

Supplement of Atmos. Meas. Tech., 11, 1501–1514, 2018
<https://doi.org/10.5194/amt-11-1501-2018-supplement>
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.



Supplement of

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

Ye Yuan et al.

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

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

S1 Applicability of ADVS

S1.1 Running frequency

We compared the ADVS derived *start time windows* of overall running frequency with seasonal running frequency at all measurement sites except for *HPB*, shown in Table S1.1.

Table S1.1: *Start time windows* by overall frequency and seasonal frequency (data of whole period), hours in advance and delayed are listed in brackets.

Frequency	<i>SSL</i>	<i>IZO</i>	<i>ZSF</i>	<i>SNB</i>	<i>JFJ</i>
Overall	0 a.m. – 5 a.m.	1 a.m. – 6 a.m.	10 p.m. – 3 a.m.	3 a.m. – 8 a.m.	11 p.m. – 4 a.m.
Winter	0 a.m. – 5 a.m. (0 hr)	2 a.m. – 7 a.m. (+1 hr)	1 a.m. – 6 a.m. (+3 hr)	6 p.m. – 11 p.m. (-9 hr)	3 a.m. – 8 a.m. (+4 hr)
Spring	0 a.m. – 5 a.m. (0 hr)	11 p.m. – 4 a.m. (-2 hr)	10 p.m. – 3 a.m. (0 hr)	5 a.m. – 10 a.m. (+2 hr)	6 p.m. – 11 p.m. (-5 hr)
Summer	0 a.m. – 5 a.m. (0 hr)	1 a.m. – 6 a.m. (0 hr)	0 a.m. – 5 a.m. (+2 hr)	3 a.m. – 8 a.m. (0 hr)	1 a.m. – 6 a.m. (+2 hr)
Autumn	1 a.m. – 6 a.m. (+1 hr)	1 a.m. – 6 a.m. (0 hr)	9 p.m. – 2 a.m. (-1 hr)	5 a.m. – 10 a.m. (+2 hr)	0 a.m. – 5 a.m. (+1 hr)

As a result, most of the seasonal derived *start time windows* differ moderately from the overall derived *start time windows*. However, two exceptions are observed, which are winter at *SNB* and spring at *JFJ*. Therefore, graphical examination was taken in Fig. S1.1 (a) and (b).

For winter time at *SNB*, the seasonal derived *start time window* is with the least standard deviation, but doesn't exhibit the minimal value from the diurnal cycle. For winter, we expect the most representative level of CO₂ to be around the minimal values with relatively small standard deviations. Therefore, the overall derived *start time window* is more suitable to be the ideal *start time window* in this case.

Similar case happens to the spring time at *JFJ*. For spring, the vegetation activities have been already influenced the CO₂ diurnal cycle. The *start time window* in this case is not suitable as the minimal values included. Again, the overall derived *start time window* is more practical.

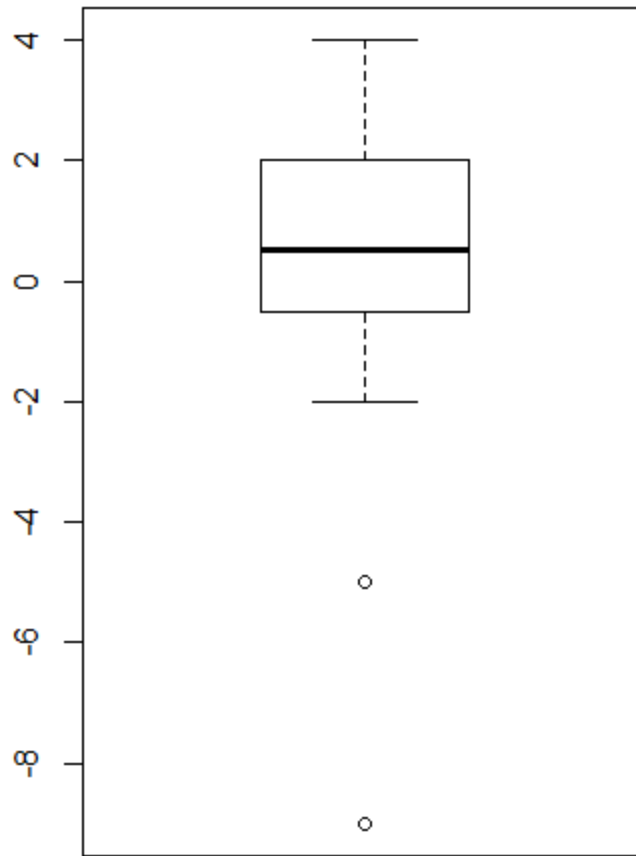


Figure S1.1 (a): Boxplot for hourly shifts of seasonal derived *start time windows* from overall derived *start time windows*, corresponding to Table S1.1.

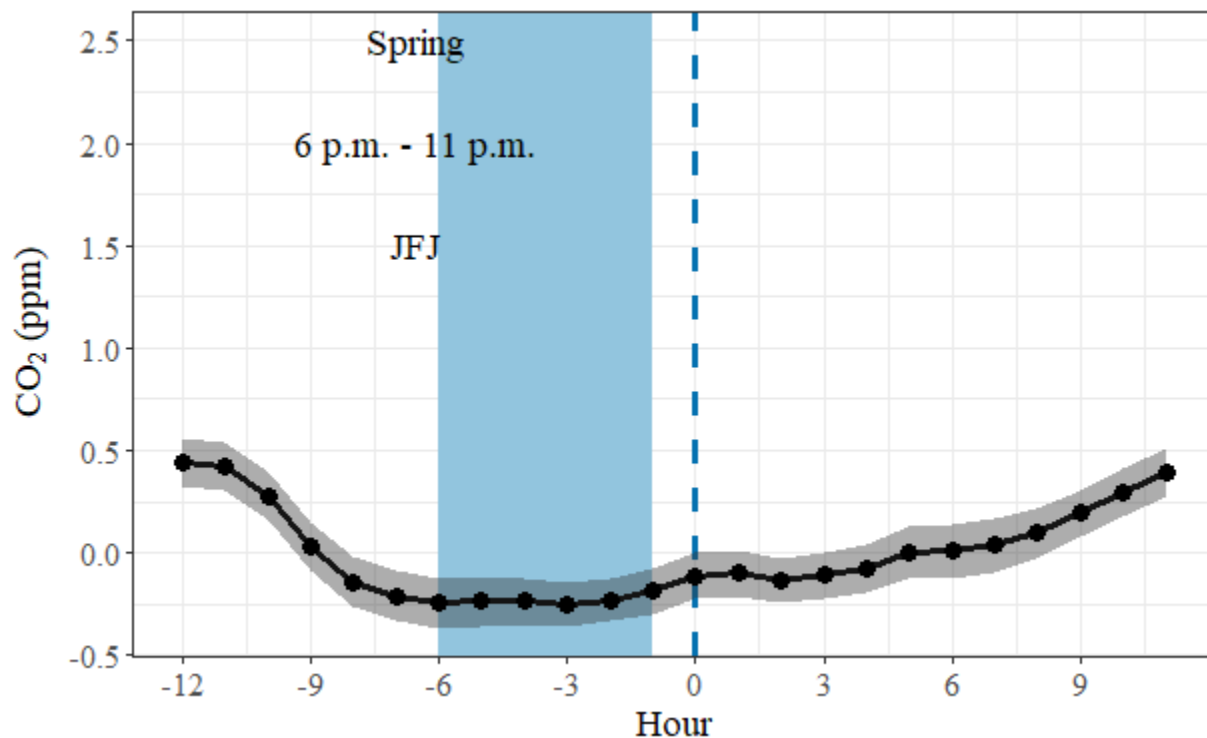
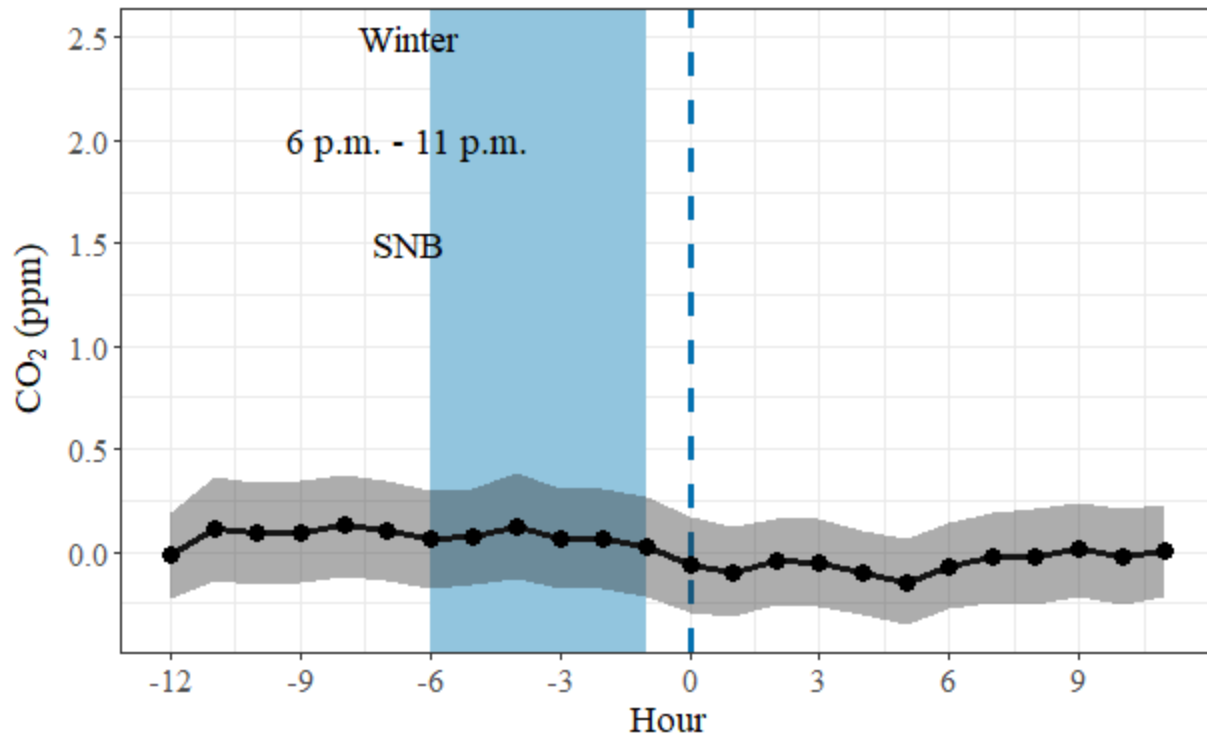


Figure S1.1 (b): Detrended mean diurnal variations of validated CO₂ mole fractions (black) with 95% confidence intervals (grey) for winter time at SNB (upper) and for spring time at JFJ (below). The corresponding seasonal derived *start time windows* are shown both in blue shades and texts.

S1.2 Standard deviation threshold $s_{i,threshold}$

The threshold of standard deviation determines the degree of variation for the selection time windows. In the study, we applied 0.3 ppm to all the stations for inter-comparison. But it is clear that for low altitude stations like *SSL*, this value may be not appropriate. We tried 1.0 ppm threshold for station *SSL* and resulted in the same *start time window* and a much higher percentage of selected data (26.51% compared with 4.0% in the main text), however with a few irregular spikes observed below.

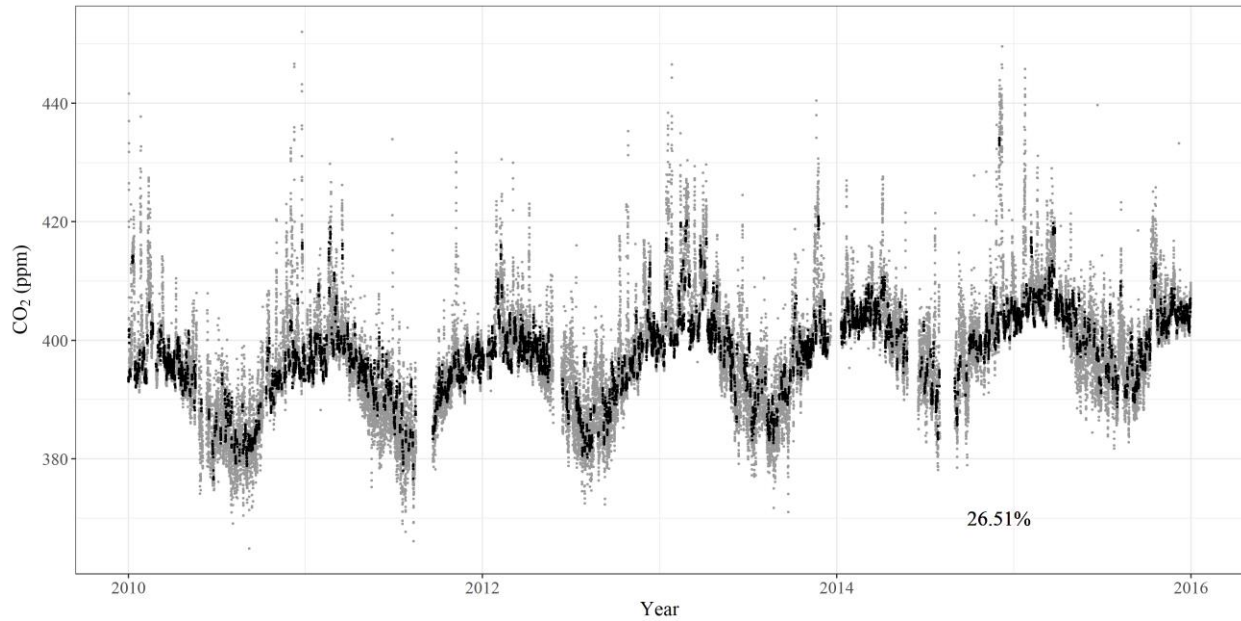


Figure S1.2: Time series plot of validated CO₂ dataset (grey) and ADVS-selected dataset (black) at *SSL* from 2010 to 2015.

S1.3 Time resolution tr

Data at station *ZSF* were taken to evaluate the differences in data selection on different time resolution. To compare with the selected result based on hourly averages, we applied ADVS directly to the 30-min validated dataset. The resulting *start time windows* are the same, but the percentages of selected data are significantly different at 95% confidence interval, shown in table below.

Table S1.3 (a): Percentages of selected data by ADVS applied to ZSF datasets.

Time resolution	Percentage of selected data \pm 95% confidence interval
30-min	13.46 ± 0.22
1-hour	14.76 ± 0.33

For station *JFJ*, the 10-min validated dataset was available for more detailed comparison. We prepared the ADVS-selected datasets on 10-min, 20-min and 30-min time scale to compare with the hourly selected results. The resulting *start time windows* are again the same. The resulted percentages of selected data show again significant differences (at 95% confidence interval) between each two datasets, shown in table below.

Table S1.3 (b): Percentages of selected data by ADVS applied to JFJ datasets.

Time resolution	Percentage of selected data \pm 95% confidence interval
10-min	18.73 ± 0.14
20-min	20.02 ± 0.21
30-min	20.75 ± 0.26
1-hour	22.14 ± 0.37

As a result, these percentages clearly indicate a significantly smaller percentage of selected data with finer time resolution. This can be possibly explained based on the statistical property of ADVS. While ADVS evaluates the standard deviation within a time window, averaged data with larger time intervals show more statistical robustness. For ADVS, it is applicable for datasets with time resolution equal to or finer than one hour.

S2 Detrended diurnal cycles

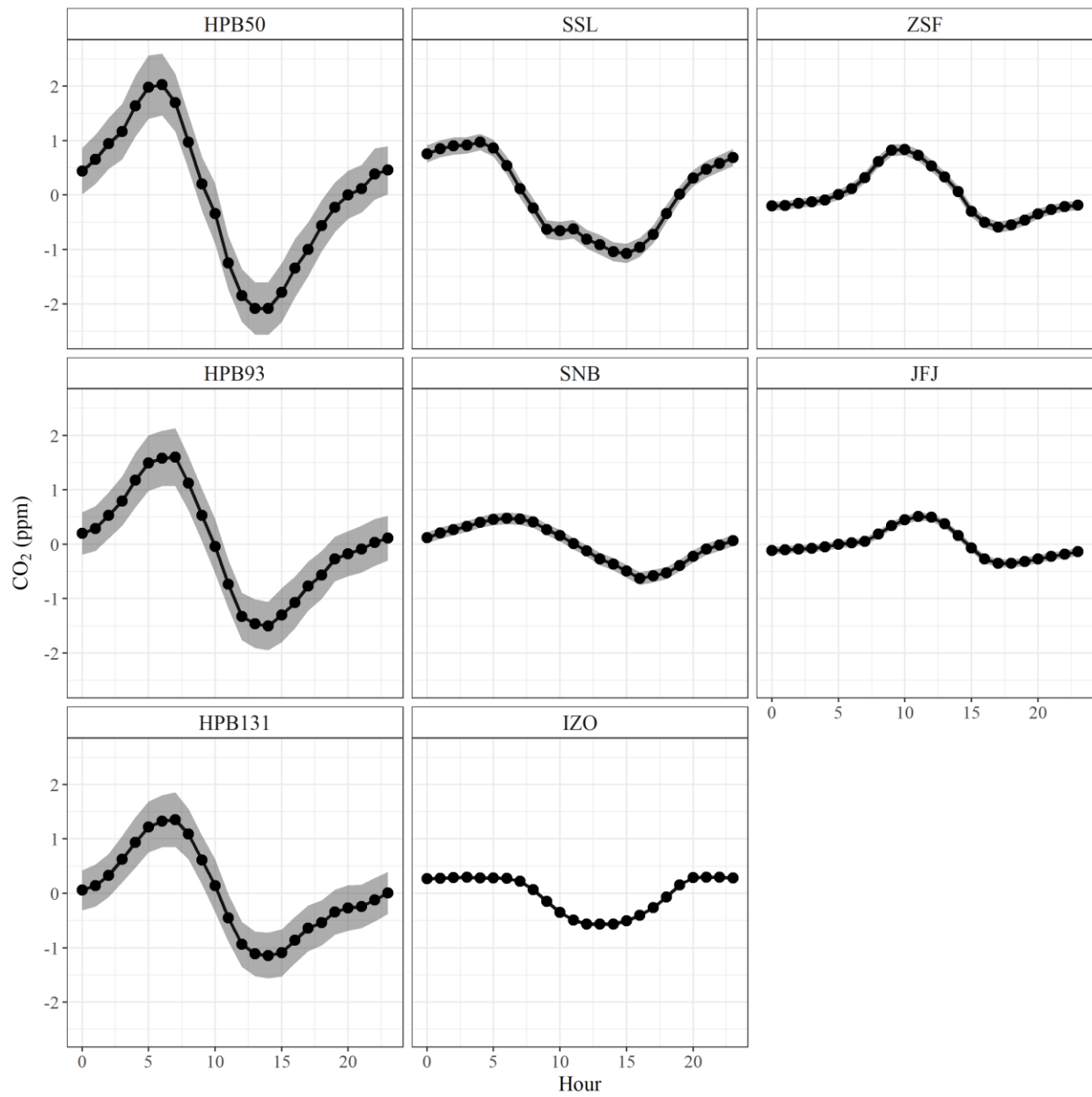


Figure S2: Detrended mean diurnal variation of validated CO₂ mole fractions (black) with 95% confidence intervals (gray) at six European GAW stations.

S3 Percentage of selected data

S3.1 Linear regression of station altitudes and percentages of selected data

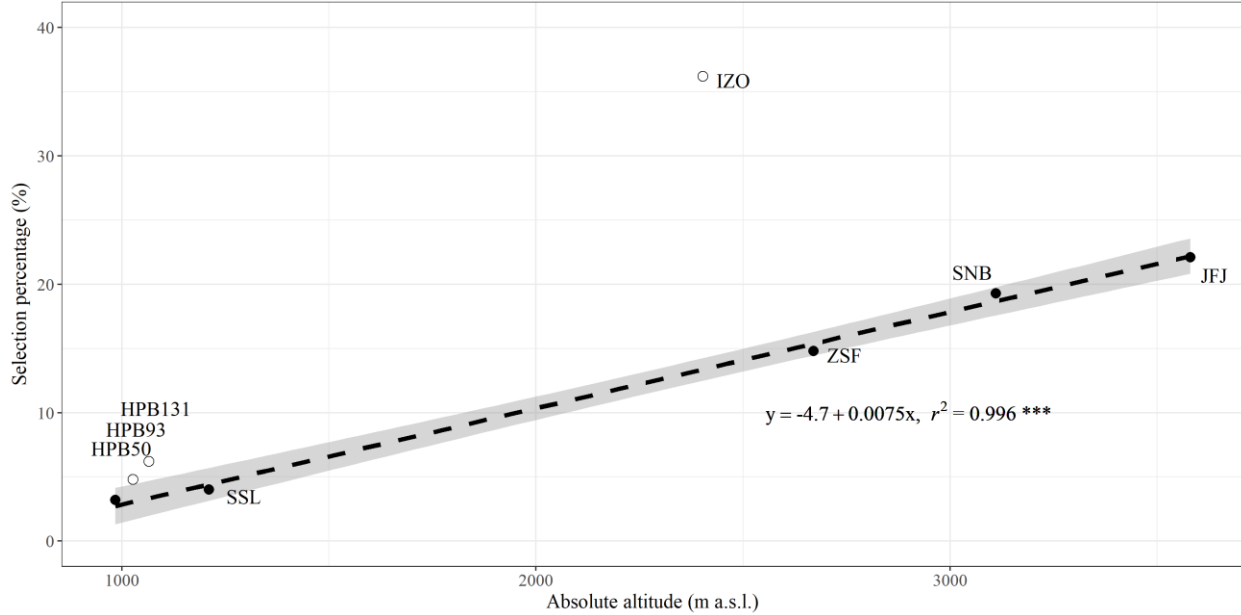


Figure S3.2: Linear regression between the absolute altitudes and the final percentages of selected data by ADVS for all continental sites (excluding *IZO*). For *HPB*, only *HPB50* is chosen as the demonstration sampling height.

S3.2 Table of percentages of selected data

Table S3.2: List of percentages of selected data (%) during ADVS data selection process (π_1 – selected days with valid *start time window* in all days; π_2 – selected hours with valid *start time window* in all hourly data; π_3 – selected hours after *forward adaptive selection* in all hourly data; ADVS – final percentages of selected data). The bottom shows the linear regression coefficients of stations (*HPB* is represented by *HPB50*; *IZO* is excluded) altitudes and the percentages of selected data at significance level of 0.05 (***).

Station ID	π_1	π_2	π_3	ADVS*
<i>HPB50</i>	15.2	2.1	2.6	3.2
<i>HPB93</i>	22.7	3.2	3.9	4.8
<i>HPB131</i>	29.6	4.3	5.2	6.2
<i>SSL</i>	14.3	2.6	3.2	4.0
<i>IZO</i>	85.2	20.0	26.1	36.2
<i>ZSF</i>	52.8	8.9	12.3	14.8
<i>SNB</i>	47.3	10.9	14.9	19.3
<i>JFJ</i>	69.3	12.1	17.4	22.1
Linear regression coefficient (γ^2)	0.941***	0.996***	0.998***	0.996***

*For ADVS, the final percentage of selected data is equivalent to the percentage of selected hours after backward adaptive selection in all hourly data.

S3.3 Comparison of percentages of selected data among data selection methods

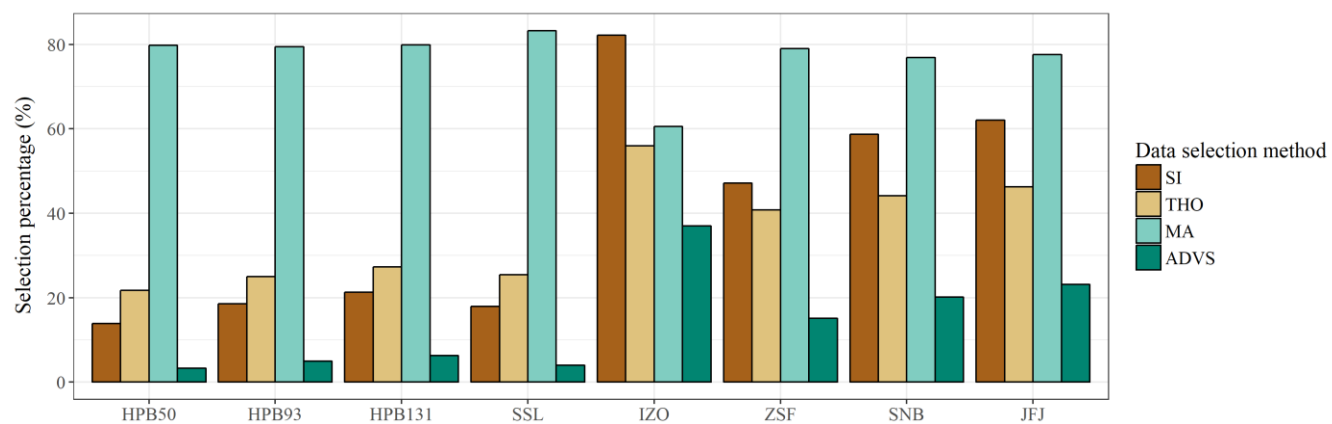


Figure S3.3: Comparison of the percentages of selected data by four statistical data selection methods applied to validated CO₂ datasets at six GAW stations.

S4 Mean monthly variations at *SSL*, *SNB* and *JFJ*

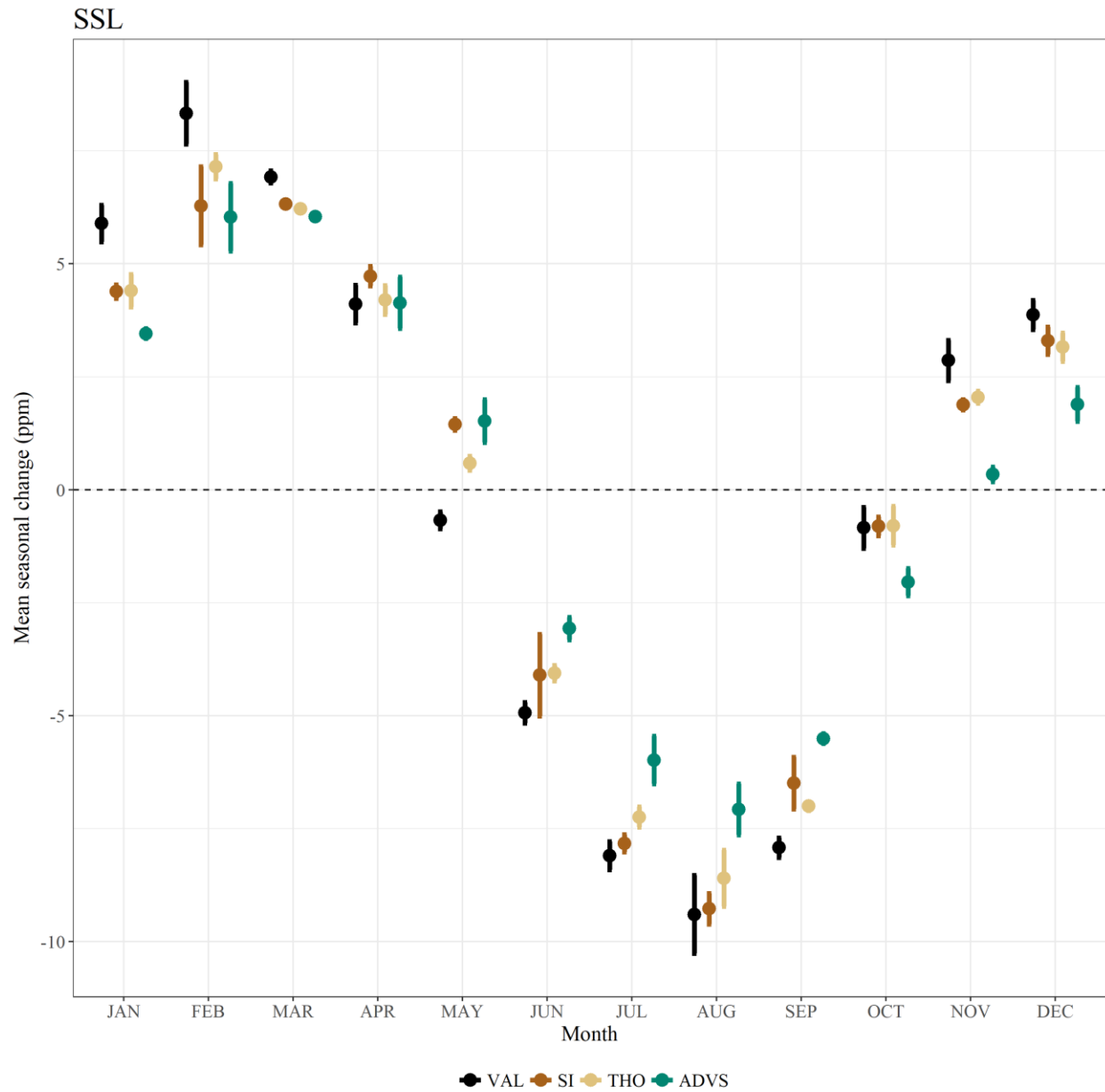


Figure S4.1: Mean monthly variation of the seasonal component decomposed by STL at *SSL* over the whole time period.

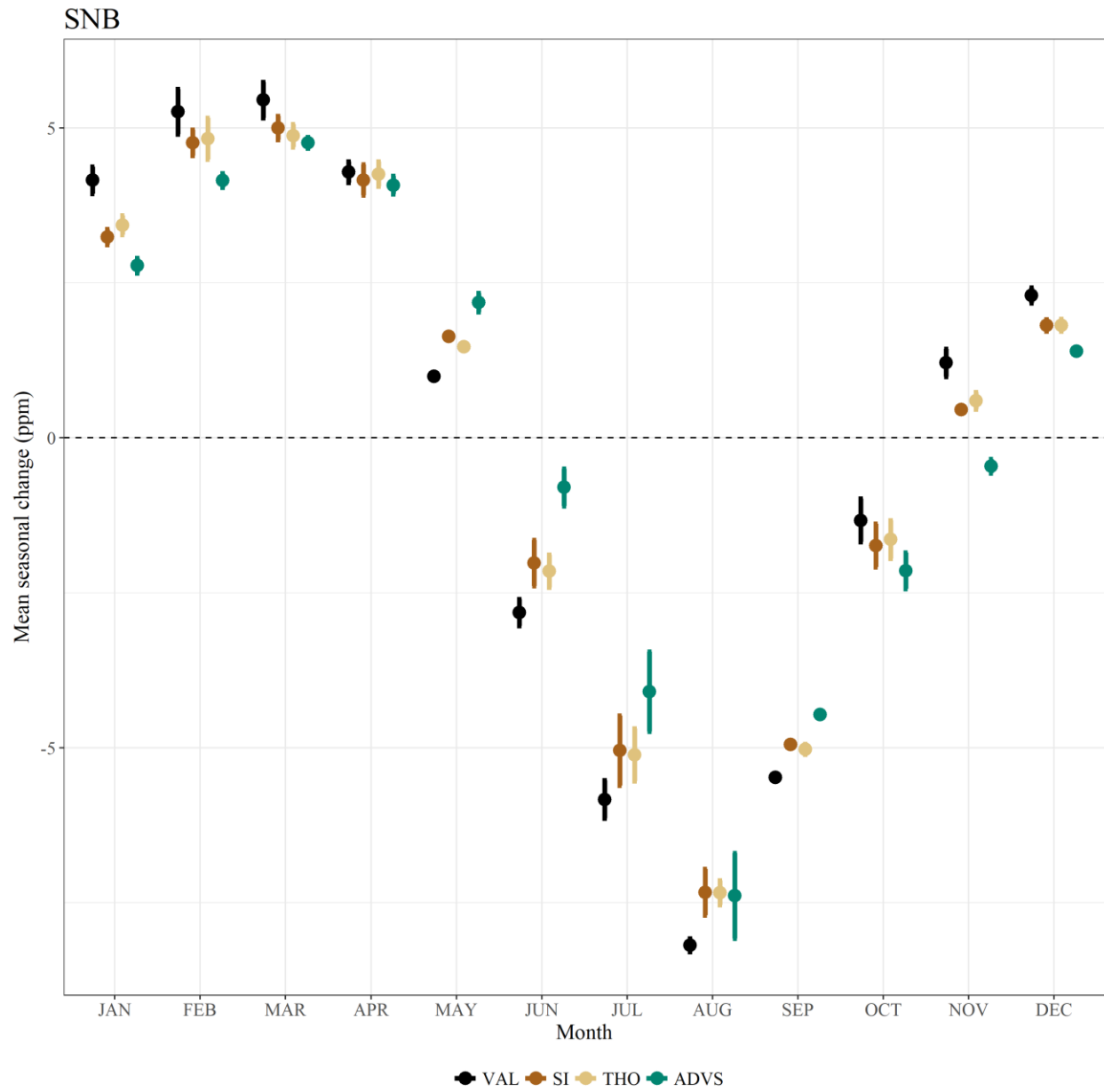


Figure S4.2: Mean monthly variation of the seasonal component decomposed by STL at *SNB* over the whole time period.

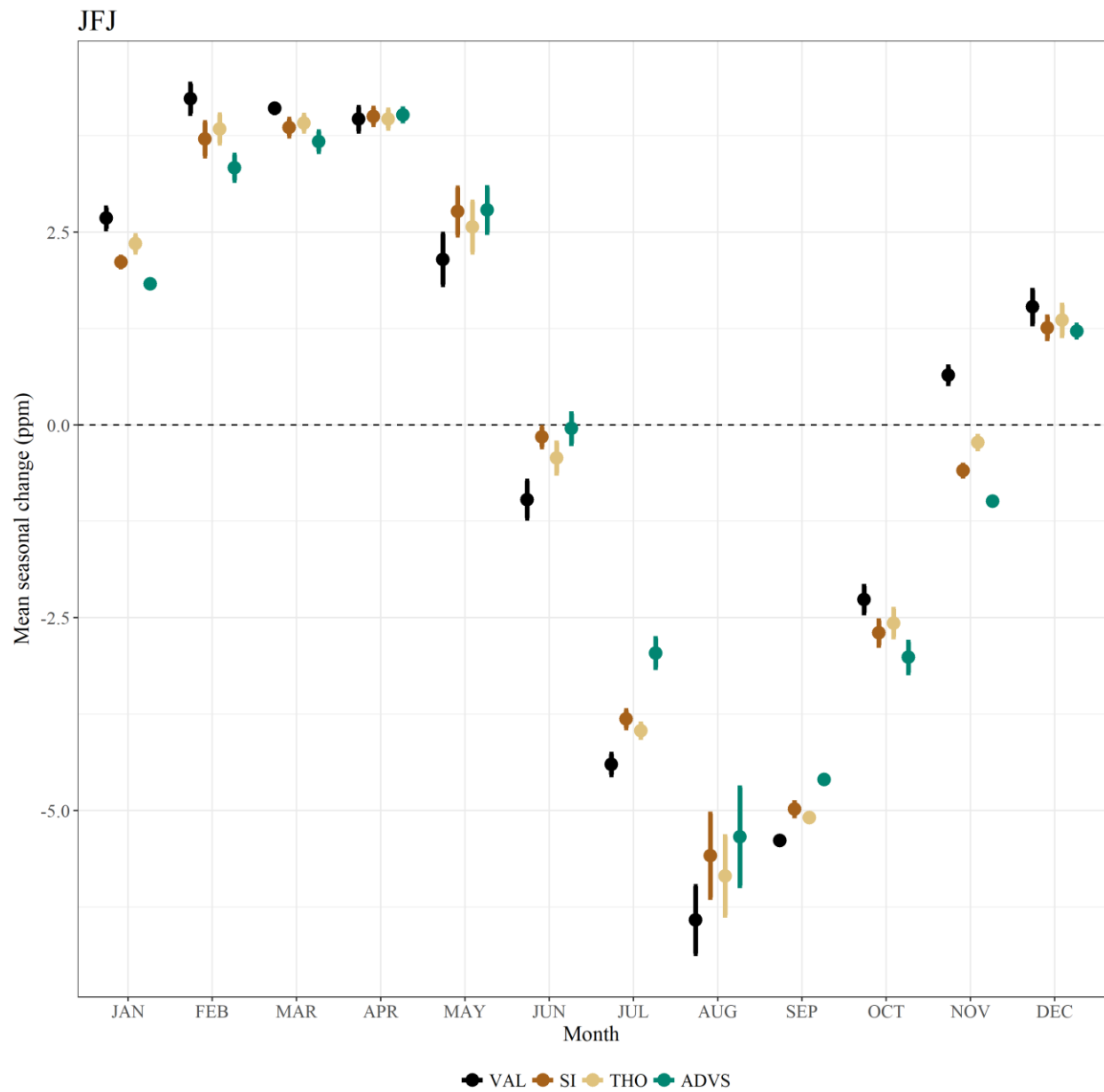


Figure S4.3: Mean monthly variation of the seasonal component decomposed by STL at *JFJ* over the whole time period.