Evaluation of two algorithms for the automatic identification of spikes on continuous atmospheric observations of CO_2, CH_4 and CO

P. Cristofanelli[1], C. Fratticioli[1], L. Hazan[2], P. Trisolino[1], M. Chariot[2], O. Gazetas[3], D. Kubistin[4], A. Laitinen[5], A. Leskinen[8,9], T. Laurila[5], M. Lindauer[4], G. Manca[6], M. Ramonet[2], M. Steinbacher[7]

[1] Institute of Atmospheric Sciences and Climate, National Research Council of Italy, Bologna, Italy; [2] Laboratoire des Sciences du Climat et de l'Environnement (LSCE), IPSL, CEA-CNRSUVSQ, Université Paris-Saclay, Gif-sur-Yvette, France; [3] Scottish Universities Environmental Research Centre (SUERC), East Kilbride,United Kingdom; [4] Deutscher Wetterdienst, Hohenpeißenberg Meteorological Observatory, Hohenpeißenberg, Germany; [5] Finnish Meteorological Institute, Helsinki, Finland; [6] European Commission Joint Research Centre, Ispra, Italy; [7] Swiss Federal Laboratories for Materials Science and Technology (Empa), Dübendorf, Switzerland; [8]University of Eastern Finland, Kuopio, Finland; [9]Finnish Meteorological Institute, Kuopio, Finland

Contact: *p.cristofanelli@isac.cnr.it*

Introduction

Within the European Integrated Carbon Observation System (ICOS) research infrastructure (www.icos-ri.eu), a network composed of 36 labelled atmospheric stations is providing preliminary calibrated and automatically screened data in near-real time (i.e. with a 24-hour delay) and final fully quality controlled data of the atmospheric mixing ratios of greenhouse gases (CO_2, CH_4, N2O) and CO. In this framework, an automated procedure to detect the occurrence of "spikes", i.e. sudden and short-lasting increases in GHG mixing ratios due to very local emissions, is required to separate local influence from regional and global signals.

To evaluate different approaches, a working group including station Principal Investigators (PIs) and the ICOS Atmospheric Thematic Center (ATC) has been launched within the ICOS Atmosphere Monitoring Station Assembly (MSA) to assess the effectiveness of two widely used spike detection algorithms ("Standard Deviation - SD" and "Robust Extraction of Baseline Signal – REBS" methods, see El Yazidi et al., 2018).

Detection algorithms (SD and REBS)

Standard Deviation (SD)

$$C_i \geq C_{unf} + \alpha \times \sigma + \sqrt{n} \times \sigma$$

C_{unf} is the first data point not selected as spike. C_i is the next data point evaluated against C_{unf} plus an additive value.

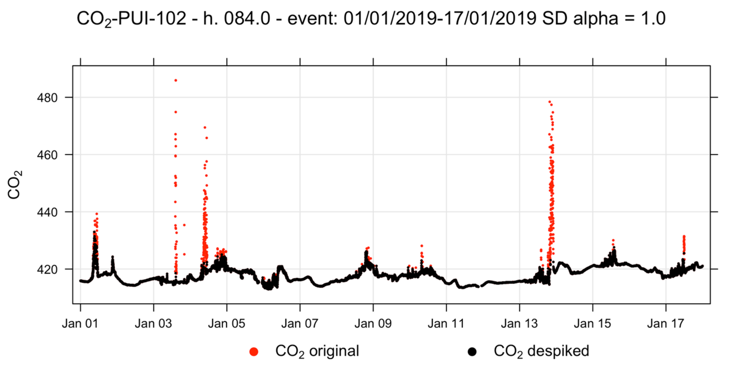
For CO_2 and CH_4, α was set to 1 and for CO it was set to 3. σ is the standard deviation of the data falling between the first and the third quartile of the dataset during the 10 previous days. n is the number of data points between C_{unf} and C_i

REBS (Robust extraction of baseline signal)

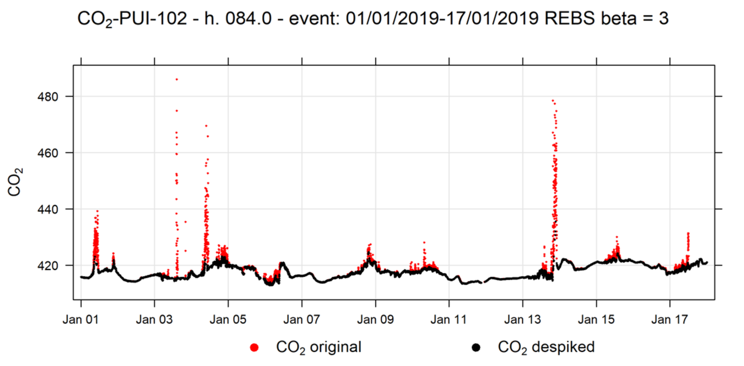
$$Y(t_i) > \hat{g}(t_i) + \beta \times \gamma$$

The data points are evaluated against a threshold value defined by a "reference" baseline $\hat{g}(t_i)$ plus two parameters β and γ.

For CO_2 and CH_4, β was set to 3 and for CO it was set to 8. The baseline was calculated on a bandwidth of 1 day to detect spikes during up to a few hours.



Example of application of SD algorithm for station Puijo (1-17 January 2019). Red points denote the data selected as "spikes"



Example of application of REBS algorithm for station Puijo (1-17 January 2019). Red points denote the data selected as "spikes"

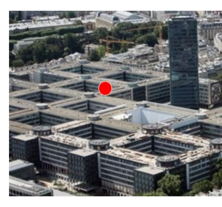
Test stations

We considered a subset of ICOS stations characterised by different typologies (i.e. continental, coastal and mountain) and one urban station (JUS) to test the two algorithms and to perform sensitivity and comparative experiments. 1-minute near real time data, i.e. preliminary calibrated and (automatically) quality controlled data, were considered over the period 2019 - 2020.

Jungfraujoch (JFJ)
Switzerland
Latitude/Longitude: 46.547N, 7.985E
Elevation: 3580 m
Mountain site

Monte Cimone (CMN)
Italy
Latitude/Longitude: 44.193N, 10.699E
Elevation: 2165 m
Mountain site

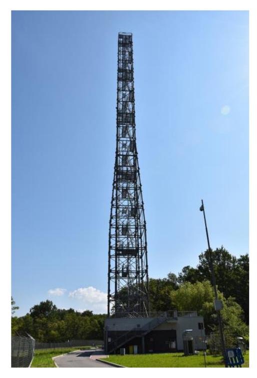
(JUS) Jussieu
France
Latitude/Longitude: 48.846N, 2.354E
Elevation: 37 m
Urban: 34 m (single level)



Puijo (PUI)
Finland
Latitude/Longitude: 62.909N, 27.654E
Elevation: 232 m
Tall tower: 84 m (highest level)

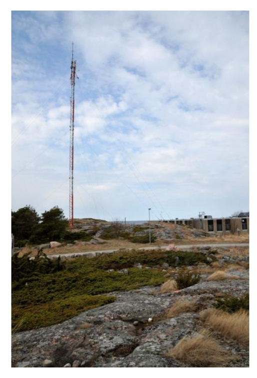
Karlsruhe (KIT)
Germany
Latitude/Longitude: 49.091N, 8.424E
Elevation: 110 m
Tall tower: 200 m (highest level)

Ispra (IPR)
Italy
Latitude/Longitude: 45.814N, 8.636E
Elevation: 210 m
Tall tower: 100 m (highest level)



Saclay (SAC)
France
Latitude/Longitude: 48.722N, 2.142E
Elevation: 160 m
Tall tower: 100 m (highest level)

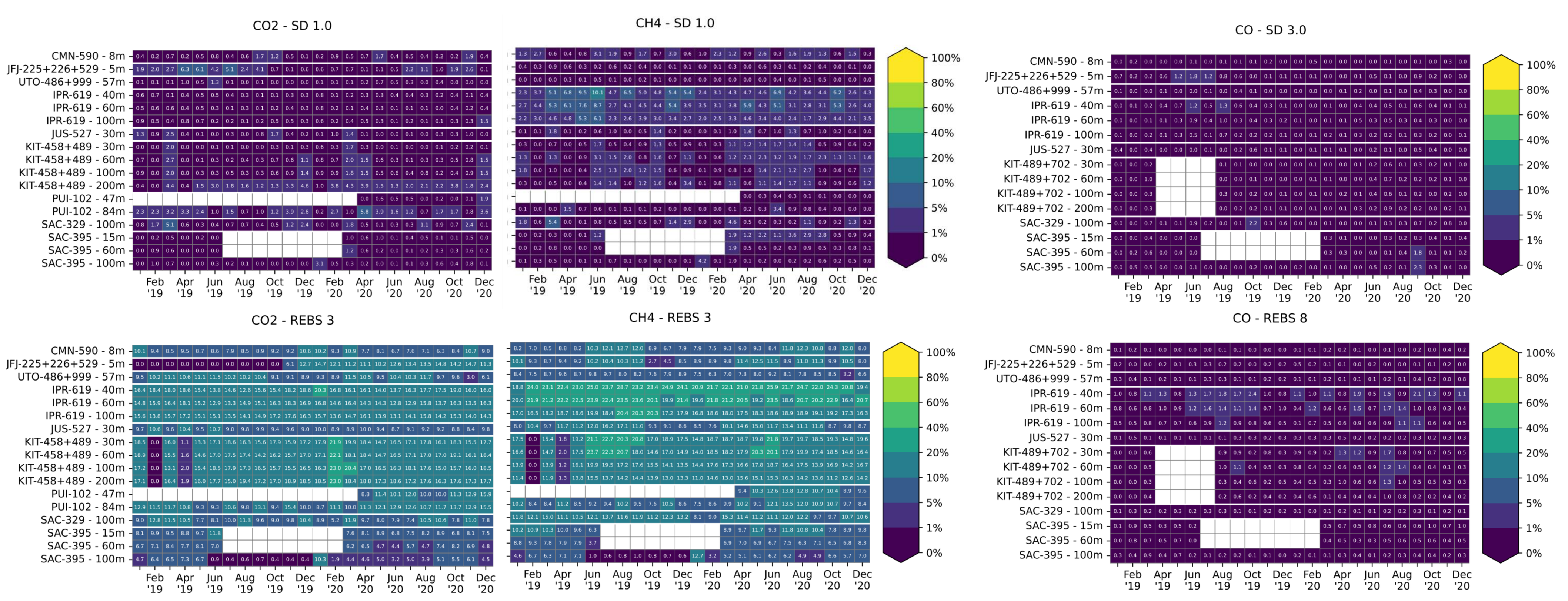
Utö (UTO)
Finland
Latitude/Longitude: 59.783N, 21.367E
Elevation: 8 m
Coastal/Marine: 60 m (single level)



ICOS INTEGRATED CARBON OBSERVATION SYSTEM

Sensitivity test: impact on spike frequency

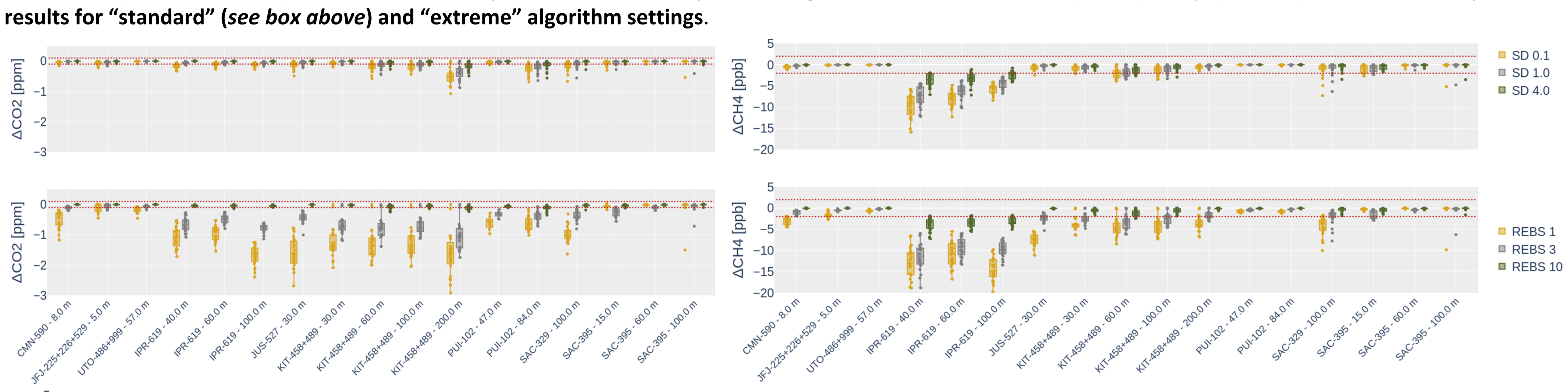
To investigate the impact of the algorithm settings on the spike detection, for each site and sampling height, we performed a sensitivity test by calculating the percentage of the data identified as "spikes" on the 1-minute dataset for CO_2, CH_4 and CO by using SD and REBS and by changing the setting parameters. For each algorithms (SD and REBS), 10 different runs were performed varying the sensitivity parameters of the algorithms: α changed from 0.1 to 4 for SD, β was changed from 1 to 10 for REBS. **Here we provide results for "standard" (*see the box above*) algorithm settings.**

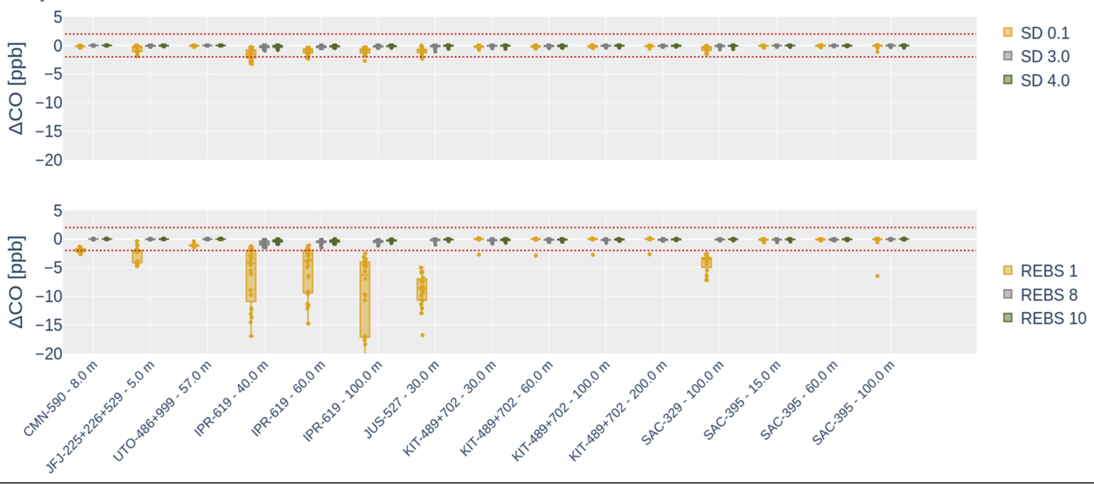


Heatmap of CO_2 (left), CH_4 (center) and CO (right) using SD (top) and REBS (bottom). For CO_2 and CH_4, α = 1 and β = 3 were used (α = 3 and β = 8 for CO). The numbers indicate the percentage of the data identified as "spikes" by the algorithms on each single month (x-axis). The y-axis denotes the sites and the sampling heights .

Sensitivity test: impact on monthly mean values

To investigate the impact of the algorithm settings on the monthly mean values of CO_2, CH_4 and CO, for each site and sampling heights, we calculated the differences (ΔCO_2, ΔCH_4, ΔCO) between the monthly mean values of de-spiked vs original dataset for different α (for SD) and β (for REBS) values. **Here we provide results for "standard" (*see box above*) and "extreme" algorithm settings.**





Boxplots of the differences between the monthly mean CO_2, CH_4 and CO calculated for the de-spiked and the original dataset. The horizontal red lines represent the WMO intercompatibility goals. For each site, different sampling heights are reported. The boxplots report the main percentiles of differences as well as outliers.

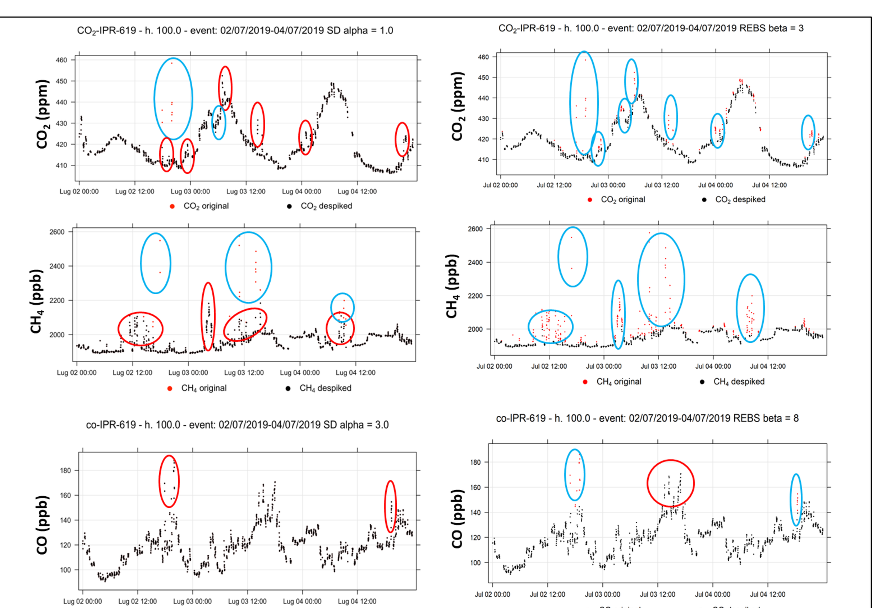
For CO_2 and CH_4, **REBS** (β=1, 3) provided larger deviations than SD in respect to the original dataset. This was especially evident for IPR, JUS, KIT and PUI. For CO, significant deviations were only found for **REBS** (β=1) at IPR and JUS. The remote sites (CMN, JFJ, UTO) and SAC-395 appeared less impacted by the de-spiking.

For the continental sites, larger deviations in respect to the original dataset were generally found after de-spiking with **REBS**, while lower deviations were generally observed for **SD** algorithm.

Analyses of case studies

For the test stations, PIs inspected the despiking results for representative spike periods and provided their comments about the performances. **Here we summarised the main outcomes:**

- IPR, PUI, UTO: with standard settings, REBS performed better than SD
- JFJ: with standard settings, both SD and REBS tended to overestimate spikes
- SAC: both SD and REBS not optimal spike selections. SD appeared to have more skills than REBS
- JUS: with standard settings, both SD and REBS tended to over detect CO_2 and CH_4 spikes. For CO, REBS with standard configuration looks better than SD.



Example from IPR (right): CO_2, CH_4 and CO measurements on 2 July 2019. Black data points are the retained measurements; red points represent the flagged using SD (left) and REBS (right) methods. Blue (red) circles represent data attribution manually confirmed (not confirmed) by site PI.

"Dichotomous" analysis

For five sites, PIs made a manual spike selection on selected measurement periods. These manual identification were compared with the SD and REBS results and objective metrics (Stephenson, 2000; Thornes & Stephenson, 2001) were calculated: BIAS, false alarm rate (F) and hit rate (H).

Each data could be either:
A - Flagged as spike by the algorithm and by the station PI
B - Flagged as spike by the algorithm but not flagged as spike by the station PI
C - Not flagged as spike by the algorithm but flagged as spike by the station PI
D- Not flagged as spike neither by the algorithm, nor by the station PI

In the following formulas, A, B, C and D represent the number of data that respect the above conditions: BIAS=(A+B)/(A+C); H= A/(A+C) ;F=B/(B+D)

All spikes

station	specie	algorithm	BIAS	F	H
CMN	CH4	REBS	13.3	0.12	0.59
		SD	2.0	0.02	0.18
	CO	REBS	2.6	0.00	0.29
		SD	1.8	0.00	0.07
	CO2	REBS	50.2	0.08	0.28
		SD	6.1	0.01	0.01
IPR	CH4	REBS	1.1	0.10	0.59
		SD	0.2	0.00	0.20
JUS	CH4	REBS	381.1	0.13	0.96
		SD	29.8	0.01	0.69
	CO	REBS	2.9	0.00	0.38
		SD	1.2	0.00	0.32
	CO2	REBS	2309.0	0.09	1.00
		SD	231.3	0.01	0.33
PUI	CH4	REBS	0.7	0.08	0.23
		SD	0.0	0.00	0.02
	CO2	REBS	0.9	0.08	0.43
		SD	0.2	0.00	0.14
UTO	CH4	REBS	5.4	0.09	0.17
		SD	0.2	0.00	0.02
	CO	REBS	0.1	0.00	0.03
		SD	0.0	0.00	0.01
	CO2	REBS	5.5	0.10	0.22
		SD	0.4	0.01	0.05

High spikes

station	specie	algorithm	BIAS	F	H
CMN	CH4	REBS	24.5	0.12	0.90
		SD	3.6	0.02	0.31
	CO	REBS	4.0	0.00	0.44
		SD	2.8	0.00	0.11
	CO2	REBS	119.9	0.08	0.67
		SD	14.7	0.01	0.02
IPR	CH4	REBS	2.2	0.12	0.89
		SD	0.4	0.01	0.36
JUS	CH4	REBS	412.9	0.13	1.00
		SD	32.3	0.01	0.75
	CO	REBS	3.1	0.00	0.41
		SD	1.2	0.00	0.34
	CO2	REBS	2309.0	0.09	1.00
		SD	231.3	0.01	0.33
PUI	CH4	REBS	7.3	0.09	0.90
		SD	0.2	0.00	0.16
	CO2	REBS	3.6	0.10	0.95
		SD	0.6	0.01	0.42
UTO	CH4	REBS	83.9	0.09	0.93
		SD	3.3	0.00	0.20
	CO	REBS	0.6	0.00	0.21
		SD	0.3	0.00	0.09
	CO2	REBS	36.5	0.10	0.84
		SD	2.5	0.01	0.33

A further analysis was then conducted only considering "high" spikes (**right**). "High" spikes were defined as the 1-minute data points whose distance from a one hour rolling mean baseline is higher than 0.5 ppm for CO_2 and 2 ppb for CH_4 and CO. A sensitivity study was performed by changing the value of the spike selection thresholds: not evident deviations were pointed out in respect to using 0.5 ppm and 2 ppb.

In general, **REBS** got higher "H" than **SD**, but also higher "F". Both the algorithms got higher "H" when "high" spikes are considered.

Summary

- For CO_2 and CH_4, REBS selected more spikes than SD with standard settings. This implied a stronger impact of REBS on monthly mean values. Similar performances were pointed out for CO between the two algorithms.
- Based on the case studies analysis, to identify one common algorithm/configuration for all the sites/species still looks challenging.
- For specific sites (e.g. IPR, PUI, UTO), REBS appeared to perform better than SD.
- The dichotomous analysis suggested that the REBS detected a larger fraction of the spikes detected by PIs (*but also more spikes NOT detected by the PIs*)
- The agreement of automatic and manual detection (both SD and REBS) increased for "high" spikes.

References:
El Yazidi, et al.: Identification of spikes associated with local sources in continuous time series of atmospheric CO, CO_2 and CH_4, Atmos. Meas. Tech., 11, 1599–1614, https://doi.org/10.5194/amt-11-1599-2018, 2018.
Thornes, J. E. & Stephenson, D. B.: How to judge the quality and value of weather forecast products, Meteorol. Appl., 8, 307–314, 2001.
Stephenson, D. B.: Use of the "Odds Ratio" for Diagnosing Forecast Skill, Weather and Forecasting, 2000.

Cosimo Fratticioli's and Pamela Trisolino's grants are funded by the National Project Rafforzamento del Capitale Umano CIR01_00019 – PRO-ICOS_MED *"Potenziamento della rete di osservazione ICOS-Italia nel Mediterraneo – Rafforzamento del Capitale Umano"*