Spain.

Correspondence to: Ye Yuan (yuan@wzw.tum.de)

Abstract. Critical data selection is essential for determining representative baseline levels of atmospheric trace gases even at remote measurement sites. Different data selection techniques have been used around the world, which could potentially lead to reduced compatibility when comparing data from different stations. This paper presents a novel statistical data selection method named Adaptive Diurnal least Variation Selection (ADVS) Adaptive Diurnal Minimum Variation (ADMV) based on CO₂ diurnal patterns occurring typically at high elevated mountain stations. Its capability and applicability were studied on records of atmospheric CO₂ observations at six Global Atmosphere Watch stations in Europe, namely, Zugspitze-Schneefernerhaus (Germany), Sonnblick (Austria), Jungfraujoch (Switzerland), Izaña (Spain), Schauinsland (Germany), and Hohenpeissenberg (Germany). Three other frequently applied statistical data selection methods were included for comparison. Among the studied methods, our ADVS ADMV method resulted in lower selection percentages with lower maxima during winter and higher minima during summer in the selected data, which can be understood as a better representation of the lower free troposphere. The measured time series were analyzed for long-term trends and seasonality by seasonal-trend decomposition technique. Compared with unselected data, mean annual growth rates of all selected datasets were not significantly different, except for the data recorded for Schauinsland. However, clear differences were found in the annual amplitudes as well as the seasonal time structure. Based on correlation analysis, results obtained by ADVS ADMV selection showed a better representation of the lower free tropospheric conditions.

1 Introduction

Continuous in situ measurements of greenhouse gases (GHGs) at remote locations have been established since 1958 (Keeling, 1960). Knowledge of background atmospheric GHG concentrations is key to understanding the global carbon cycle and its effect on climate as well as the GHG responses to a changing climate. A crucial issue when using data from remote stations remains the identification of time periods that are representative of larger spatial areas and their differentiation from periods influenced by local and regional pollution. If these two regimes are well disaggregated, the available datasets can represent more reliable information about long-term changes of undisturbed atmospheric GHG levels or be used to investigate local and regional GHG sources and sinks when specifically analyzing the deviations from the baseline conditions. In this study, the baseline conditions refer to a selected subset of data from the validated dataset, representing well-mixed air masses with minimized short-term external influences (Elliott, 1989; Calvert, 1990; Balzani Lööv et al., 2008; Chambers et al., 2016).

Measurement results depend on sampling methods, analytical instrumentation, and data processing. Validated data (labeled as VAL in this study to differentiate from the selected data) are usually obtained after signal correction, for example, due to interferences from other GHGs such as water vapor, calibration accounting for sensitivity changes of the analyzer, and validation based on plausibility checks. Data selection starts with validated data and identifies in subsequent steps a final subset of the validated data set based on predefined criteria for specific qualities such as representativeness. With a particular focus on CO_2 in this study, it is commonly accepted that data selection methods can be categorized into meteorological, tracer, and statistical selection methods (Ruckstuhl et al., 2012; Fang et al., 2015).

Meteorological data selection makes use of the meteorological information at the measurement sites, which provides valuable information about the surrounding environment as well as air mass transport (Carnuth and Trickl, 2000; Carnuth et al., 2002). Forrer et al. (2000), Zellweger et al. (2003), and Kaiser et al. (2007) intensively studied the relationship between measured trace gases (such as O_3, CO, and NO_x) and meteorological processes at Zugspitze, Jungfraujoch, Sonnblick, and Hohenpeissenberg. For CO_2, the most common parameters applied in the literature are wind speed and wind direction. They can provide information on critical variations at stations with sources and sinks in their vicinity, while these parameters are less suited at stations in largely pristine environments. For example, Lowe et al. (1979) performed a pre-selection on the CO_2 record at Baring Head (New Zealand) during the southerly wind period only (clean marine air). In addition, Massen and Beck (2011) found that the CO_2 versus wind speed plot can be valuable for baseline CO_2 estimation without a local influence of continental measurements. Besides, fixed time window selection has been widely used, by selecting data in a certain time interval of the day based on local and mesoscale mechanisms of air mass transportation. For selecting well-mixed air at elevated mountain sites, nighttime is usually chosen with a special focus on the exclusion of afternoon periods due to the influence of convective upward transport (Bacastow et al., 1985). Brooks et al. (2012) also limited their mountaintop CO_2 results in the Rocky Mountains (USA) by "time-of-day" from 0 a.m. till 4 a.m. local time (LT) to increase the likelihood of sampling the free tropospheric environment at the station. Apart from this, modeling techniques such as backward

trajectories are very helpful for analyzing in detail the origins and transport processes of air masses arriving at the station (Cui et al., 2011). Uglietti et al. (2011) focused on the origins of atmospheric CO_2 at Jungfraujoch (Switzerland) by the FLEXible PARTicle dispersion model (FLEXPART). Using tracers, data selection can be performed by investigating the correlations between the air components of interest. Many tracers have been applied and compared with CO_2. Threshold

5 limits of 300 ppb for CO and 2000 ppb for CH_4 were defined by Sirignano et al. (2010) to perform a regional analysis of CO_2 data at Lutjewad (the Netherlands) and Mace Head (Ireland). Similar approaches with black carbon and CH_4 were performed by Fang et al. (2015) at Lin'an (China). Moreover, Chambers et al. (2016) developed and applied a data selection technique to identify baseline air masses using atmospheric radon measurements at the stations Cape Grim (Australia), Mauna Loa (Hawaii, USA), and Jungfraujoch (Switzerland).

10 Unlike most of the methods mentioned above, which require additional data or advanced transport modeling, statistical data selection only relies on the time series of interest and typically investigates the variability of signal. It is usually assumed that the most representative CO_2 data are found during well-mixed conditions revealing small variations in time (Peterson et al., 1982) and in space (Sepúlveda et al., 2014). For continuous measurements, it is possible to investigate within-hour and hour-to-hour variability in the datasets. The within-hour variability is often expressed as the standard deviation of the measured

15 data within 1 h. The hour-to-hour variability compares the differences between hourly averaged concentrations either during a certain time period, or from one hour to the next. Pales and Keeling (1965) marked ambient data as "variable" when the within-hour variability for the air sample is significantly larger than the within-hour variability for the reference gas. Consequently, they only selected CO_2 data in "steady" conditions for 6 h or more. Besides, Peterson et al. (1982) also rejected sampled CO_2 data values for adjacent hours when the hour-to-hour variability exceeded 0.25 ppm. Thoning et al.

20 (1989) combined these two strategies using an iterative approach by selecting data according to deviations of daily averages from a spline curve fit. Ruckstuhl et al. (2012) developed a method based on robust local regression, called Robust Extraction of Baseline Signal, to estimate the baseline curves generalized for atmospheric compounds, which is available in the R package IDPmisc (Locher and Ruckstuhl, 2012).

The present study focuses on the comparison of results from statistical data selection methods and the Adaptive Diurnal least

25 Variation Selection (ADVS) Adaptive Diurnal Minimum Variation (ADMV). The ADVS ADMV is seen as a possible alternative to already known data selection methods as discussed above. By applying ADVS ADMV to the atmospheric CO_2 records from six European elevated mountain stations, the selection selected results are compared with those derived from three other statistical data selection methods. To investigate the potential influences of the data selection method on the trend and seasonality, further analyses focus on the decomposition of validated and selected datasets in trend and seasonal

30 components. Finally, differences between ADVS ADMV and other data selection methods were assessed by correlation analysis.

2 Methods

2.1 CO$_2$ measurements at elevated European sites

CO$_2$ measurements from six ~~selected~~ European mountain stations (see Fig. 1) within the Global Atmosphere Watch (GAW) network were used ~~to test the data selection algorithms~~. The data ~~have been~~ were taken from mountain stations due to their

5   remote location~~s~~, ~~which results in them~~ being subjected to limited anthropogenic influence and ~~having~~ increased representativeness. Three high alpine measurement sites were included: Zugspitze-Schneefernerhaus (*ZSF*, DE, 47°25′ N, 10°59′ E, 2670 m a.s.l.), Jungfraujoch (*JFJ*, CH, 46°33′ N, 7°59′ E, 3580 m a.s.l.), and Sonnblick (*SNB*, AT, 47°03′ N, 12°57′ E, 3106 m a.s.l.). They are often above the planetary boundary layer (PBL) and thus exposed to free and assumed clean lower tropospheric air masses, but ~~are~~ periodically influenced by regional emissions from lower altitudes. Additionally,

10   to test data selection for a less remote environment, CO$_2$ measurements from Schauinsland (*SSL*, DE, 47°55′ N, 7°55′ E, 1205 m a.s.l.) at a clearly lower elevation in the mid-range Black Forest were investigated. Data selection was also applied to three recently started CO$_2$ time series from different sampling heights above ground at a tall tower at the Hohenpeissenberg observatory (*HPB*, DE, 47°63′ N, 11°01′ E, 934 m a.s.l.), located in the northern foothills of the Alps. Henne et al. (2010) presented a method of categorizing site representativeness based on the influence and variability of population and

15   deposition by the surface fluxes. *JFJ* and *SNB* ~~are~~ were classified as "mostly remote," while *ZSF* ~~is~~ was considered as "weakly influenced, constant deposition," and *SSL* and *HPB* ~~are~~ were considered as "rural" (Henne et al., 2010). Finally, Izaña station on Tenerife Island ~~in the North Atlantic~~ (*IZO*, ES, 28°19′ N, 16°30′ W, 2373 m a.s.l.) in the North Atlantic was chosen as a reference ~~for comparison~~ due to its location above the subtropical temperature inversion layer, which means that the station is rarely affected by any local or regional CO$_2$ sources and sinks (Gomez-Pelaez et al., 2013).

20   For this study unless otherwise indicated, hourly data were used consistently for the purpose of evaluating the data selection method as practical as possible~~, unless otherwise indicated~~. The validated CO$_2$ hourly averages from all stations were downloaded from the World Data Centre for Greenhouse Gases (http://ds.data.jma.go.jp/gmd/wdcgg/). Data with higher time resolution required for this study were provided directly by the station investigators. In addition, the time stamp was defined as the beginning of the averaging interval. Descriptions of the sampling elevation and time period of available data are given

25   in Table 1. Further information on each station can be found in ~~the report by~~ Schmidt et al. (2003) for *SSL*, Gilge et al. (2010) for *HPB* and *SNB*, Gomez-Pelaez et al. (2010) for *IZO*, Risius et al. (2015) for *ZSF*, and Schibig et al. (2015) for *JFJ*. Practical data selections and analyses in this study were performed in the R Statistical Environment (R Core Team, 2017).

2.2 ADVS ~~ADMV~~

ADVS ~~ADMV~~ is a tool for automated and systematic analysis of diurnal CO$_2$ cycles at elevated mountain stations in order to

30   select consecutive time sequences with minimum variation, which can be regarded as representing well-mixed air conditions. Even though such measurement sites are remotely located, the CO$_2$ levels are still influenced by local sources and sinks. For example, at *ZSF*, these can be characterized by anthropogenic CO$_2$ sources, detectable especially in winter during the day,

4

whereas in summer the convective upwind transport results in a stronger impact of air masses with depleted CO_2 concentrations due to photosynthesis at lower altitudes. Plant respiration activities, which may contribute small amounts, are primarily not visible in the convective upwind air masses (which arrive at mountain sites predominantly in the afternoon). Although high elevated mountain stations do not have vegetation in their surroundings, mountain stations at lower altitudes

5 but still in the vegetation zone may be influenced by plant respiration, especially at night. ~~This~~ It points out the importance of finding a certain diurnal time window representing the most stable ~~and representative~~ CO_2 level, which in turn can be used for selecting representative data. However, the duration of this time window during the day varies with the season and from day to day because of variations in the dynamics of transport to the site (e.g., Birmili et al., 2009; Herrmann et al., 2015). In summer, larger variabilities in the CO_2 signal are observed due to more prevalent convective boundary layer air mass

10 injections influencing the diurnal pattern, resulting in shorter periods of stable conditions, whereas in winter, significantly longer stable periods occur. ~~In winter, no~~ No upwind air masses with depleted CO_2 levels due to photosynthesis by vegetation are recorded in winter. To receive as much representative data as possible, it is desirable to select the time window dynamically. ADVS ~~ADMV~~ is constructed to select a subset from the measured data, being best representative for baseline conditions with an adaptive selected time window specific for every day.

15 The algorithm is based on two basic assumptions. First, air masses measured at altitude stations contain well-mixed air, closest to baseline levels, within a certain time window of several hours during the day. For the elevated mountain stations discussed in this paper, this time interval is around midnight. Different diurnal patterns are apparent at each station, so the selected time window should be adjusted accordingly. Second, it is assumed that the real baseline conditions are not subject to local influences and thus represent air masses originating only from the uninfluenced lower free troposphere. This

20 indicates that the variability of the measured CO_2 signal should be minimal within this selected time window. The methodological steps of ADVS ~~ADMV~~ are introduced in detail below in the two sections *starting selection* and *adaptive selection*.

2.2.1 *Starting selection*

For a given validated hourly dataset, ADVS ~~ADMV~~ starts data selection by finding a *start time window* for all days. The

25 standardized selection procedure for the *start time window* results from site-specific parameters. This time interval is set as the most stable period from the diurnal variation. The step is referred to as *starting selection*. It begins by analyzing the mean diurnal cycle of the data input.

Step 1: Detrending is done by subtracting a 3-day average for each day, including the neighboring two days. It is the shortest possible time window to remove sudden changes in the time series related to the previous and posterior days while

30 preserving the diurnal pattern.

Step 2: The overall mean diurnal variation, $\bar{d}_i$ ($i = 0$ to 23 h), is calculated from the complete set of detrended data.

5

Step 3: The standard deviations s_{Δ_j} from the overall mean diurnal variation $\bar{d}_i$ are calculated on a moving window Δ_j ($j = 6\,h$). To be able to place a full set of 24 moving time windows over the overall mean diurnal variation, time windows across midnight (e.g., 6 h from 11 p.m. to 4 a.m. LT) are also included, that is, its first j hours are appended to the end of the 24 h in the overall mean diurnal variation. The time window with the smallest standard deviation is selected as the *start time window*.

Result: The *start time window* $[i_{start}, \ldots, i_{end}]$.

With the focus on elevated mountain stations, *starting selection* is purposely designed with the moving window Δ_j of 6 h, and the starting hour i_{start} to be between 6 p.m. and 5 a.m. LT for this study. For other stations with possibly different diurnal patterns, *starting selection* can be adjusted accordingly. For instance, at urban stations or stations completely within the continental PBL, the *start time window* can be chosen based on their best mixing conditions, which often occur in the afternoon with a shorter moving window, when the PBL reaches its maximum depth after "ingesting" free tropospheric air during its growth. Being aware that calculating the *start time window* from overall data could differ from the *start time windows* calculated by season, the overall generated *start time windows* have been compared with seasonally generated *start time windows* for high~~ly~~ elevated mountain stations (see Supplement S1.1). Because these differences ~~are~~ were mostly minimal to moderate and this work aims at a methodical comparison under identical conditions, the *start time windows* are always derived from overall data. ~~constant generation of *start time window* from overall data has been chosen.~~

2.2.2 *Adaptive selection*

The second part *adaptive selection* is designed to determine the most suitable time window for each day, based on the data variability. The length of the *start time window* is adapted (expanding only) in both directions in time. *Adaptive selection* is performed on a daily basis, starting with the first day of the given dataset. The following steps only describe the *forward adaptive selection*. ADVS ~~ADMV~~ runs the *backward adaptive selection* in an analogous manner but going back in time.

Step 1: The mean molar fraction $\bar{x}_i$, standard deviation s_i, and the proportion of missing values $\pi_{missing}$ are calculated from data in the *start time window* $[i_{start}, \ldots, i_{end}]$.

Step 2: If $s_i \leq 0.3$ ppm (CO_2) and $\pi_{missing} \leq 0.5$, ADVS ~~ADMV~~ continues to advance in time to examine whether the next data point x_f can be included in the selected time window W with $f = i_{end} + 1$. Otherwise, it is considered that the *start time window* does not fulfill the assumptions. No data are selected for this day. One should proceed on to **Next Day**.

Step 3: The absolute difference between x_f and $\bar{x}_i$ is calculated, and the following threshold criterion is applied: $|x_f - \bar{x}_i| \leq \kappa \cdot s_i$, where κ is the threshold parameter. If this criterion holds, x_f is included in W and ADVS ~~ADMV~~ continues. Otherwise, ADVS ~~ADMV is finished only with the *start time window*~~ stops for this day with only the *start time window*, and one should proceed on to **Next Day**.

Step 4: Mean $\bar{x}_W$ and standard deviation s_W for the new selected time window W are calculated. If $s_W \leq 0.3$ ppm (CO_2), ADVS ADMV continues with the next data point x_f while $f = f + 1$. Otherwise, ADVS ADMV is finished stops for this day with the previous selected time window and one should proceed on to **Next Day**.

Step 5: The new absolute difference between x_f and $\bar{x}_W$ is calculated, as well as the new threshold criteria. If condition $|x_f - \bar{x}_W| \leq \kappa \cdot s_W$ holds, x_f is included in W and ADVS ADMV goes back to **Step 4**. Otherwise, ADVS ADMV is finished stops for this day and one should proceed on to **Next Day**.

When data selections for all days is are finished, ADVS ADMV continues with *backward adaptive selection*. Afterwards, one should proceed on to **Result**.

Result: This is the final selected time window, which is a combination of $W_{forward}$ and $W_{backward}$ for the referring day.

The following limitations of the forward and backward expansions of the time window should be considered. ADVS ADMV always runs for no longer than 24 h including the *start time window*, namely, $f \leq 24 \cdot tr$, where tr is the time resolution in data points per hour of the input data. This sometimes results in an overlap of "selected" and "unselected" data for two consecutive days. We always label the data as "selected" once it has been selected by ADVS ADMV. The threshold parameter κ is the controlling factor of ADMV for the length of the selected time window. As κ increases, the length of the selected time window increases becomes larger. The value of 2 was chosen heuristically for this study as a compromise between selecting as many data points as possible and achieving the least data variability. Similar values of sensitivity-controlling parameters in other data selection methods can be found (Thoning et al., 1989; Sirignano et al., 2010; Uglietti et al., 2011; Satar et al., 2016). In **Step 2**, values of 0.3 ppm and 0.5 indicate the threshold values for s_i and $\pi_{missing}$. We denote them as $s_{i,threshold}$ and $\pi_{missing,threshold}$. It has been shown that less remote stations at lower altitudes require a larger value than 0.3 ppm because of different mixing conditions. When performing ADVS ADMV data selection at lower sites such as *HPB* and *SSL*, we recommend a higher $s_{i,threshold}$, such as 1.0 ppm instead of 0.3 ppm. However, throughout this study, we used the described parameter setting (0.3 ppm) for a methodical inter-comparison of selection methods at all stations. Potential influences of these parameter sizes ($s_{i,threshold}$ and tr) are were discussed in Supplements S1.2 and S1.3.

2.3 Other statistical data selection methods for comparison

We compared ADVS ADMV with three statistical data selection methods. The first method named SI is based on "steady intervals" (Lowe et al., 1979; Stephens et al., 2013). Steady intervals, which are considered as baseline conditions, are defined by a standard deviation being lower than or equal to 0.3 ppm for 6 or more consecutive hours.

Second, we adopted a method applied by NOAA ESRL, which originated from Thoning et al. (1989). This selection routine has been applied specifically for measurements of background CO_2 levels at Mauna Loa. This method (referred to as THO) was applied as described on the website http://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html. The first step of THO examines the within-hour variability by selecting hours with hourly standard deviation less than 0.3 ppm. Second, it computes hourly averages and checks the hour-to-hour variability by retaining any two consecutive hourly values where the

hour-to-hour difference is less than 0.25 ppm. The last step is based on the diurnal pattern (similar to ADVS ~~ADMV~~), by excluding data from 11 a.m. to 7 p.m. LT due to transported air influenced by photosynthesis.

The last method compared is a moving average technique (MA). A moving time window of 30 days and a threshold criterion of two standard deviations from the moving averages were applied to discard outliers. Afterwards, new moving averages and

5 new threshold criteria were calculated for data exclusion. This step ~~is~~ was repeated until no more outliers can be found. A more detailed description can be found in Uglietti et al. (2011) and Satar et al. (2016).

2.4 Seasonal-trend decomposition STL

To analyze ~~and compare~~ the selected results from different data selection methods and compare them with ~~as well as~~ the original validated datasets, we applied the seasonal-trend decomposition technique based on locally weighted regression

10 smoothing (Loess), named STL (Cleveland, 1979; Cleveland et al., 1990). STL has been widely used on measurements of atmospheric CO_2 and other trace gases (Cleveland et al., 1983; Carslaw, 2005; Brailsford et al., 2012; Hernández-Paniagua et al., 2015; Pickers and Manning, 2015). It decomposes a time series of interest into a trend component T, a seasonal component S, and a remainder component R, which allows separate detailed analyses ~~and comparisons~~ of trend and seasonality. Two recursive procedures are included in the STL technique: an inner loop where seasonal and trend smoothing

15 based on Loess are performed and updated in each pass, and an outer loop that computes the robustness weights to reduce the influences of extreme values for the next run of the inner loop (Cleveland et al., 1990).

For this study, we used the implemented function `stl` in R (R Core Team, 2017). Owing to functional limitation~~s~~ of ~~function~~ `stl`, full time coverage of monthly data is needed in order to reduce the risk of large time gaps or unequal spacing (Pickers and Manning, 2015). All data ~~results~~ were first aggregated to monthly averages. Then, missing data were substituted

20 by linear interpolation, using R function `na.approx` (Zeileis and Grothendieck, 2005). For the application of STL, two parameters need to be specified, which are the seasonal smoothing parameter $n_{(s)}$ ($s.window$ in function `stl`) and the trend smoothing parameter $n_{(t)}$ ($t.window$ in function `stl`). As $n_{(s)}$ and $n_{(t)}$ increase, the seasonal and trend components get smoother (Cleveland et al., 1990). For optimal compatibility in this study, the same parameters were chosen for all stations as $n_{(s)} = 7$ and $n_{(t)} = 23$, based on the recommendation of Cleveland et al. (1990). Another parameter combination of

25 $n_{(s)} = 5$ and $n_{(t)} = 25$ was also tested according to Pickers and Manning (2015), but with no significant differences in results.

3 Results and discussion

3.1 *Start time window*

ADVS ~~ADMV~~ was applied to the validated hourly averages from all six stations with the parameter settings as described

30 above. The detrended mean diurnal cycles were obtained with the *start time window* for each station by *starting selection*

(see Fig. 2, for conventional mean diurnal plots see Supplement S2). The observed differences in the *start time windows*, as well as in the widths of the confidence intervals (gray shades ~~in Fig. 2~~), reflect the characteristics of differently situated measurement sites and different sampling levels. ~~can be explained by different site environments and thus differing data variabilities.~~ The first subplot column (*HPB50*, *HPB93*, and *HPB131*), representing the three sampling heights at *HPB*, shows similar detrended diurnal patterns with similar *start time windows*. The slightly different *start time window* at *HPB131* potentially indicates different dynamics of the atmospheric transport at higher elevation. The decreasing amplitude with increasing sampling height indicates that the higher the sampling inlet is above the ground, the less it is affected by the local surface fluxes. The three *start time windows* suggest that the most stable period at *HPB* occurs during the last few hours of a day, also including midnight. However, in contrast to all other stations covering at least a full year, *HPB* data are only from September of 2015 to June of 2016. The results may not be fully comparable, but instead it shows that the data selection method is also applicable to data with time periods shorter than one year.

Regarding the second subplot column (*SSL*, *SNB*, and *IZO*), the *start time windows* can be found from midnight on or later in the morning. The *start time window* for *SSL* encompasses its diurnal maximum, indicating that data variability is considerably smaller in the early morning than in the afternoon because of its vicinity to the Black Forest region, which has strong influence due to local photosynthetic activities (Schmidt et al., 2003). A similar diurnal pattern can be found at *SNB*. The influence of CO_2 sources is not as prominent as the effect of distant CO_2 sinks, since it is situated at the single summit peak of Hoher Sonnblick only surrounded by mountains and glaciers, with a negligibly small number of tourists, thus anthropogenic activities are minimal. *IZO* is a special case, since it is located on a remote mountain plateau on the Island of Tenerife above the strong subtropical temperature inversion layer. Even though the *start time window* is limited to 6 h, *IZO* presents an ideal mean diurnal cycle for data selection from a potentially much longer time window.

In the right subplot column, both *ZSF* and *JFJ* find their *start time windows* around midnight (with more hours after midnight). *ZSF* shows higher diurnal CO_2 amplitude than *JFJ*, but the two sites show similar diurnal patterns. For the choice of the *start time window* from the mean diurnal variation, relatively close or even local anthropogenic sources may influence the CO_2 at these two stations, possibly due to touristic influences.

3.2 Selection percentage

With the determined *start time windows*, ADVS ~~ADMV~~ selected the data for all stations (see Fig. 3). In addition, we calculated the percentages of ADVS ~~ADMV~~ selected data values among all values of the complete datasets for all stations, which are listed in the first column of Table 2. The higher the selection percentage is the more well-mixed air is measured at the station, which is assumed to be a representation of lower free tropospheric conditions. This holds especially for *IZO*. Because of this the greatest amount of accepted data points with 36.2% was found at this station. The sites with intermediate percentages are *JFJ* (22.1%), *SNB* (19.3%), and *ZSF* (14.8%). For the three sampling heights at *HPB*, only 3.2% (50 m), 4.8% (93 m), and 6.2% (131 m) of the data were selected by ADVS ~~ADMV~~. Finally, a similar percentage was found for *SSL* (4.0%), probably due to its higher data variability.

9

Since the stations were listed according to their altitudes, it ~~was~~ is clear that all four selection percentages increase with altitude, indicating ~~which indicated~~ that measurements at higher altitudes ~~could~~ can capture progressively well-mixed and hence representative air. Therefore, linear least-squares regression was applied between the absolute altitudes and the selection percentages for continental stations. *IZO* ~~was~~ is on a remote island and therefore not comparable. This approach

5   ~~revealed~~ reveals a significant positive linear trend (see coefficient in Table 2). The related figure of linear regression can be found in Supplement ~~S3.2~~ S3.1.

To examine the characteristic growth of ADVS ~~ADMV~~ selection percentages during the selection process, we additionally calculated selection percentages after completing both the *starting selection* and *adaptive selection* steps mentioned in Section 2.2 (see Supplement ~~S3.1~~ S3.2). All ~~of the~~ results of percentages show an order of stations similar to that above, and

10 the selection percentages increase steadily step by step for all stations. The selection percentages of ADVS ~~ADMV~~ were then compared ~~again~~ with those of the ~~already~~ mentioned statistical data selection methods SI, THO, and MA ~~in~~ (see Table 2, with the corresponding figure shown in Supplement S3.3).

Since the selection percentages indicate not only the amount of data declared as representative but also show the characteristics of the selection methods, this criterion ~~was~~ is used for further assessment. All other methods except for MA

15 ~~resulted~~ results in higher selection percentages for ~~more highly~~ higher elevated stations (*IZO*, *ZSF*, *SNB*, and *JFJ*) than for less elevated ones (*HPB* and *SSL*). ADVS ~~ADMV~~ always performs the strictest in all cases. Based on the stepwise study of the selection percentages (see Supplement ~~S3.1~~ S3.2), such low percentages are due to the precise definition of the *start time window*. With *adaptive selection*, the selection percentages have grown but are still lower than those of the other methods. SI and THO, on the other hand, show differences between stations at high and low elevations. Compared with SI, THO is

20 higher at stations at lower elevations, but lower at high ones. A major limitation of SI seems to be the requirement for consecutive hours, in our case of 6 h with 0.3 ppm standard deviation threshold, which might be too restrictive for stations at lower elevations. However, this criterion results in a fairly large percentage for stations at high elevations. At *ZSF*, *SNB*, and *JFJ*, it results in the second largest, and even the largest in the case of *IZO*.

The highest selection percentages of approximately 80% were obtained with MA at most stations except for *IZO*. ~~However,~~

25 ~~owing to the minimal data variability of CO_2 measurements at *IZO*, the selection interval in MA becomes so small that the selection percentage becomes considerably smaller than at all other stations.~~ However, *IZO* obtains the largest selection percentages from all other selection methods. This is probably caused by the very low data variabilities of CO_2 measurements at *IZO*, resulting in too strict moving average thresholds for MA method. Thus, we conclude that MA does not work properly in the case of very well-mixed air (*IZO*). At all other stations, it is possible that MA declares too much data as

30 representative. Therefore, MA was excluded from further analyses.

3.3 STL decomposed results

STL was applied to the validated datasets before and after selection with SI, THO, and ADVS ~~ADMV~~, except for *HPB* due to its limited length of time (less than one year). Depending on data availability, STL was performed on CO_2 data from 2012

to 2015 at *SNB*, while data inputs at *SSL, IZO*, *ZSF*, and *JFJ* cover the whole period from 2010 to 2015. Figure 4 gives an overview of the decomposition in each component by STL. The following sections discuss the resulting components obtained by STL, namely the trend component ~~over the observation period~~, the seasonal component, and ~~finally~~ the remainder component.

3.3.1 Trend component

From the trend component~~s~~, the mean annual growth rate~~s~~ ~~is~~ were estimated by linear regression (see Table 3). Based on the 95% confidence interval~~s~~ for the slope, ~~a~~ positive trend~~s~~ i.e. increasing CO_2 concentrations are observed. Owing to the overlap of the confidence intervals, differences in the mean annual growth rates among VAL and selected datasets at the same station are all in good agreement. This indicates that the trend component is not influenced by the statistical data selection method, which agrees well with the finding of Parrish et al. (2012) from a study of baseline ozone concentrations that there were no significant differences of the long-term changes between the baseline and unfiltered datasets. Moreover, the following fact is observed for all sites except for *SSL*. Compared to unselected data (VAL), the mean annual growth rates based on selected datasets are systematically higher approaching the growth rates at *IZO*. *IZO* can be considered as better representing the lower free tropospheric conditions and agrees well with the mean annual global CO_2 growth rates (2.31 ppm) during the same time period (2010-2015) based on data from https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html. ~~the mean annual absolute increase during last 10 years (2.21 ppm yr⁻¹) reported by WMO (2017a).~~ The exception at *SSL* is probably caused by stronger local influences as a result of its lower elevation. Besides, the confidence intervals of the mean annual growth rates are always smaller after data selection, which improves the precision of trends.

3.3.2 Seasonal component

The resulting seasonal components show systematic differences between VAL and selected datasets. The mean monthly variations were calculated on a monthly scale over the entire period from the analyzed data. Figure 5 (a) and (b) present the results at stations *ZSF* and *IZO*. At most stations (except for *IZO*), the seasonal amplitudes have been substantially reduced compared to VAL (see also Fig. 4). At *ZSF*, the averaged peak-to-peak seasonal amplitude, defined as mean seasonal maximum minus seasonal minimum, drops the most by 18.9% from VAL with the ADVS ~~ADMV~~ selected dataset. An explanation of this reduction is CO_2 signal exclusion from local sources and sinks by data selection. When taking a closer look at the monthly averages, lower CO_2 values are found in the selected datasets in the winter months from October to April, indicating that the CO_2 concentrations estimated by VAL are above the background levels ~~level is overestimated by VAL~~ because of more dominant anthropogenic activities and no active vegetation. Higher values in the summer months from May to September explain underestimation of VAL due to intensified upward transport of photosynthetic signatures resulting from vegetation. Similar patterns can be found at stations *SSL*, *SNB*, and *JFJ* (see Supplement S4). *IZO*~~, as expected by its location,~~ always shows the smallest seasonal amplitude and nearly uninfluenced monthly results between

11

VAL and selected datasets. Based on this consideration, it is very likely that the lower free troposphere will react with a delay to CO_2 concentration changes of effective sources and sinks on the ground, acting like an atmospheric memory.

A time delay of one month in the mean seasonal maximum is shown in Fig. 5 (a) at *ZSF* with selected datasets by SI and ADVS ~~ADMV~~ (March), compared with the maximum from the validated data (February). In addition, a similar time shift

5 can also be found by other selection methods at stations *SSL* (one-month delay from February to March by SI and ADVS ~~ADMV~~) and *JFJ* (two-month~~s~~ delay from February to April by SI, THO, and ADVS ~~ADMV~~). As for station *IZO* (April) in Fig. 5 (b) and station *SNB* (March), the seasonal maxima stay the same. The magnitude of these delays may be related to mixing features in the lower free troposphere. Rapid changes are usually observed close to sources and sinks, e.g., from anthropogenic and biogenic activities. Thus, the higher the station is above the boundary layer, the later the maxima during

10 the winter can be observed, because of the late response due to inhibited mixing conditions. However, this delay does not occur for the minima during the summer because of the very effective upward transport and more favorable mixing conditions at that time of year. Consequently, no change~~s~~ in the seasonal minima ~~are~~ is observed at all measurement sites, which is taken as an indicator of enhanced thickness of the mixing layer as good mixing conditions. Taking *ZSF* as an example, Birmili et al. (2009) observed low concentrations of particle number in winter and found it representative for the

15 free tropospheric air by analyzing the annual and diurnal cycles. From spring on, the warmer it gets the higher the PBL goes. The intense vertical atmospheric exchange during summer months results in a daily air mass transport from the boundary layer to reach *ZSF* due to thermal convection (Reiter et al., 1986; Birmili et al., 2009). Thus there are optimal transportation and mixing conditions. Therefore after data selection, the timing of seasonal peaks better corresponds among the stations.

3.3.3 Remainder component

20 The remainder component contains data with external and random influences. It has characteristics of random noise, being basically different from site to site and statistically uncorrelated with the general signal of CO_2 concentrations in the lower free troposphere (Thoning et al., 1989). The standard deviation of the remainder component is taken here as a measure for external influences (see Fig. 4). Table 4 shows the calculated standard deviations from the remainder components at each station. Comparable results are derived from all selected datasets. *SSL*, as the lowest elevated station, exhibits the most

25 variation. *IZO* with the smallest standard deviations in the remainder components proves to be the station least influenced by its surrounding environment. The three alpine measuring stations (*ZSF*, *SNB* and *JFJ*) exhibit intermediate variability. From this perspective, STL performs well to show the site characteristics. Consequently, the noise of the remainder components, given in Table 4, decreases with increasing altitude~~s~~ of the continental mountain stations, which is in inverse relation to the selection percentage~~s~~ (Table 2). *IZO* was excluded in both regressions against altitude because of its maritime character.

30 **3.4 Correlation analysis**

As mentioned above, data selection is defined here as an approach of extracting a group of data to be the best representative for the lower free troposphere. Consequently, the selected CO_2 data~~-~~sets from all stations should theoretically agree better

among themselves. For validation of this, we took the combination of the trend and seasonal components from STL and examined the correlations between each pair of stations in a Pearson correlation matrix (see the upper panel of Fig. 6). The trend and seasonal components of all VAL and selected datasets were first compiled, and then Pearson's correlation coefficients were calculated assuming normal distribution of data examined by the Anderson Darling test (P < 0.05). The

5 correlation matrices are shown for each type of dataset individually. Data used for correlation were chosen only when available at all stations (2012–2015). In general, most pairs show higher correlation coefficients with selected data from the different selection methods, especially between the three Alpine stations (*ZSF*, *SNB*, and *JFJ*). This evaluation hence shows a similar result to the method presented by Sepúlveda et al. (2014) for identifying baseline conditions based on the correlation between distant measuring stations. Pairs including *IZO* after data selection by ADVS ~~ADMV~~ show a notable

10 increase in the correlation coefficients, meaning a better coherence between the reference station *IZO* and the others.

On the other hand, when selecting representative data more effectively, the results should contain less local and regional influences. Therefore, we compared the remainder components derived from STL pairwise to check whether the Pearson correlation coefficients decreased after data selection~~, as shown in~~ (see the lower panel of Fig. 6). The number of insignificant correlations between the station pairings is the greatest for ADVS ~~ADMV~~. For the only two coefficients

15 significant at the 0.05 confidence level (*ZSF-SNB* and *ZSF-JFJ*), they drops largely from 0.75 to 0.48, and from 0.75 to 0.40, respectively, which cannot be observed by the other selection methods. This means that by ADVS ~~ADMV~~ the combination of trend and seasonal components correlate best and the remaining unselected data have the lowest correlation among the methods. If these two criteria are used to separate the representative part of the data from the unrepresentative part, the ADVS ~~ADMV~~ method produces the best results.

4. Conclusions and outlook

We presented a novel statistical data selection method, the ADVS ~~ADMV~~, for CO_2 measurements at elevated GAW mountain stations. For validation and assessment of the data selection procedure, we applied the method to six CO_2 datasets measured at GAW mountain stations in the European Alps. The ADVS ~~ADMV~~ method resulted in an increasing number of selection percentages representing the background conditions with growing altitude of measurement sites, ~~percentage of data~~

25 ~~with growing altitude~~ which is reasonable due to the underlying atmospheric dynamics. Comparing ADVS ~~ADMV~~ with three other well-known statistical data selection methods, all methods yielded rather consistent characteristics across different stations. Nevertheless, among all the methods, ADVS ~~ADMV~~ is the most restrictive in terms of the number of selected data in the overall datasets.

In addition, we applied the time series decomposition tool STL to all validated and selected datasets. All statistical data

30 selection methods resulted in the same annual trend in terms of the 95% confidence interval from the validated datasets, while the seasonal signal varied substantially with smaller seasonal amplitudes and delayed occurrences of seasonal maxima. We also presented an additional assessment of the proposed new method compared with the other statistical data selection

methods based on correlation analysis. Both higher correlation coefficients of the trend and seasonal components by STL and inversely lower coefficients of the remainder indicate a better performance of ADVS ~~ADMV~~ than the other methods SI and THO.

The presented method ADVS ~~ADMV~~ is useful for data selection of atmospheric CO_2 data representative of the lower free troposphere. It requires only data from a single measurement site. It is easily adjustable to the local conditions and it runs automatically. The method can also be applied to historical datasets. The results provide evidence that the proposed ADVS ~~ADMV~~ method confers the possibility of selecting data that are representative of CO_2 concentrations of a larger area of the lower free troposphere. This is an elementary prerequisite for application of the method to a larger number of different stations and an essential step toward generalization. It directly supports the objective of GAW to extrapolate from a set of point measurements from single stations to a larger representative area or region in the lower free troposphere (WMO, 2017~~b~~). In future, there is a need to test whether such results could be used for additional tasks, such as ground calibration of satellite measurements. Finally, it would be extremely interesting to test as the next step, whether this presented method is applicable to stations ~~concept also holds~~ in other regions and on other continents. Moreover, the issue of whether and how to include coastal stations in a systematic and practically generalizable approach for selecting representative data at GAW stations will be a particular concern.

14

Table 1: Information of measured CO_2 data-sets at six GAW mountain stations.

Station (GAW ID)	Sampling elevation (a.s.l.)	Time period (yyyy.mm)	Data provider
Hohenpeissenberg (*HPB*)	984/1027/1065 m	2015.09–2016.06	DWD
Schauinsland (*SSL*)	1210 m	2010.01–2015.12	UBA-De
Izaña (*IZO*)	2403 m	2010.01–2015.12	AEMET
Zugspitze-Schneefernerhaus (*ZSF*)	2670 m	2010.01–2015.12	UBA-De
Sonnblick (*SNB*)	3111 m	2010.01–2015.12	UBA-At
Jungfraujoch (*JFJ*)	3580 m	2010.01–2015.12	Empa

Table 2: Selection percentages of selected data in ~~among~~ all data by different data selection methods. The bottom shows the linear regression coefficients of stations~~²~~ (*HPB* is represented by *HPB50*; *IZO* is excluded) altitudes and the selection percentages at the ~~a~~ significance level of 0.05 (***).

Station ID	ADVS ~~ADMV~~	SI	THO	MA
HPB50	3.2	13.9	21.7	79.8
HPB93	4.8	18.5	25.0	79.4
HPB131	6.2	21.3	27.3	79.8
SSL	4.0	17.9	25.4	83.2
IZO	36.2	82.2	56.0	60.5
ZSF	14.8	47.1	40.8	79.0
SNB	19.3	58.7	44.2	76.9
JFJ	22.1	62.1	46.3	77.6
Linear regression coefficient (γ^2)	0.996***	0.992***	0.985***	0.645

Table 3: Mean annual growth rates (ppm y $^{-1}$) with 95% confidence intervals from linear regression applied on the trend components by STL over 2010 to 2015, except for *SNB*. Data at *SNB* were decomposed over 2012 to 2015 due to missing data from 2010 to 2011 and thus shown in gray.

Station ID	VAL	SI	THO	ADVS ADMV
SSL	2.04 ± 0.09	1.89 ± 0.06	2.04 ± 0.06	2.03 ± 0.09
IZO	2.24 ± 0.03	2.26 ± 0.02	2.25 ± 0.02	2.25 ± 0.02
ZSF	2.13 ± 0.08	2.16 ± 0.05	2.17 ± 0.06	2.19 ± 0.06
SNB	2.02 ± 0.07	2.06 ± 0.06	2.06 ± 0.06	2.08 ± 0.04
JFJ	2.13 ± 0.03	2.15 ± 0.02	2.14 ± 0.02	2.14 ± 0.02

Table 4: Standard deviations of the remainder components by STL over 2010 to 2015, except for *SNB*. Data at *SNB* were decomposed over 2012 to 2015 due to missing data from 2010 to 2011 and thus shown in gray.

Station ID	VAL	SI	THO	ADVS ~~ADMV~~
SSL	1.61	1.16	1.26	1.99
IZO	0.34	0.33	0.30	0.30
ZSF	0.89	0.75	0.72	0.73
SNB	0.66	0.56	0.55	0.70
JFJ	0.56	0.45	0.48	0.47

Figure 1: Locations of six European elevated mountain stations. Symbols from left to right stand for: *IZO* **– Izaña, Spain;** *SSL* **– Schauinsland, Germany;** *JFJ* **– Jungfraujoch, Switzerland;** *HPB* **– Hohenpeissenberg, Germany;** *ZSF* **– Schneefernerhaus-Zugspitze, Germany;** *SNB* **– Sonnblick, Austria.**
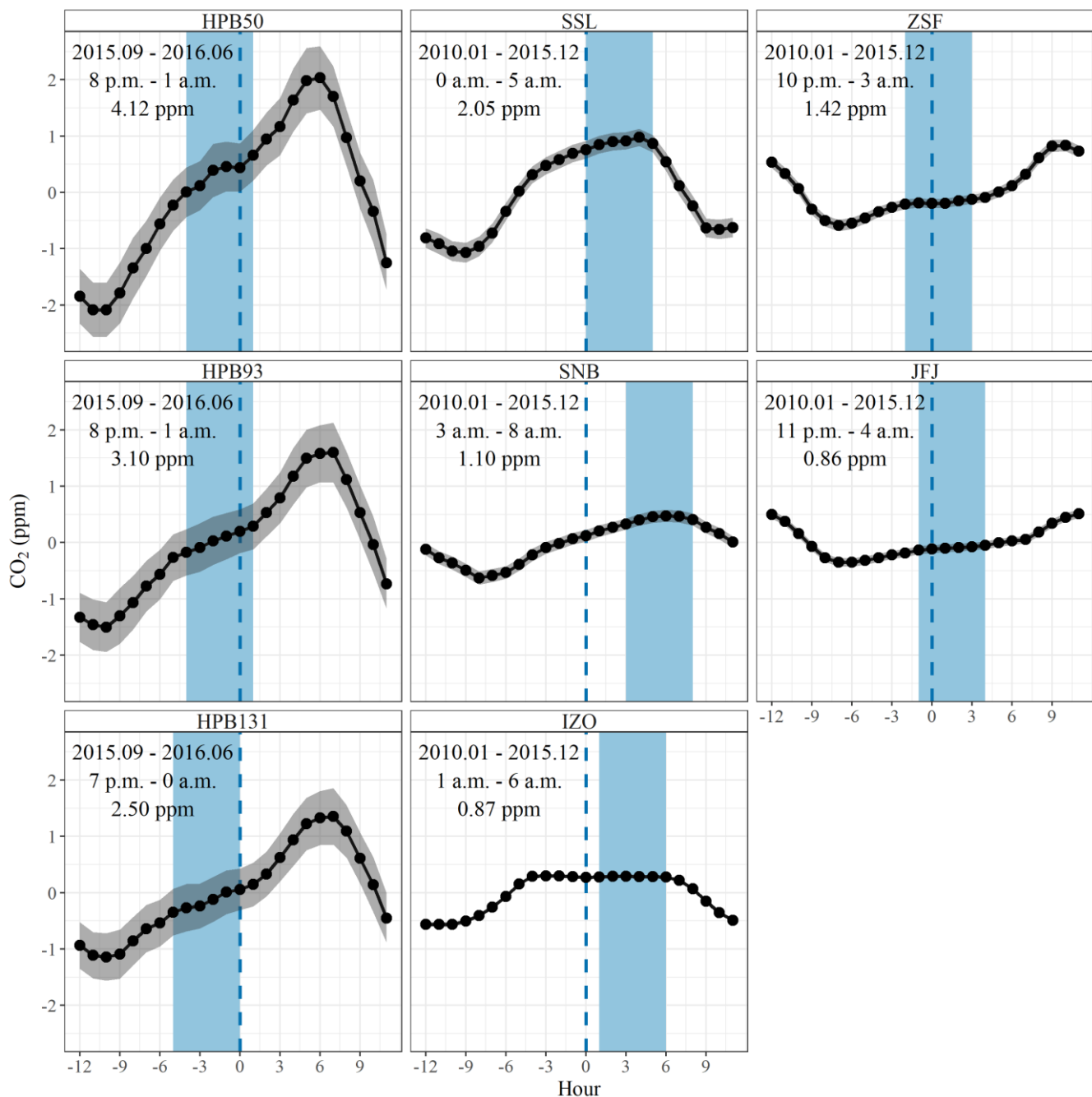
5

Figure 2: Detrended mean diurnal cycles of validated CO_2 datasets (black) with 95% confidence intervals (gray) from six GAW stations (hours in LT). Measurements at *HPB* are differentiated by the sampling heights (e.g., *HPB50* for 50 m a.g.l.). The covered time periods (top text), resulting *start time windows* (middle text, also in light blue shades), and mean diurnal amplitudes (bottom text) are shown in each subplot.
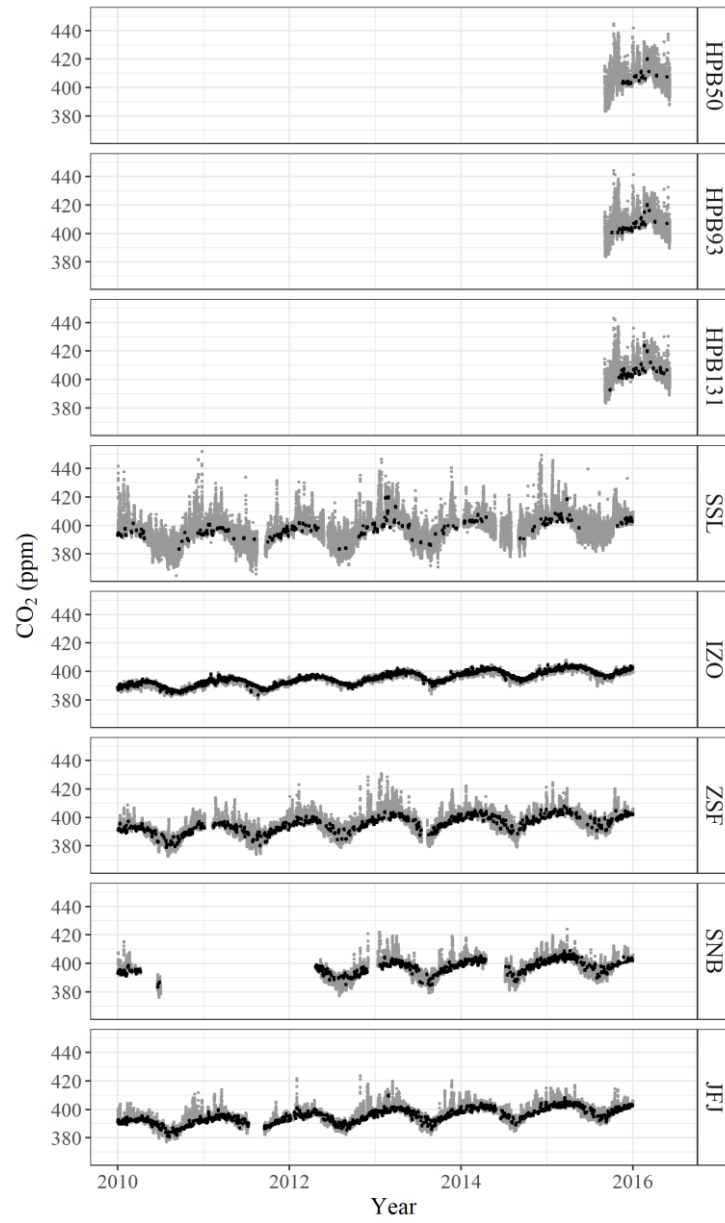
Figure 3: Time series plots of validated CO_2 datasets (gray), and selected data–sets by ADVS ~~ADMV~~ (black) at six GAW stations.
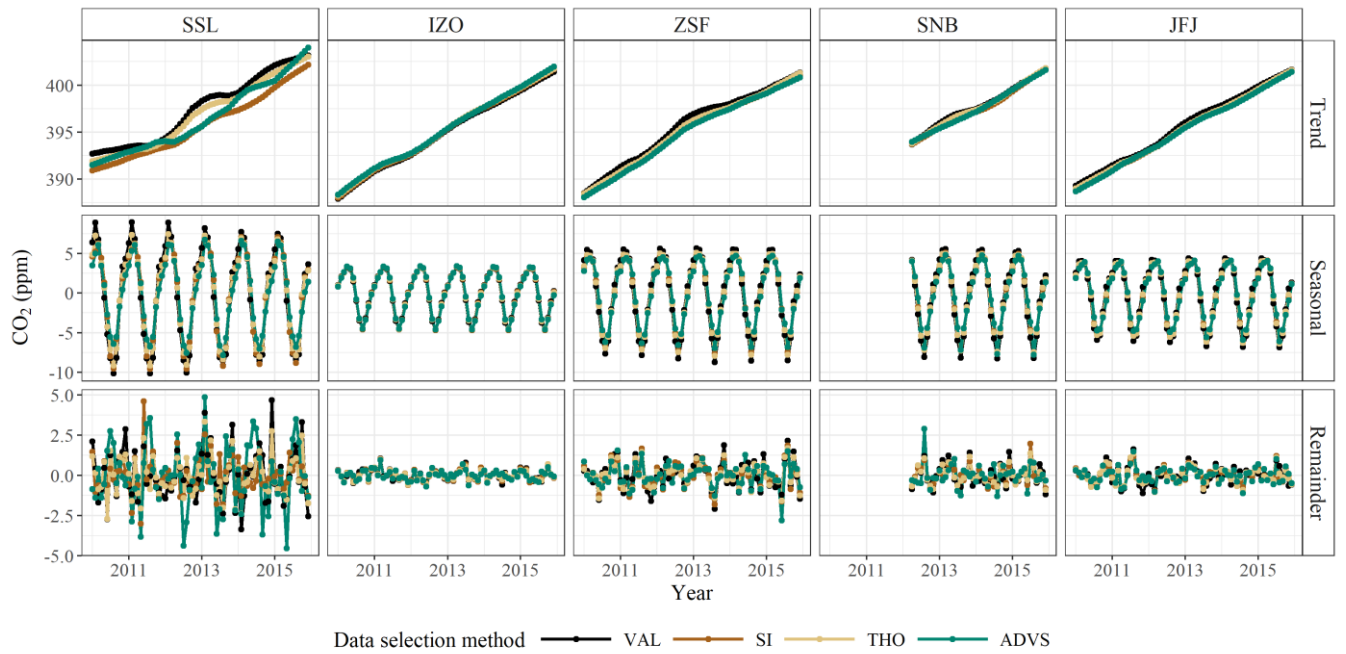
Figure 4: STL decomposition results from VAL (black), SI-selected (brown), THO-selected (yellow), and ADVS ADMV -selected (green) data-sets at five GAW stations.
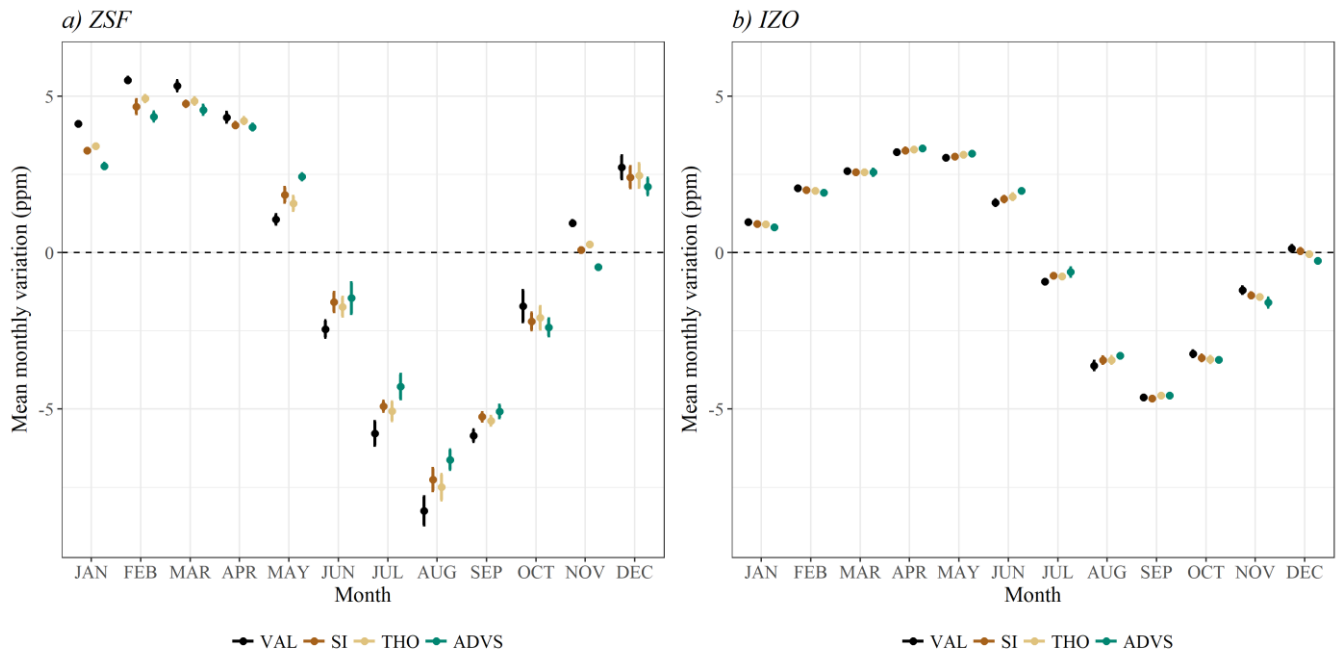
Figure 5: Mean monthly variation of the seasonal component decomposed by STL at a) *ZSF* **and b)** *IZO* **over the whole period. For better visualization of the results of selection methods, dots have been separated horizontally equidistantly. The 95% confidence intervals are shown as error bars.**
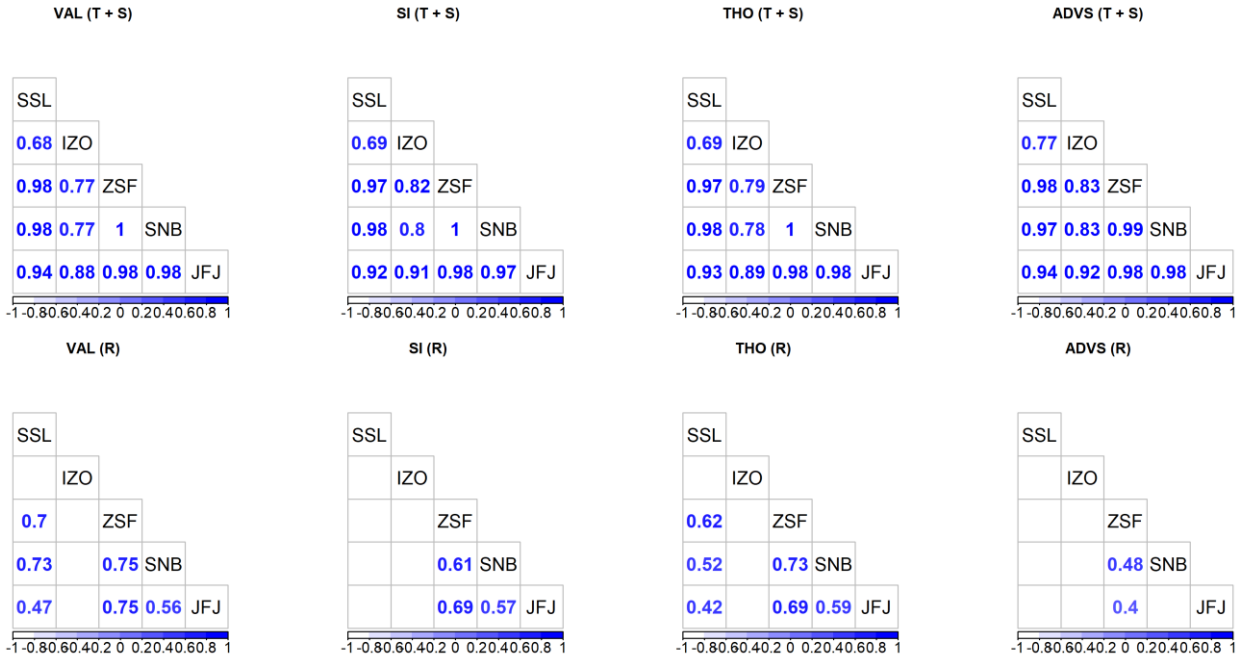
5

Figure 6: Pearson's correlation matrices of combinations of trend (T) and seasonal (S) components (upper panel), and only remainder (R) components (lower panel) at stations *SSL*, *IZO*, *ZSF*, *SNB*, and *JFJ* by different selection methods, as indicated on the top. The color scale reflects the strength of correlation. Correlations with no significant coefficient at the 0.05 confidence level were left blank.

References

Bacastow, R. B., Keeling, C. D., and Whorf, T. P.: Seasonal Amplitude Increase in Atmospheric CO_2 Concentration at Mauna Loa, Hawaii, 1959-1982, J. Geophys. Res., 90, 10529–10540, doi:10.1029/JD090iD06p10529, 1985.

Balzani Lööv, J. M., Henne, S., Legreid, G., Staehelin, J., Reimann, S., Prévôt, A. S. H., Steinbacher, M., and Vollmer, M. K.: Estimation of background concentrations of trace gases at the Swiss Alpine site Jungfraujoch (3580 m asl), J. Geophys. Res., 113, doi:10.1029/2007JD009751, 2008.

Birmili, W., Ries, L., Sohmer, R., Anastou, A., Sonntag, A., König, K., and Levin, I.: Feine und ultrafeine Aerosolpartikeln an der GAW-Station Schneefernerhaus/Zugspitze, Gefahrst. Reinhalt. L, 69, 1/2, 31–35, 2009.

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

Brooks, B. G. J., Desai, A. R., Stephens, B. B., Bowling, D. R., Burns, S. P., Watt, A. S., Heck, S. L., and Sweeney, C.: Assessing filtering of mountaintop CO_2 mole fractions for application to inverse models of biosphere-atmosphere carbon exchange, Atmospheric Chemistry and Physics, 12, 2099–2115, doi:10.5194/acp-12-2099-2012, 2012.

Calvert, J. G.: Glossary of atmospheric chemistry terms, Pure and Applied Chemistry, 62, 2167–2219, doi:10.1351/pac199062112167, 1990.

Carnuth, W., Kempfer, U., and Trickl, T.: Highlights of the tropospheric lidar studies at IFU within the TOR project, Tellus B, 54, 163–185, doi:10.1034/j.1600-0889.2002.00245.x, 2002.

Carnuth, W. and Trickl, T.: Transport studies with the IFU three-wavelength aerosol lidar during the VOTALP Mesolcina experiment, Atmospheric Environment, 34, 1425–1434, doi:10.1016/S1352-2310(99)00423-9, 2000.

Carslaw, D. C.: On the changing seasonal cycles and trends of ozone at Mace Head, Ireland, Atmos. Chem. Phys., 5, 3441–3450, doi:10.5194/acp-5-3441-2005, 2005.

Chambers, S. D., Williams, A. G., Conen, F., Griffiths, A. D., Reimann, S., Steinbacher, M., Krummel, P. B., Steele, L. P., van der Schoot, M. V., Galbally, I. E., Molloy, S. B., and Barnes, J. E.: Towards a Universal "Baseline" Characterisation of Air Masses for High- and Low-Altitude Observing Stations Using Radon-222, Aerosol Air Qual. Res., 16, 885–899, doi:10.4209/aaqr.2015.06.0391, 2016.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A seasonal-trend decomposition procedure based on Loess, 6, 3–73, 1990.

Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association, 74, 829–836, doi:10.1080/01621459.1979.10481038, 1979.

Cleveland, W. S., Freeny, A. E., and Graedel, T. E.: The Seasonal Component of Atmospheric CO_2: Information From New Approaches to the Decomposition of Seasonal Time Series, J. Geophys. Res., 88, 10934–10946, doi:10.1029/JC088iC15p10934, 1983.

Cui, J., Pandey Deolal, S., Sprenger, M., Henne, S., Staehelin, J., Steinbacher, M., and Nédélec, P.: Free tropospheric ozone changes over Europe as observed at Jungfraujoch (1990–2008): An analysis based on backward trajectories, J. Geophys. Res., 116, doi:10.1029/2010JD015154, 2011.

Elliott, W. P. (Ed.): The Statistical treatment of CO_2 data records, NOAA Technical Memorandum ERL ARL, 173, U.S.
5   Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Silver Spring, Md., ix, 131, 1989.

Fang, S.-X., Tans, P. P., Steinbacher, M., Zhou, L.-X., and Luan, T.: Comparison of the regional CO_2 mole fraction filtering approaches at a WMO/GAW regional station in China, Atmospheric Measurement Techniques, 8, 5301–5313, doi:10.5194/amt-8-5301-2015, 2015.

10   Forrer, J., Rüttimann, R., Schneiter, D., Fischer, A., Buchmann, B., and Hofer, P.: Variability of trace gases at the high-Alpine site Jungfraujoch caused by meteorological transport processes, J. Geophys. Res., 105, 12241–12251, doi:10.1029/1999JD901178, 2000.

Gilge, S., Plass-Duelmer, C., Fricke, W., Kaiser, A., Ries, L., Buchmann, B., and Steinbacher, M.: Ozone, carbon monoxide and nitrogen oxides time series at four alpine GAW mountain stations in central Europe, Atmos. Chem. Phys., 10,
15   12295–12316, doi:10.5194/acp-10-12295-2010, 2010.

Gomez-Pelaez, A. J., Ramos, R., Cuevas, E., and Gomez-Trueba, V.: 25 years of continuous CO_2 and CH_4 measurements at Izaña Global GAW mountain station: annual cycles and interannual trends, in: Proceedings of the Symposium on Atmospheric Chemistry and Physics at Mountain Sites (ACP Symposium 2010), Interlaken, Switzerland, June 8-10, 2010, 157–159, 2010.

20   Gomez-Pelaez, A. J., Ramos, R., Gomez-Trueba, V., Novelli, P. C., and Campo-Hernandez, R.: A statistical approach to quantify uncertainty in carbon monoxide measurements at the Izaña global GAW station: 2008–2011, Atmos. Meas. Tech., 6, 787–799, doi:10.5194/amt-6-787-2013, 2013.

Henne, S., Brunner, D., Folini, D., Solberg, S., Klausen, J., and Buchmann, B.: Assessment of parameters describing representativeness of air quality in-situ measurement sites, Atmos. Chem. Phys., 10, 3561–3581, doi:10.5194/acp-10-
25   3561-2010, 2010.

Herrmann, E., Weingartner, E., Henne, S., Vuilleumier, L., Bukowiecki, N., Steinbacher, Coen, F., Collaud Conen, M., Hammer, E., Jurányi, Z., Baltensperger, U., and Gysel, M.: Analysis of long-term aerosol size distribution data from Jungfraujoch with emphasis on free tropospheric conditions, cloud influence, and air mass transport, J. Geophys. Res., 120, doi 10.1002/2015JD023660, 2015.

30   Hernández-Paniagua, I. Y., Lowry, D., Clemitshaw, K. C., Fisher, R. E., France, J. L., Lanoisellé, M., Ramonet, M., and Nisbet, E. G.: Diurnal, seasonal, and annual trends in atmospheric CO_2 at southwest London during 2000–2012: Wind sector analysis and comparison with Mace Head, Ireland, Atmospheric Environment, 105, 138–147, doi:10.1016/j.atmosenv.2015.01.021, 2015.

Kaiser, A., Scheifinger, H., Spangl, W., Weiss, A., Gilge, S., Fricke, W., Ries, L., Cemas, D., and Jesenovec, B.: Transport of nitrogen oxides, carbon monoxide and ozone to the Alpine Global Atmosphere Watch stations Jungfraujoch (Switzerland), Zugspitze and Hohenpeissenberg (Germany), Sonnblick (Austria) and Mt. Krvavec (Slovenia), Atmospheric Environment, 41, 9273–9287, doi:10.1016/j.atmosenv.2007.09.027, 2007.

5 Keeling, C. D.: The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere, Tellus, 12, 200–203, doi:10.1111/j.2153-3490.1960.tb01300.x, 1960.

Locher, R. and Ruckstuhl, A.: IDPmisc: Utilities of Institute of Data Analyses and Process Design (www.idp.zhaw.ch): https://CRAN.R-project.org/package=IDPmisc, 2012.

Lowe, D. C., Guenther, P. R., and Keeling, C. D.: The concentration of atmospheric carbon dioxide at Baring Head, New
10 Zealand, Tellus, 31, 58–67, doi:10.1111/j.2153-3490.1979.tb00882.x, 1979.

Massen, F. and Beck, E.-G.: Accurate Estimation of CO_2 Background Level from Near Ground Measurements at Non-Mixed Environments, in: The Economic, Social and Political Elements of Climate Change, Leal Filho, W. (Ed.), Climate Change Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 509–522, 2011.

Pales, J. C. and Keeling, C. D.: The Concentration of Atmospheric Carbon Dioxide in Hawaii, J. Geophys. Res., 70, 6053–
15 6076, doi:10.1029/JZ070i024p06053, 1965.

Parrish, D. D., Law, K. S., Staehelin, J., Derwent, R., Cooper, O. R., Tanimoto, H., Volz-Thomas, A., Gilge, S., Scheel, H.-E., Steinbacher, M., and Chan, E.: Long-term changes in lower tropospheric baseline ozone concentrations at northern mid-latitudes, Atmos. Chem. Phys., 12, 11485–11504, doi:10.5194/acp-12-11485-2012, 2012.

Peterson, J. T., Komhyr, W. D., Harris, T. B., and Waterman, L. S.: Atmospheric carbon dioxide measurements at Barrow,
20 Alaska, 1973-1979, Tellus, 34, 166–175, doi:10.1111/j.2153-3490.1982.tb01804.x, 1982.

Pickers, P. A. and Manning, A. C.: Investigating bias in the application of curve fitting programs to atmospheric time series, Atmos. Meas. Tech., 8, 1469–1489, doi:10.5194/amt-8-1469-2015, 2015.

R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria: https://www.R-project.org/, 2017.

25 Reiter, R., Sladkovic, R., Kanter, H.-J., 1986. Concentration of trace gases in the lower troposphere, simultaneously recorded at neighboring mountain stations. Meteorl. Atmos. Phys. 35 (4), 187–200. 10.1007/BF01041811.

Risius, S., Xu, H., Di Lorenzo, F., Xi, H., Siebert, H., Shaw, R. A., and Bodenschatz, E.: Schneefernerhaus as a mountain research station for clouds and turbulence, Atmos. Meas. Tech., 8, 3209–3218, doi:10.5194/amt-8-3209-2015, 2015.

Ruckstuhl, A. F., Henne, S., Reimann, S., Steinbacher, M., Vollmer, M. K., O'Doherty, S., Buchmann, B., and Hueglin, C.:
30 Robust extraction of baseline signal of atmospheric trace species using local regression, Atmos. Meas. Tech., 5, 2613–2624, doi:10.5194/amt-5-2613-2012, 2012.

Satar, E., Berhanu, T. A., Brunner, D., Henne, S., and Leuenberger, M.: Continuous CO_2/CH_4/CO measurements (2012–2014) at Beromünster tall tower station in Switzerland, Biogeosciences, 13, 2623–2635, doi:10.5194/bg-13-2623-2016, 2016.

Schibig, M. F., Steinbacher, M., Buchmann, B., van der Laan-Luijkx, I. T., van der Laan, S., Ranjan, S., and Leuenberger, M. C.: Comparison of continuous in situ CO_2 observations at Jungfraujoch using two different measurement techniques, Atmos. Meas. Tech., 8, 57–68, doi:10.5194/amt-8-57-2015, 2015.

Schmidt, M., Graul, R., Sartorius, H., and Levin, I.: The Schauinsland CO_2 record: 30 years of continental observations and
5   their implications for the variability of the European CO_2 budget, Journal of Geophysical Research D: Atmospheres, 108, ACL 14-1 - 14-7, 2003.

Sepúlveda, E., Schneider, M., Hase, F., Barthlott, S., Dubravica, D., García, O. E., Gomez-Pelaez, A., González, Y., Guerra, J. C., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Strong, K., Weaver, D., Palm, M., Sadeghi, A., Deutscher, N. M., Warneke, T., Notholt, J., Jones, N., Griffith, D. W. T., Smale, D., Brailsford, G. W., Robinson, J., Meinhardt, F.,
10   Steinbacher, M., Aalto, T., and Worthy, D.: Tropospheric CH_4 signals as observed by NDACC FTIR at globally distributed sites and comparison to GAW surface in situ measurements, Atmos. Meas. Tech., 7, 2337–2360, doi:10.5194/amt-7-2337-2014, 2014.

Sirignano, C., Neubert, R. E., Rödenbeck, C., and J. Meijer, H. A.: Atmospheric oxygen and carbon dioxide observations from two European coastal stations 2000-2005: continental influence, trend changes and APO climatology, Atmospheric
15   Chemistry and Physics, 10, 1599–1615, 2010.

Stephens, B. B., Brailsford, G. W., Gomez, A. J., Riedel, K., Mikaloff Fletcher, S. E., Nichol, S., and Manning, M.: Analysis of a 39-year continuous atmospheric CO_2 record from Baring Head, New Zealand, Biogeosciences, 10, 2683–2697, doi:10.5194/bg-10-2683-2013, 2013.

Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric Carbon Dioxide at Mauna Loa Observatory: 2. Analysis of
20   the NOAA GMCC Data, 1974-1985, J. Geophys. Res., 94, 8549–8565, doi:10.1029/JD094iD06p08549, 1989.

Uglietti, C., Leuenberger, M., and Brunner, D.: European source and sink areas of CO_2 retrieved from Lagrangian transport model interpretation of combined O_2 and CO_2 measurements at the high alpine research station Jungfraujoch, Atmospheric Chemistry and Physics, 11, 8017–8036, doi:10.5194/acp-11-8017-2011, 2011.

WMO: WMO Greenhouse Gas Bulletin, No. 13, ISSN 2078-0796, 30 October, 2017a.
25 WMO: WMO Global Atmosphere Watch (GAW) Implementation Plan: 2016-2023, 2017b.

Zeileis, A. and Grothendieck, G.: zoo: S3 Infrastructure for Regular and Irregular Time Series, J. Stat. Soft., 14, doi:10.18637/jss.v014.i06, 2005.

Zellweger, C., Forrer, J., Hofer, P., Nyeki, S., Schwarzenbach, B., Weingartner, E., Ammann, M., and Baltensperger, U.: Partitioning of reactive nitrogen (NO_y) and dependence on meteorological conditions in the lower free troposphere,
30   Atmos. Chem. Phys., 3, 779–796, doi:10.5194/acp-3-779-2003, 2003.