

## Progetto NextDATA

### WP 1.1 High-altitude climatic observation system and climate station network

D1.1.B: Launch of experimental services for data quality check and near-real time data delivery/early warning at the Mt. Cimone GAW global station

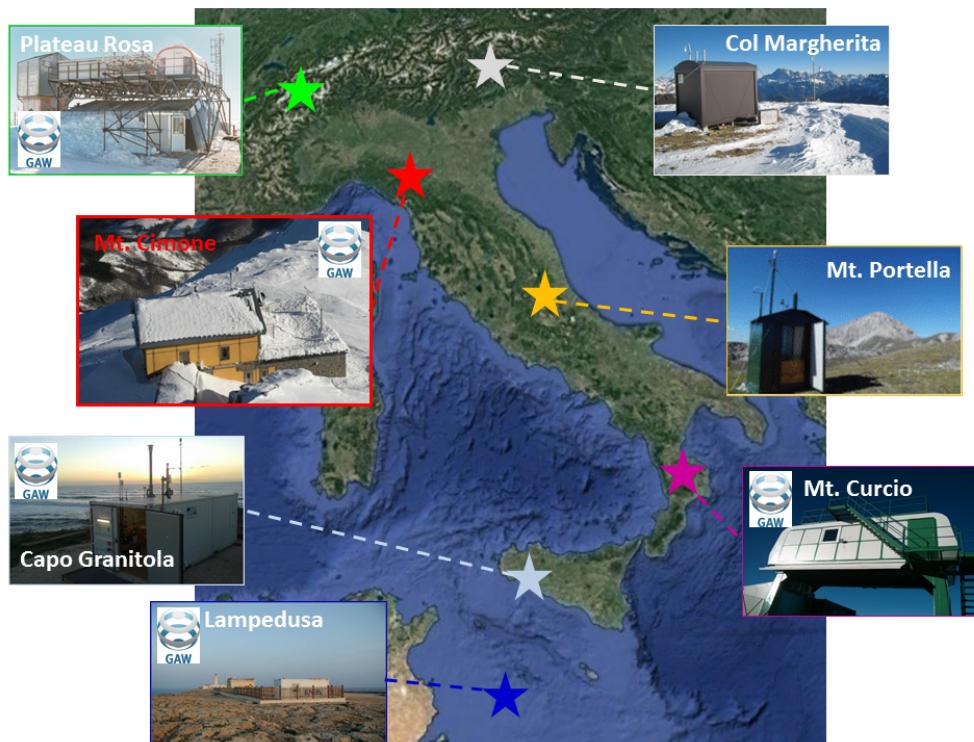
#### 1. Introduction

As reported in the executive plan, the NextDATA project aims at creating a climate observational network in mountain and remote areas, based on climate observatories for the monitoring of meteo-climatic conditions and atmospheric composition (Fig. 1). The network for monitoring the background atmospheric composition comprises five high-mountain atmospheric observatories: Monte Cimone (CMN, Northern Apennines, the only WMO/GAW global station in Italy; 2165 m), Plateau Rosa (PRS, Western Alps, WMO/GAW regional station; 3480 m), Col Margherita (MRG, Eastern Alps; 2550 m), Monte Portella (CMP, Central Apennines; 2912 m), and Monte Curcio (CUR, Southern Apennines, WMO/GAW regional station; 1796 m). In addition to these observatories, the WMO/GAW regional stations Capo Granitola (south-western Sicily, CGR) and Lampedusa (LMP), although not located at high-altitude, will provide complementary information on the background conditions of the Mediterranean Basin area.

The project activity, aimed at monitoring the background atmospheric conditions, is based on the upgrade and support to already existing atmospheric observatories, managed by different research organizations and CNR institutes. Several of these observatories are already part of international projects/research programs for the monitoring of Essential Climate Variables (ECVs). More specifically, in the framework of the WMO/GAW activity, observations of greenhouse and reactive gases are carried out at Plateau Rosa, Mt. Cimone, Mt. Curcio, Capo Granitola and Lampedusa (Table 1). Moreover, Plateau Rosa, Mt. Cimone and Lampedusa are also included among the potential sites in the European Research Infrastructure ICOS (Integrated Carbon Observation System), a pan-European research infrastructure which provides harmonized and high precision scientific data on carbon cycle and greenhouse gas budget and perturbations. Measurements of physical properties of atmospheric aerosols are performed at Mt. Cimone, Mt. Curcio, and Capo Granitola. Mt. Cimone, is part of ACTRIS-2 (Aerosols, Clouds and Trace gases Research Infrastructure) project which implements ACTRIS-RI, a pan-European initiative consolidating actions amongst European partners producing high-quality observations of aerosols, clouds and trace gases. Col Margherita and Mt. Curcio are part of GMOS (Global Mercury Observation System), a global observational network providing comparable data on mercury levels in ambient air and deposition, from monitoring sites worldwide distributed in the Southern and Northern Hemispheres. The GMOS network, through the GEO-Flagship GOS<sup>4</sup>M (Global Observation System for Mercury) in the framework of the GEO Strategic Plan (2016-2025), aims to support the goals of GEOSS and other on-going international programs (e.g., UNEP Mercury Program) as well as the Minamata Convention implementation (Arts. 19 and 22). These measurements constitute the basis for a more effective integration between observations carried out at these 7 measurement sites.

To contribute to the implementation of the background observational network, WP1.1 aims at setting-up a system for the automatic processing of ECV data coming from the different sites of the NextDATA network. In particular we propose a common approach for the data validation (data flagging) and common data formats for submission to NextDATA archives and to the reference international programs/projects. Indeed, due to the large amount of data recorded by the measurement stations, it is not feasible to perform the data validation/flagging by the so-called “visual inspection”, and by manual manipulation of data files. For these reasons, here we propose to develop and implement automatic procedures for data verification and formatting. The implementation of such automatic procedures will represent a powerful resource to help the researchers in spending more time for the scientific purposes, rather than data verification and formatting. Besides making the data creation process faster and favoring a timely data submission, the adoption of standardized validation procedures will also assure a more subjective flagging of data, as well as the possibility to trace back the actions which led to data validation (i.e., data revisions will be easier).

On a daily basis, raw data from measurement sites are transferred to a server located at CNR-ISAC HQs (in Bologna) for automatic processing and storage. The automatic processing encompasses a preliminary harmonization of file formats which is a pre-requisite for data flagging, data aggregation (on common temporal frames: 1 and 60 minutes) and final harmonization of file format following the guidelines from international reference programs or projects (i.e., WMO/GAW, ACTRIS-2. See also Deliverable 2.1.1). The automatic data processing also encompasses the production of many data products (updated daily), which provide the Principal Investigators with an overview of the instruments and data behavior, to support both the quality control of data, as well as the data inspection for scientific or operational purposes (i.e., the identification of event of interest or to perform preliminary data analysis). Currently, the automatic processing of data has been activated for a subset of ECVs and measurement sites (see Table 2). In particular, the Mt. Cimone WMO/GAW global station has been selected as a “proof-of-concept site”, due to the large number of ECVs observed and the large diversity of data format produced by the measurement systems.



**Figure 1.** Geographical location and pictures of the monitoring stations belonging to the atmospheric background observational network supported by the NextDATA project.

**Table 1.** List of Essential Climate Variables (ECVs) collected at the atmospheric background monitoring observatories to be part of the automatic processing (MRG: Col Margherita; CMN: Monte Cimone; CMP: Monte Portella/Campo Imperatore; CUR: Monte Curcio; CGR: Capo Granitola; LMP: Lampedusa). Green cells indicate the ECVs for which automatic processing has been activated, yellow cells indicate ECVs for which routines have been produced but still not activated. Crosses indicate ECVs for which processing is planned to be activated before the end of the Project. <sup>1</sup>AOD measurements are processed in the framework of AERONET and GAW-NRT programmes.

ECV	Reference programmes	Monitoring stations					
		MRG	CMN	CMP	CUR	CGR	LMP
$CO_2$	WMO/GAW ICOS-RI		X		X	X	
$CH_4$			X		X	X	
$CO$			X		X	X	
$O_3$		X	X	X		X	X
$SO_2$			X			X	
$NO$	WMO/GAW ACTRIS		X			X	
$NO_2$			X			X	
<i>Particle scattering</i>			X		X	X	
<i>Particle absorption</i>			X		X	X	
<i>Particle size distribution (by SMPS)</i>					X		
<i>Particle concentration</i>					X	X	
<i>Coarse particle size distribution (by OPC)</i>				X	X	X	
$AOD^1$			X				X
Total $O_3$	WMO/GAW						
UV radiation							
Meteorological parameters	WMO/GAW	X	X	X	X	X	X
Solar radiation		X	X	X	X	X	X

A further effort to strengthen the implementation of the background observational network is related to the activation of near-real time data delivery and early-warning systems, based on the data collected at the background atmospheric stations. A list of priorities was defined in the deliverable D1.1.1 (June 2017, see Table 2). Here we will describe a series of test cases that have been exploited or activated before December 2017.

**Table 2.** List of Essential Climate Variables (ECVs) collected at the atmospheric background monitoring observatories to be part of the near-real time data delivery service (PRS: Plateau Rosa, MRG: Col Margherita; CMN: Monte Cimone; CMP: Monte Portella/Campo Imperatore; CUR: Monte Cuccio; CGR: Capo Granitola; LMP: Lampedusa). Green cells indicate the ECVs for which data delivery has been activated. Crosses indicate ECVs for which data delivery services are planned to be activated before the end of the Project.

ECV	Reference programs	Monitoring stations						
		PRS	MRG	CMN	CMP	CUR	CGR	LMP
CO	CAMS COPERNICUS			X			X	
O <sub>3</sub>		X	X	X	X		X	X
NO				X			X	
NO <sub>2</sub>				X			X	
Particle scattering	ACTRIS WMO/GAW			X		X	X	
Particle absorption				X		X	X	
Particle size distribution (by SMPS)				X		X		
Particle concentration				X		X	X	
Coarse particle size distribution (by OPC)	WMO SDS-WAS			X				
AOD	AERONET GAW-NRT			X				X

## 2. Automatic processing of ECVs

### 2.1 Data collection

The instrumental raw data are transferred (at least) once a day from the measurement stations to a “test” server located at CNR-ISAC HQs in Bologna. During the current development phase, different transfer strategies are applied, as a function of the specific requirements of the measurement stations. As an instance, for CMN data files are downloaded from the station server, while for MRG data files are uploaded by the station server to the CNR-ISAC server. To facilitate the participation of the stations, in this development phase we let the personnel of each station decide which strategy to adopt for data transfer. The files stored in the CNR-ISAC server already contain information of measured quantities in geophysical units, as well as internal diagnostic parameters used for automatic QA/QC.

All data files are centralized to the CNR-ISAC server as a function of originating station, ECV class (gas, aerosol, meteorology, and photometry) and ECV sub-class (i.e., instruments, see Table 3). Then, the data files are processed to obtain a homogeneous file system in terms of nomenclature and format, which will be ingested by the automatic procedure for QA/QC and flagging. For the file name, we adopted the following coding:

SSS\_PPP\_YYYYMMDD\_TTT.dat

where SSS is the station code, PPP is the ECV code (see Table 4 for the complete list of adopted codes), YYYYMMDD is the file production date, and TTT is the native time resolution of the measurements (i.e., 1-minute: “01M”, 30-minute: “30M”, 60-minute: “60M”).

**Table 3.** List of nomenclature and codes used for automated QA/QC

ECV class	Instrument/ECV	Code
Gas	Ozone	O3
	Carbon Monoxide	CO
	Sulphur Dioxide	SO2
	Nitrogen oxides	NOx
	Picarro	GHG
Aerosol	Nephelometer	Neph
	Multi Angle Absorption Photometer	MAAP
	Optical Particle Counter	OPC
	Size Mobility Particle Sizer	SMPS
Meteorology	Automated weather station	Meteo
	Radiometer	RAD
Photometry	Solar photometer	AOD

**Table 4.** List of station codes

Measurement station	Code
Plateau Rosa	PRS
Col Margherita	MRG
Mt. Cimone	CMN
Mt. Portella – Campo Imperatore	CMP
Mt. Curcio	CUR
Capo Granitola	CGR
Lampedusa	LMP

In this current implementation phase, specific routines have been implemented for the different stations and ECV classes, to harmonize the different files to the common format. It is foreseen that data will be provided by each station to the CNR-ISAC server already in the homogenized format.

As an example, in the following, we will describe the processing chain and the routines developed for the Thermo 42i-TL, a state-of-art instrument for the continuous determination of nitrogen monoxide (NO) and dioxide (NO<sub>2</sub>) mixing ratios based on chemiluminescence detection (CLD). This instrument belongs to a class of instruments widely diffused among the NextDATA measurement stations and, due to the complexity of steps necessary to obtain the final data, it represents an effective case study for implementing mature QA/QC routines. These routines will be useful also for other ECVs, as near-surface O<sub>3</sub>, SO<sub>2</sub>, CO, greenhouse gases (GHG) and punctual aerosol measurements (near-surface scattering and absorption coefficients, particle number concentration). Figure 2 reports the different steps of the data processing.

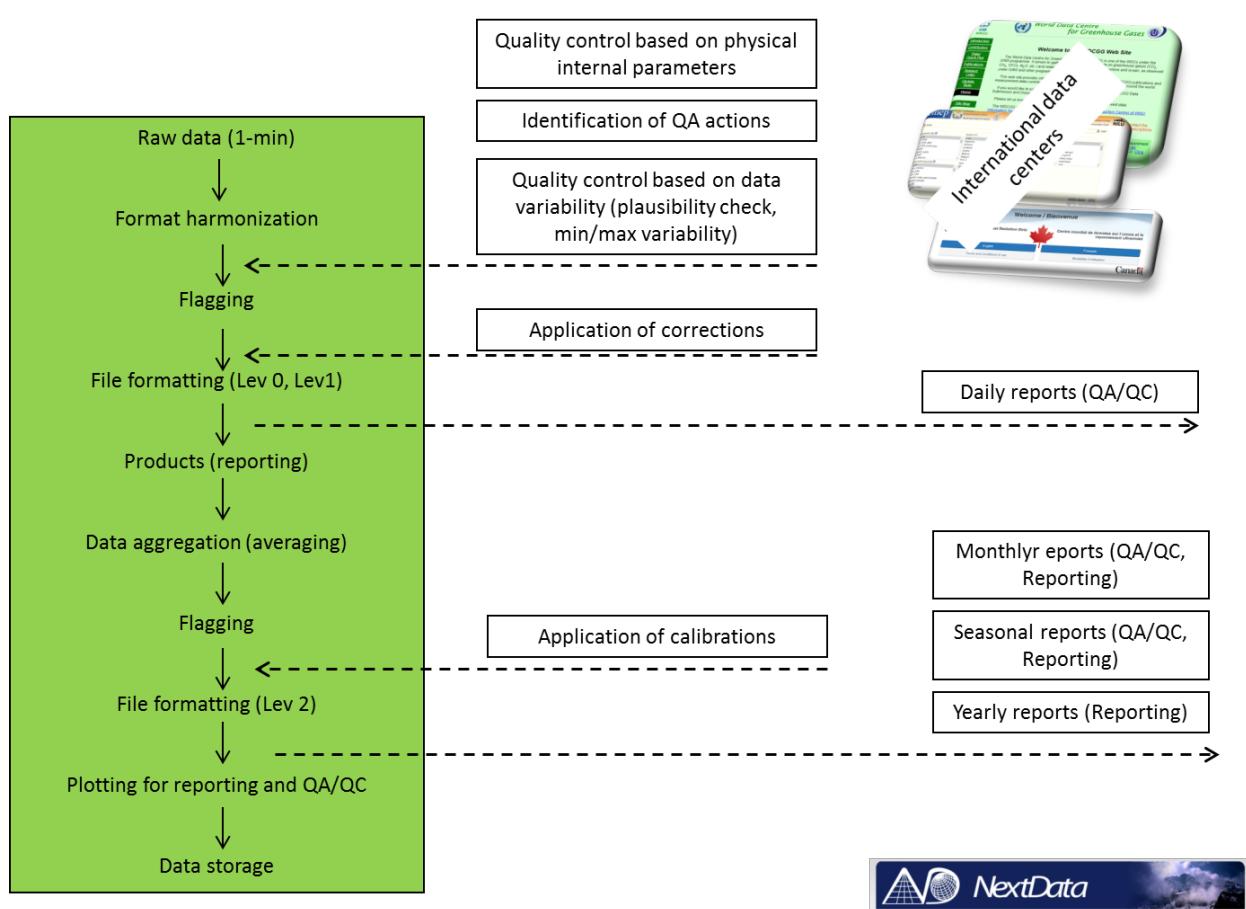
## 2.2 Data formatting

To optimize the interoperability of the data system, the processed data are formatted in agreement with the guidelines of the WMO/GAW data-centers. The greenhouse gases (CO<sub>2</sub>, CH<sub>4</sub>) and the carbon monoxide (CO) data are created in agreement with formats and metadata indicated by the World Data Centre for Greenhouse Gases (WDCGG), as reported in “Revision of the WDCGG Data Submission and Dissemination Guide” (GAW Report No. 188). “Near-surface” reactive gases (O<sub>3</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>) and meteorological parameters are formatted following the NASA-Ames standard, as indicated by the World Data Centre for Reactive Gases (WDCRG). This format is based on the textual format ASCII NASA-Ames 1001 with additional

metadata (as a function of the different ECVs). The templates to be used for the different ECV can be found at the web page <http://ebas-submit.nilu.no/Submit-Data/Reporting-Templates/all-templates-temporary>. This standard is also used for the aerosol physical properties data (particle number size distribution, particle number concentration, coarse mode particle size distribution, particle light absorption coefficient, particle light scattering coefficient) and AOD (Aerosol Optical Depth), in agreement with the World Data Centre for Aerosol (WDCA). The files created by the automatic processing chain contain all of the observations carried out during a full solar year; they are updated on a daily basis.

### 2.3 Data processing

Specific processing chains have been developed for each instrument, although the general workflow is the same (Fig. 2): data-formatting, data control and flagging, data correction (if needed), data aggregation (time averaging), data flagging, data formatting and report production (Fig. 2).



**Figure 2.** Workflow of automatic data processing for Thermo 42iTL.

#### 2.3.1 Data control and flagging

The first step of the automatic processing chain consists in the identification of the different measurement modes. Usually, a gas or aerosol analyzer runs in 4 different modes: “sample” (when ambient air is measured), “calibration” (when air from one or more laboratory standards is measured), “zero” (when a gas mixture scrubbed by the molecules to be quantified is measured, typically to determine the instrumental zero off-set for routine quality checks) and “span” (when dry air enriched by a specific amount of the molecules to be quantified is measured, typically to determine the instrumental span factor for routine quality checks). Systematic variations with time of the zero off-set and the span factor are used to

timely detect instrumental problems, while a “full” calibration is used to link the measurement to a reference calibration scale hosted by a central laboratory. Depending on the ECV and the instrument, the zero source can be either tanks with pure dry air or a dry air generator, while span sources can be a tank with assigned mixing ratio or a permeation tube or an internal generator (like a UV-lamp in the case of O<sub>3</sub>). The identification of the measurement mode, which leads to the attribution of a specific flag to the data records, is fundamental for two reasons: (1) data affected by calibrations or quality checks must be discharged for the production of the final “validated” data-set, and (2) data recorded during calibrations or quality checks are used to obtain correction factors or quality control metrics.

The identification of measurement periods affected by “calibrations” or “zero/span” can be performed considering two general cases: (1) by working with internal diagnostic parameters of analyzers or “calibration units” (as an instance, some classes of instruments provide the acquisition system with the information related to their current “mode”), or (2) by searching the existence of a log-file that indicates the occurrence of QC exercises.

The second step is the analysis of instrumental data. Such step is based on general criteria, but, at the same time, it is adapted as a function of specific measurement stations and ECVs. The following checks are implemented in the data control process:

- *Diagnostic/instrumental checks*: the internal diagnostic parameters (e.g., temperatures, flows, pressures) are considered, and these values are compared with their typical ranges, which are reported on the instrument manual. For each measurement, if at least one parameter fails a check, the data is flagged as invalid (and thus not considered for the successive data aggregation). As an example, of the flags used for NO<sub>x</sub> are reported by Table 5.
- *Plausibility checks*: identification of measurement periods with measured values exceeding the expected variability (i.e., definition of plausible ranges). These ranges are defined as a function of the measurement stations (e.g., the plausible range for atmospheric pressure at a coastal station like CGR is different from that of a mountain station like MRG). Currently, two different processes for the definition of these plausibility ranges are considered. The first is the adoption of specific threshold values, defined upon existing literature and in collaboration with scientists in charge of the instruments. The second process is related to the on-line calculation of variable threshold values basing on statistical analysis of data over specific time periods (1 hour, 1 day, 1 month or a full year), e.g., percentiles of the data-set or confidence intervals like n-sigma above or below the average values.
- *Variability checks*: verification of the variability (i.e., rate of change with respect to time) of the observed ECV. Depending on the considered ECV and the site features, a range for minimum and maximum ECV variability is defined (typically on hourly basis).
- *Comparison among parallel/simultaneous observations*: if available at the same measurement site, the time series of the same ECV recorded by different instruments can be compared to point out possible anomalies.

For Thermo42iTL, the variability ranges used for the aforementioned checks are reported in Table 6.

**Table 5.** Example of descriptive flags (following the NASA-Ames format) for Thermo 42iTl instrument.

Flag	Description
0.999000000	Missing data
0.147000000	Under detection limit
0.682000000	Calibration
0.664000000	Low sampling flow
0.456000000	High variability
0.440000000	Derived value (corrected for night-time offset)
0.390000000	Less than 50% of data used for data averaging

**Table 6.** list of parameters used for data control of Thermo 42iTl instrument with defined plausible threshold values.

Parameter	Variability range
Flow sample	0.5 – 1.5 lt/min
NO variability	NO(i)-NO(i+1) < 0.5 ppb
NO <sub>2</sub> variability	NO <sub>2</sub> (i)-NO <sub>2</sub> (i+1) < 0.5 ppb
NO	0.05 – 20
NO <sub>2</sub>	0.08 – 20
Zero offset	-0.5 – 0.5
Span coeff	0.90 - 1.10
Sc	0.1 -1.0

### 2.3.2 Data correction

For NO<sub>x</sub>, the major difficulty in this task is related to the ingestion by the processing chain of the information provided by automatic calibration (zero and span). CLD instruments directly detect and quantify only the NO mixing ratio; therefore, it is necessary to convert NO<sub>2</sub> to NO to quantify NO<sub>x</sub> (and finally NO<sub>2</sub>) mixing ratios. Commercially available instruments for air-quality monitoring are usually equipped with a Molybdenum (Mo) heated converter. However, this set-up is not recommended by GAW (2011), since this kind of detector is not selective to NO<sub>2</sub>: it also converts other oxidized nitrogen compounds, such as nitric acid (HNO<sub>3</sub>), peroxyacetyl nitrate (PAN), and other organic nitrates (Steinbacher et al., 2007). For these reasons, and in agreement with ACTRIS-2 and WMO/GAW guidelines, the NextDATA instrument is equipped with a photolytic converter (Blue Light Converter, BLC, Air Quality Design Inc. USA), which uses an UV light source to selectively convert NO<sub>2</sub> to NO. Since the BLC conversion efficiency (Sc) is significantly lower than 100%, it is paramount to derive the actual value of Sc. The Sc obtained by a gas phase titration carried out during the calibration is then used to correct the NO<sub>2</sub> reading and to obtain the actual NO<sub>2</sub> mixing ratio. The calibration process is composed by the following steps:

1. Sampling of zero air: the NO reading is used to calculate instrumental zero-offset.
2. Sampling of span air (obtained by diluting 5 ppm of NO in N<sub>2</sub> to about 100 ppb): The NO (NO<sub>(1)</sub>) and NO<sub>x</sub> (NO<sub>x(1)</sub>) readings are used to calculate the NO and NO<sub>x</sub> span factors.
3. After the determination of the new calibration factors for NO (zero offset and span factor), O<sub>3</sub> is added to the mixture, so that 80% of the NO amount is titrated (gas phase titration).
4. After stabilization, the data for the NO (NO<sub>(2)</sub>) and NO<sub>2</sub> (NO<sub>x(2)</sub>) are recorded.
5. Sampling of zero air to purge the instrument after calibration.

The converter efficiency (Sc) is calculated as follows:

The effective produced NO<sub>2</sub> amount results from: [NO<sub>2</sub>] = [NO<sub>(1)</sub> - NO<sub>(2)</sub>]

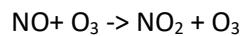
The NO<sub>2</sub> amount converted by BLC is calculated by: [(NO<sub>x(2)</sub> - NO<sub>(2)</sub>) - (NO<sub>x(1)</sub> - NO<sub>(1)</sub>)]

Accordingly, the efficiency factor is calculated by:

$$Sc = \frac{[(NOx_{(2)} - NO_{(2)}) - (NOx_{(1)} - NO_{(1)})]}{[NO_{(1)} - NO_{(2)}]}$$

The obtained zero offset, span coefficients, and Sc values are stored in a table. A linear interpolation along the time is applied between two consecutive calibration events, and then calibration parameters are applied to NO and NOx data for obtaining a “calibrated” time series. The “calibrated” NO<sub>2</sub> is obtained by subtracting the “calibrated” NO from the “calibrated” NOx. A list of threshold values are applied to the calculated calibration factors). If the check of calibration factors against threshold values fails, the last calibration factors successfully calculated are retained for data correction.

In a successive step, a further correction is applied to NO data, by calculating for each day the night-time (00:00 – 04:00 UTC+1) average value. Under excess of O<sub>3</sub> (like in a remote or rural region during night-time), NO must be completely titrated by the reaction:



In these conditions, NO is expected to decrease below the instrumental detection limit. If not, a “night-time” zero offset correction is calculated by linearly interpolating the “night-time” NO reading and by subtracting the “night-time” offset from the data series.

### 2.3.3 Data Averaging

Basing on the data screening and corrections applied in the previous step (i.e., coding of level-0 and level-1 data), 1-minute data are aggregated to hourly (60-minute) average values for obtaining level-2 data. For time aggregation, only data with a valid numflag are considered. This means that also data “under detection limit” are used for data aggregation. The measurement time associated with an average value corresponds to the beginning of the average period (e.g., the hourly mean at 1:00 is calculated by using the 1-minute data from 01:00 to 01:59). In case less than 30 (1-minute) data are used for the calculation of the hourly mean value, the data is flagged (0.390); if less than 2 (1-minute) data are used, the hourly mean is set to missing (0.999). Time is expressed in UTC.

## 2.4 Data products

To support the station personnel in carrying out the QA/QC checks, a suite of products (i.e., data plots) is produced by the automatic data processing. The data products are created or updated on a daily basis, by using specific routines based on R software. To this aim, some specific functions of the “OpenAir” package (see Carslaw and Ropkins, 2012) are used.

The data products are arranged as a function of their data coverage: daily, monthly, seasonal and yearly. In the following, we provide a brief overview of the products already available.

### 2.4.1 Daily data products

*Description:* These products are based on automated plots generated daily, using native time resolution of data. They provide a time series (along a full calendar day) of ECV raw and corrected data, together with instrumental diagnostic parameters and related flags.

*Aim:* The main purpose is to have a daily diagnostic about measurement status, as well as a high-resolution quick-view of the ECV variability on a daily time frame.

### 2.4.2 Monthly data products (1)

*Description:* These products are based on automated plots generated daily, using native time resolution of data, as well as averaged data. They provide information about corrected and averaged (level-2) ECV data, together with internal instrumental parameters and automatic flags.

*Aim:* The main purpose is to provide a medium-term diagnostic about instrument performance (e.g., detecting medium-term instrumental drifts able to affect measurements), as well as to give a quick view of level-2 data. This product also provides a table with basic statistical parameters for the corrected ECV.

#### 2.4.2 Monthly data products (2)

*Description:* These products use the “timeVariation” function of the “OpenAir” package (Carslaw and Ropkins, 2012) on level-2 data. For each calendar month, a plot representing the average diurnal variation of the considered ECV (with statistical confidence interval) is provided.

*Aim:* This product provides information about typical diurnal variability of the selected ECV. The main purpose is data reporting and QA/QC (identification of anomalous behavior for the selected ECV).

#### 2.4.3 Monthly data products (3)

*Description:* These products are based on automated plots generated month-by-month using corrected and averaged (level-2) ECV data. They provide the time series of corrected ECV values for each single calendar month.

*Aim:* The main purpose is to provide a medium-term diagnostic about ECV data, also useful to detect the occurrence of “special” events. As an instance, for each month, a reference line describing the ECV mean average value, as well as the indication of the lowest and the highest ECV values, are provided.

#### 2.4.4 Seasonal data products

*Description:* These products are based on automated plots generated month-by-month using corrected and averaged ECV (level-2) data. For each season (DJF, MAM, JJA, SON), they provide the average diurnal variation of corrected ECV values, their annual cycle and mean averaged weekly variation. For each analysis, the statistical confidence interval of averages is provided.

*Aim:* These products provide information about typical diurnal variability of the selected ECV for each season. The main purpose is the operational data reporting and QA/QC (i.e., identification of anomalous diurnal behaviors potentially due to instrumental problems).

#### 2.4.5 Yearly data products (1)

*Description:* This product is based on automated plots generated daily by using corrected and averaged data ECV (level-2). For the whole year, ECV time series, flags and internal diagnostic parameters are provided. Additional statistical information about ECV variability and the percentage of valid data is also given.

*Aim:* The main purpose is to have a medium-term diagnostic about corrected ECV data, also useful to detect the occurrence of “special” events. Moreover, also internal instrumental parameters are plotted for QA/QC purpose. This would allow the station PI to timely detected possible instrumental changes or drifts occurring over several months.

#### 2.4.6 Yearly data products (2)

*Description:* This product uses the “calendarPlot” function of the “OpenAir” package (Carslaw and Ropkins, 2012) on level-2 corrected data.

*Aim:* It provides the average daily values throughout the year for the selected ECV. The main purpose is operational data reporting and quick-view about ECV data availability/coverage at the station.

### 2.5 Routine availability

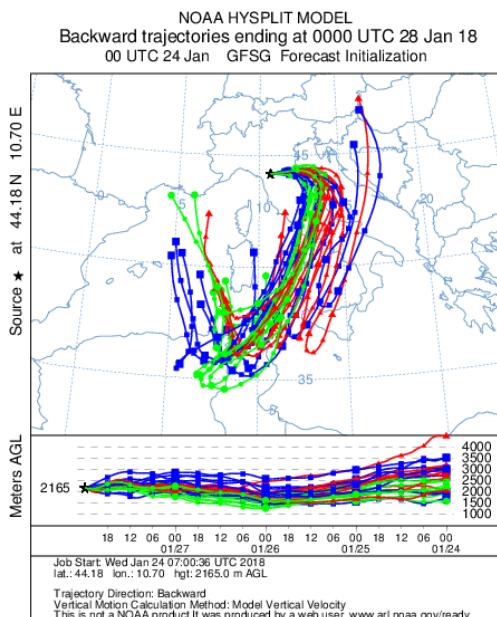
All the routines implemented for the different steps involved in the automatic processing of data are written by using the R suite. Once finally validated, these routines will be freely accessible on the NextDATA website.

### 3. Implementation of NRT data delivery and early warning services

#### 3.1 Implementation of an experimental service for detection of dust events and early warning

During the previous year of activity, an “early-warning” system for the identification of Saharan dust events (hereinafter SDEs) was tentatively implemented at the Mt. Cimone station. The system was based on the methodology presented by Duchi et al. (2016), which coupled the observation of in-situ “near-surface” coarse aerosol particle number concentration (in the range  $D_p = 1\text{--}10 \mu\text{m}$ , obtained by an Optical Particle Counter 1.108 GRIMM) with 3-D back-trajectories. More specifically, a “dusty day” was identified when a significant increase in the coarse concentration daily average value was associated with the presence of air-masses originated from northern Africa.

To implement a “near-real time” detection of SDEs at Mt. Cimone, we coupled the near-real time detection of increases in the “near-surface” coarse particles number concentration with information of air-mass origin and path, as deduced by back-trajectory forecasts calculated by the Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT), see Figure 3. HYSPLIT is developed by NOAA’s Air Resources Laboratory, one of the most widely used models for atmospheric trajectory calculations (Stein and Ngan, 2015). On each calendar day, an ensemble of forecasted 120-hour long back-trajectories is calculated for Mt. Cimone, for the next 4<sup>th</sup> calendar day (see <http://www.isac.cnr.it/cimone/hysplit>). Back-trajectory calculations are based on the operational NOAA GFSG forecast data file.



**Figure 3.** Example of 120-hour back-trajectory ensemble forecast calculated for CMN by HYSPLIT model. The ensemble describing the air-mass origin for 28<sup>th</sup> January 2018 at 00:00 UTC was calculated on 24<sup>th</sup> January 2018 at 00:00 UTC.

The daily back-trajectory ensemble file is automatically processed to point out the possible passage over a geographical box roughly representing northern Africa. The ensemble is calculated for taking into account possible uncertainties in simulation of air-mass circulation. If air-masses are forecasted to pass over northern Africa and an increase of coarse particle concentration is observed for at least three hours (to retain only robust events), a warning is launched by automatic e-mails.

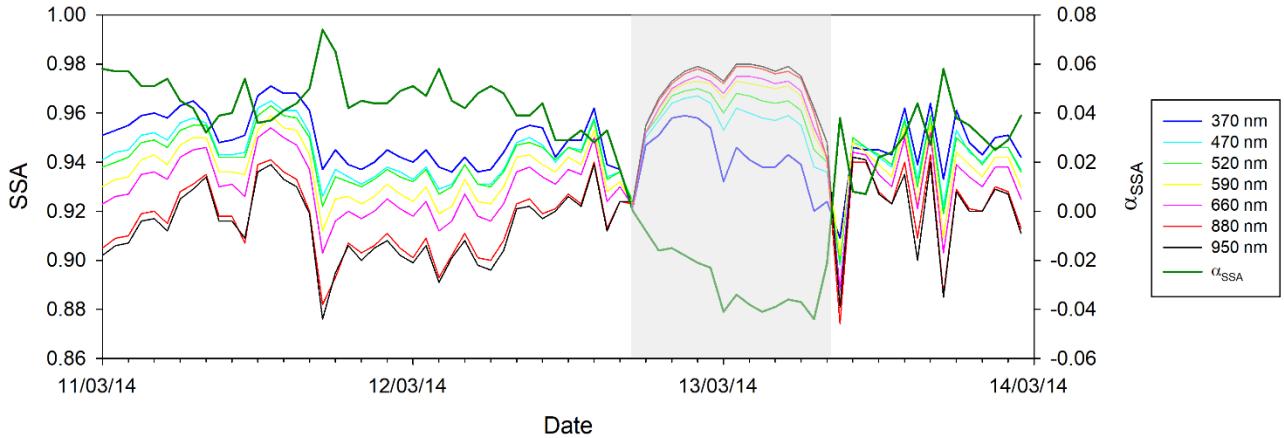
To evaluate the “robustness” of the Saharan dust event detection based on the analysis of the “near-surface” coarse particle number concentration, the aforementioned methodology was compared to another well-known methodology used to assess the occurrence of SDEs at the mountain site Jungfraujoch in the Swiss Alps (Collaud Coen et al., 2004). This reference methodology is based on the simultaneous measurements of scattering and absorption coefficients at different wavelengths, and on the calculation of

the wavelength dependence of the single scattering albedo (SSA). At Mt. Cimone, a three-wavelength integrating nephelometer (TSI 3563) is running since 2014, while a seven-wavelength aethalometer (Magee Scientific AE31) was used to carry out measurements over the period 2011–2015. The “core” of this methodology is the variation of the Ångström exponent of the single scattering albedo with wavelength (hereafter referred to as  $\alpha_{SSA}$ ): the authors indicate as SDEs the time periods that exhibit negative  $\alpha_{SSA}$  for more than 3 hours. This is also visible from the inversion of the wavelength dependence of the SSA during hours characterized by dust transport. An application of the Collaud Coen et al. (2004) methodology to Mt. Cimone measurements is shown in Figure 4, in which these two features are clearly visible. To obtain the time series (over the two-year period common to both of the instruments) of the  $\alpha_{SSA}$ , we proceeded as follows:

- Scattering coefficients from the nephelometer ( $\lambda = 450, 550$ , and  $700$  nm) were fitted with a power-law dependence, to determine the scattering exponent  $\alpha_{sca}$
- By using the  $\alpha_{sca}$  value, scattering coefficients at the seven wavelengths of the aethalometer ( $\lambda = 370, 470, 520, 590, 660, 880$ , and  $950$  nm) were then computed
- For each of these 7 wavelengths, the single scattering albedo (SSA) was evaluated, according to  $SSA = \frac{\sigma_{sca}}{\sigma_{sca} + \sigma_{abs}}$ , where  $\sigma_{sca}$  and  $\sigma_{abs}$  are the scattering and absorption coefficients, respectively
- The seven SSA values were then fitted with a power-law dependence, to obtain  $\alpha_{SSA}$

According to the Collaud Coen et al. (2004) criterion for identifying a SDE based on the variation of  $\alpha_{SSA}$ , a total of 107 days characterized by dust transport were observed at Mt. Cimone, for 2014–2015. A sensitivity study was also performed, by varying the minimum consecutive hours to be selected as SDE (i.e., we used either 2 or 4 h), giving no significant differences.

To assess the comparison between the two methodologies, the “dichotomous” approach presented in Thornes and Stephenson (2001) was followed. More specifically,  $2 \times 2$  contingency tables were computed (Table 7); in these tables, each entry encloses a list of SDE or no-SDE days, defined by each methodology considered. From the contingency tables, it is possible to evaluate several skill scores, which are useful to measure the skill of one methodology in identifying SDE compared with the other one. For a complete explanation of the parameters hereby used, please refer to Thornes and Stephenson (2001) and Wilks (2006). In all of the three comparisons considered (i.e., by using a threshold of 2, 3 and 4 h for the Collaud Coen et al. (2004) methodology), the values of the statistical parameters (i.e., accuracy, false alarm ratio and probability of false detection) were similar. The accuracy in the range 0.75–0.78 seems to be in line with what reported in Collaud Coen et al. (2004), i.e., that 71% of SDE cases identified by a negative  $\alpha_{SSA}$  are “confirmed” by the presence of trajectories originated in northern Africa. Moreover, a moderate/high value of the false alarm ratio, associated to a low probability of false detection, could be explained by the fact that a SDE is generally less likely to occur than a no-SDE. In addition to these parameters, we considered the Odds Ratio Skill Score (ORSS, see Thornes and Stephenson, 2001), a parameter which varies between  $-1$  and  $+1$ , where a score of  $1$  represents perfect skill and a score of  $0$  indicates no skill; negative values imply that values of one series are opposite to what observed by the other one. In all cases the ORSS was above the minimum ORSS required to have skill at the 99% confidence level, thus indicating that the agreement between the two methodologies is not due to chance (i.e., is statistically significant).



**Figure 4.** Temporal evolution (hourly averages) of the calculated single scattering albedo (SSA) during a SDE (shaded area) at Mt. Cimone, from March 11<sup>th</sup> to 13<sup>th</sup>, 2014. Also reported in the plot is the temporal evolution of  $\alpha_{SSA}$ , computed by a power-law fit of the SSA values at the seven wavelengths.

**Table 7.** The 2x2 contingency tables for the comparison of the DTE event time series, i.e., identified by the Duchi et al. (2016) and the Collaud Coen et al. (2004)–hereby indicated as “CC” –methodologies. For CC, several thresholds on the minimum number of consecutive hours to identify a SDE were chosen (i.e., 2, 3, and 4 h). Capital letters are defined as follows: A indicates the number of SDE days selected by both methodologies, B represents the number of days selected as SDE by the first methodology (Y) but as no-SDE (N) by the second one, C represents the number of days selected as no-SDE by the first methodology but as SDE by the second one, and D represents the number of no-SDE days selected by both methodologies. Also reported in the table are several skill scores, as defined in Thorne and Stephenson (2001), and Wilks (2006).

		Duchi et al. (2016)		Duchi et al. (2016)		Duchi et al. (2016)	
		Y	N	Y	N	Y	N
CC (3 h thresh.)	Y	A = 25	B = 73	CC (2h thresh)	Y A = 33	B = 101	CC (4 h thresh)
	N	C = 43	D = 324		N C = 35	D = 296	
Accuracy (ACC)		0.75		0.71		0.78	
False Alarm Ratio (FAR)		0.74		0.75		0.71	
Probab. of False Detect. (POFD)		0.18		0.25		0.14	
ORSS		0.44		0.47		0.51	
min ORSS		0.35		0.35		0.35	

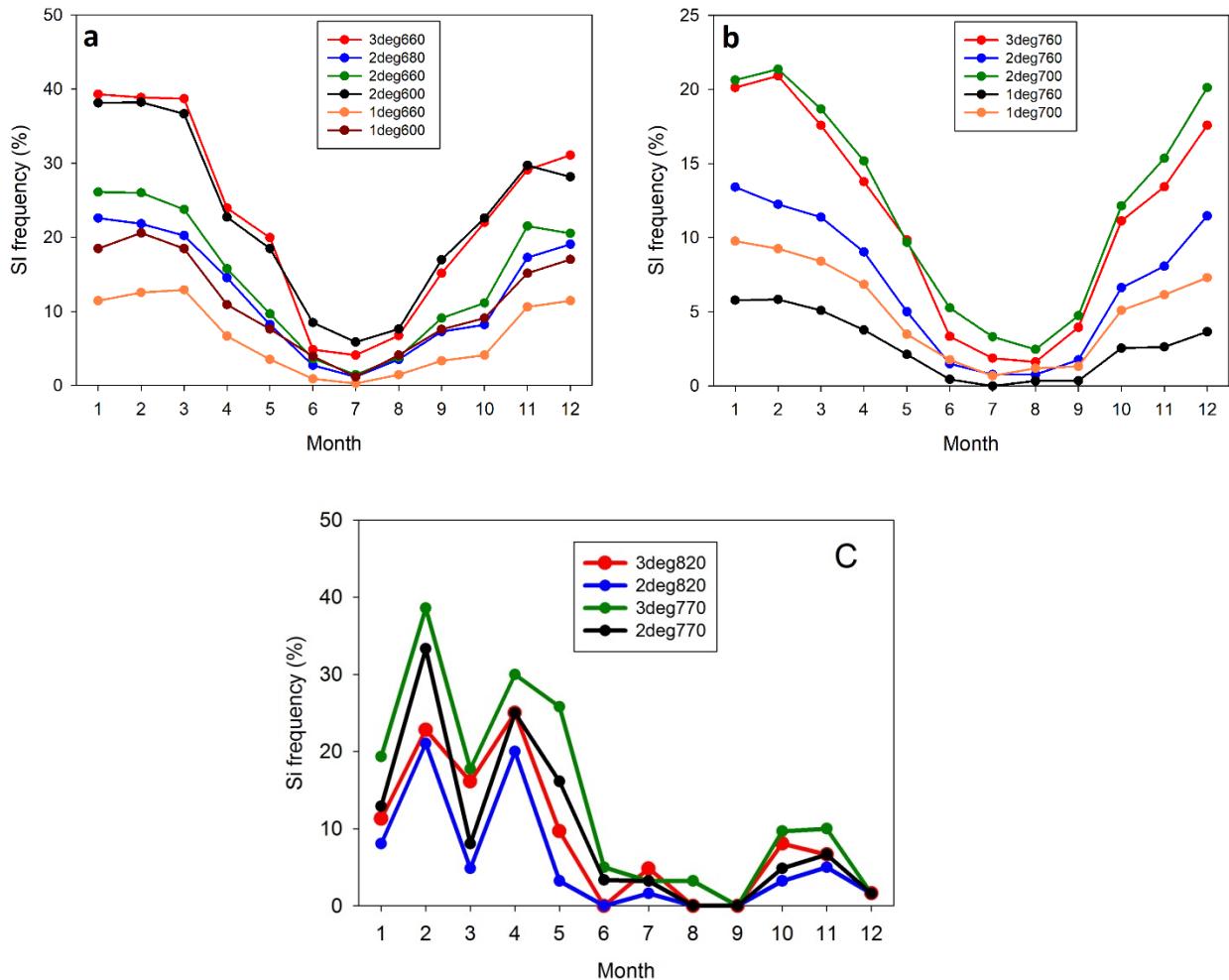
### 3.2 Application of STEFLUX to NextDATA mountain sites

STEFLUX (Stratosphere-to-Troposphere Exchange Flux, see Putero et al., 2016) is a fast-computing and reliable instrument for the identification of stratospheric intrusion (SI) occurring at a specific location and during a specific time window, developed in the framework of NextDATA. Its calculations are based on a compiled stratosphere-to-troposphere exchange (STE) climatology (Škerlak et al., 2014), which makes use of the ERA-Interim reanalysis dataset from the ECMWF, as well as a refined version of a well-established Lagrangian methodology. The STE climatology is available from 1979 and continuously updated, allowing also the climatological characterization of SI events for the chosen site.

To assess the STEFLUX ability to routinely identify the occurrence of stratospheric intrusion events at the high-altitude stations of the background network, we carried out a preliminary comparison of STEFLUX outputs and observations at Plateau Rosa (3480 m a.s.l.), Campo Imperatore – Mt. Portella (2401 m a.s.l.)

and Mt. Curcio (1832 m a.s.l.). The comparison with CMN observations was successfully carried out in the NextDATA paper by Putero et al. (2016).

STEFLUX was applied at these measurement sites, by selecting different configurations for the input parameters. As input, STEFLUX requires the coordinates of the “target box” (i.e., a geographical area including the measurement site) to be specified, together with its top lid. Depending on the underlying topography, the “ideal” extension of the box might vary from site to site. In order to choose the best configuration, we performed some sensitivity studies on these parameters. We specified several “target boxes” centered at the station locations but with different horizontal (from  $1^\circ \times 1^\circ$  to  $3^\circ \times 3^\circ$ ) and vertical boundaries. The results are displayed in Figure 5, where the annual variation of SI frequency according to different settings is presented. The time periods of study differed according to the availability of atmospheric measurements at the stations: we chose 2006–2016 for PRS (a), 2012–2016 for CMP (b) and 2015 for CUR (c). Please note that CUR presents a more noisy behavior due to the relatively short time period considered. For all the measurement sites, a clear seasonal cycle for SI frequency is discernible with highest values during winter-spring and minima during summer.



**Figure 5.** Annual average variation of SI frequency at PRS (panel a), CMP (b) and CUR (c). Displayed are different configurations for the input parameters of STEFLUX, in the following form: 3deg660 indicates that the “target box” has a horizontal extension of  $3^\circ \times 3^\circ$  around the measurement site (located as center), and a vertical top lid of 660 hPa.

The differences between the different configurations are minimized from June to September for all the measurement sites, while the seasonal cycle is more evident for the simulations characterized by the “target box” with the highest top lid or the widest horizontal boundaries. In some way, it looks that the

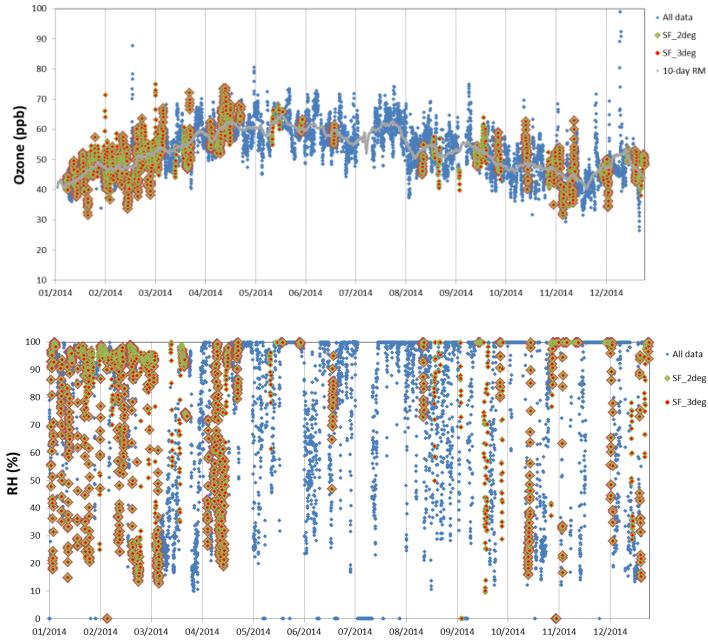
definition of the “target box” top lid and horizontal extension can compensate each other: increasing the top lid or increasing the horizontal extension lead to similar result on SI frequency variations.

To assess the STEFLUX performance in identifying period affected by SI at the NextDATA mountain stations, relative humidity (RH) and ozone ( $O_3$ ) observations were used to objectively identify SI events. According to Cui et al. (2009), enhancements over the 10-day running mean  $O_3$  value and 50% RH are taken as thresholds and are applied to  $O_3$  and RH measurements, respectively. If  $O_3$  values are continuously higher than the running mean by 10% and RH values are below 50% for a time period of at least 6 hours, SI events are identified by measurements (Cui2009\_RH+O3). Since, at these sites, a large fraction of days is characterized by the lack of simultaneous  $O_3$  and RH measurements, we also considered two less strict identification criteria, which only consider the RH (Cui2009\_RH) or the  $O_3$  (Cui2009\_O3). Even if this approach does not consider other well-known stratospheric tracers (e.g.,  $^{7}\text{Be}$ , or potential vorticity), its use in mountain sites is justified by Trickl et al. (2010).

Figures 6-8 report, for each measurement site, the time series (for selected years characterized by the highest data coverage) of  $O_3$ , RH and SI events as selected by STEFLUX. According to STEFLUX, depending on the “target box” configuration, SI events were observed for 19%-32% of time for PRS, 9%-17% for CMP and 4-12% for CUR, on a yearly basis. On the other side, according to the Cui2009\_O3 criterion, we identified 40 days (24% of the data-set) possibly affected by SI at CMP. According to the Cui2009\_RH criterion, we identified 154 days (48% of the data-set) at CUR and 861 days (49%) at PRS. Lastly, according to Cui2009\_RH+O3, we found 25 days (7%) at CUR and 187 days (11%) at PRS as possibly affected by SI.

In agreement with Cui et al. (2009) and Putero et al. (2016), a quantitative assessment of STEFLUX performance was accomplished by using the following approach: if during the period of a measured SI event at least one SI event is identified by the model simulation, we consider that STEFLUX well captures the measured SI event. We call this approach “forward comparison”. By means of this approach, we found that STEFLUX captured from 32% of Cui2009\_RH days to 14% of Cui2009\_RH+O3 days at PRS, from 22% to 3% of Cui2009\_O3 days at CMP, and from 15% of Cui2009\_RH days to 1% of Cui2009\_RH+O3 days at CUR (see Tables 8 -10). This indicated that the STEFLUX performance is decreasing with decreasing latitude and height of the measurement sites.

In a second step, the comparison between STEFLUX identified events and the measured events was conducted in a reverse way, namely “backward comparison”. We firstly collected SI days identified with different STEFLUX configurations (i.e., different extensions of the “target boxes”) and then we check the fractions of days also identified by the measurement criteria. For PRS, we found that the Cui2009\_RH+O3 (Cui2009\_RH) criterion confirmed from 8% to 10% (48% to 50%) of STEFLUX days. For CMP, Cui2009\_O3 confirmed from 24% to 29% of STEFLUX days. For CUR, we found that the Cui2009\_RH+O3 (Cui2009\_RH) criterion confirmed from 3% to 9% (32% to 38%) of STEFLUX days.

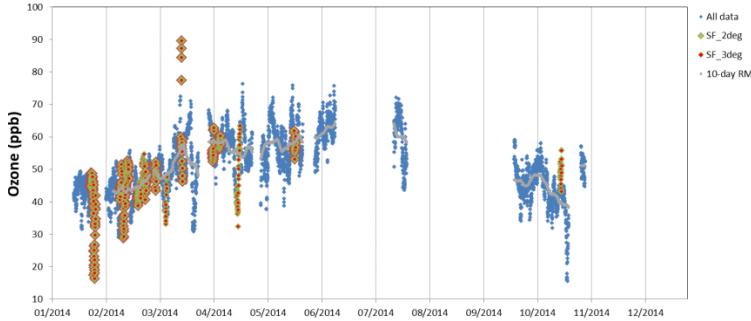


**Figure 6.** PRS, year 2014: measured  $O_3$  (top) and RH (bottom) with colored dots representing the match with different STEFLUX SI events (top lid @ 660 hPa, see legend for horizontal box extension). The  $O_3$  10-day running mean (top: grey line) is also reported.

**Table 8.** Upper plate: agreement (in number of days and in percentage) between measured SI events at PRS (based on different selection criteria) and STEFLUX outputs for different “target box” top lid (680 hPa and 660 hPa) and horizontal extensions ( $2^\circ \times 2^\circ$  and  $3^\circ \times 3^\circ$ ). Bottom plate: Agreement (in number of days and in percentage) between STEFLUX outputs for different “target box” top lid (680 hPa and 660 hPa) and horizontal extensions ( $2^\circ \times 2^\circ$  and  $3^\circ \times 3^\circ$ ) and measured SI events at PRS (based on different selection criteria).

Selection	Days	Cui2009_RH	Cui2009_RH+O3
SF_2deg_680hPa	331 (18.8%)	166 (50.1%)	28 (8.4%)
SF_2deg_660hPa	388 (22.0%)	196 (50.5%)	37 (9.5%)
SF_3deg_660hPa	572 (32.4 %)	278 (48.7%)	57 (10.0%)

Selection	Top lid	Days	SF_2deg_680	SF_2deg_660	SF_3deg_660
Cui2009_RH	680hPa	861 (48.8%)	167 (19.4%)	196 (22.7%)	279 (32.4%)
Cui2009_RH+O3	680hPa	187 (10.6%)	29 (15.5%)	37 (19.8%)	57 (30.5%)

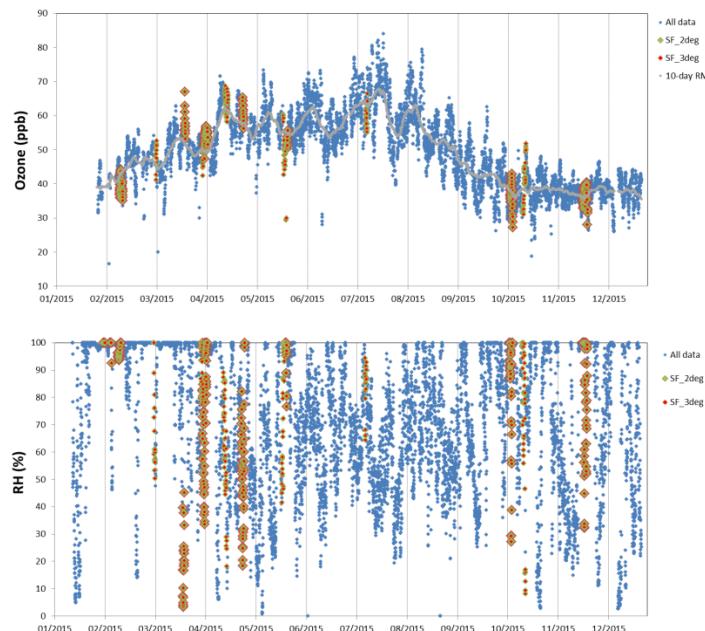


**Figure 7.** CMP, year 2014: measured  $O_3$  (top) and RH (bottom) with colored dots representing the match with different STEFLUX SI events (top lid @ 760 hPa, see legend for horizontal box extension). The  $O_3$  10-day running mean (top: grey line) is also reported.

**Table 9.** Upper plate: agreement (in number of days and in percentage) between measured SI events at CMP (based on different selection criteria) and STEFLUX outputs for different “target box” top lid (760 hPa and 700 hPa) and horizontal extensions ( $2^\circ \times 2^\circ$  and  $3^\circ \times 3^\circ$ ). Bottom plate: agreement (in number of days and in percentage) between STEFLUX outputs for different “target box” top lid (750 hPa and 700 hPa) and horizontal extensions ( $2^\circ \times 2^\circ$  and  $3^\circ \times 3^\circ$ ) and measured SI events at CMP (based on different selection criteria).

Selection	Days	Cui2009_RH
SF_2deg_760hPa	15 (8.9%)	4 (27%)
SF_3deg_760hPa	21 (12.6 %)	6 (29%)
SF_2deg_700hPa	29 (17.4%)	10 (24%)

Selection	Days	SF_2deg_760hPa	SF_3deg_760hPa	SF_2deg_700hPa
Cui2009_O3	40 (23.9%)	4 (3%)	6 (15%)	9 (22%)



**Figure 8.** CUR, year 2015: measured  $O_3$  (top) and RH (bottom) with colored dots representing the match with different STEFLUX SI events (top lid @ 820 hPa, see legend for horizontal box extension). The  $O_3$  10-day running mean (top: grey line) is also reported.

**Table 10.** Upper plate: Agreement (in number of days and in percentage) between measured SI events at CUR (based on different selection criteria) and STEFLUX outputs for different “target box” top lid (820 hPa and 770 hPa) and horizontal extensions (2°x2° and 3°x3°). Bottom plate: agreement (in number of days and in percentage) between STEFLUX outputs for different “target box” top lid (820 hPa and 770 hPa) and horizontal extensions (2°x2° and 3°x3°) and measured SI events at CUR (based on different selection criteria).

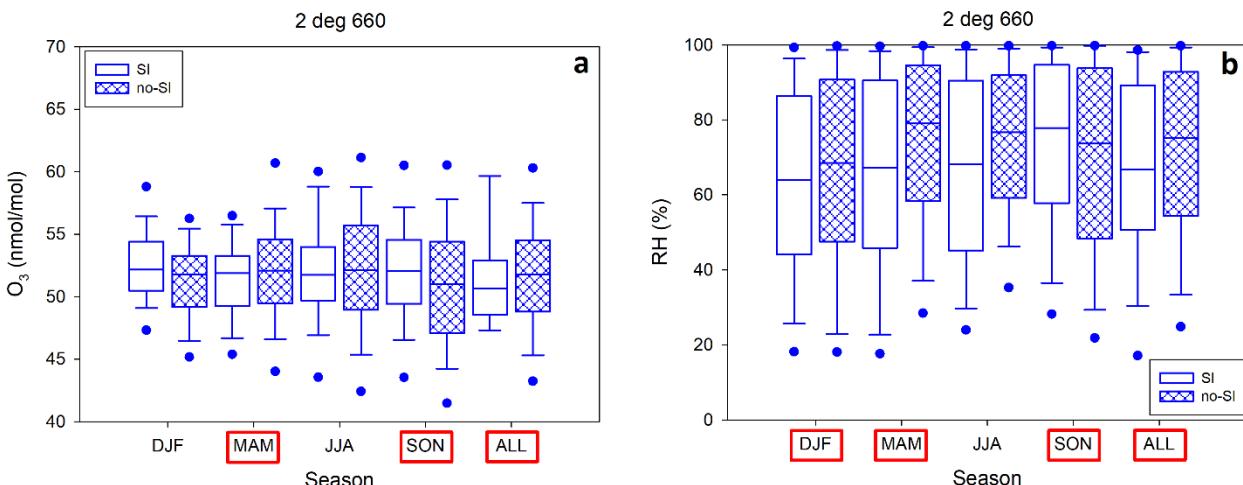
Selection	Days	Cui2009_RH	Cui2009_RH+O3
SF_2deg_820hPa	13 (4.1%)	5 (38%)	1 (7%)
SF_3deg_820hPa	22 (6.9 %)	8 (36%)	2 (9%)
SF_2deg_770hPa	28 (7.4%)	9 (32%)	1 (3%)
SF_3deg_770hPa	42 (12.0 %)	15 (36%)	2 (5%)

Selection	Top lid	Days	SF_2deg	SF_3deg
Cui2009_RH	820hPa	154 (48.5%)	5 (3.2%)	8 (5.2%)
Cui2009_RH+O3	820hPa	25 (7.1%)	1 (4%)	2 (8%)
Cui2009_RH	770hPa	154 (4.4%)	9 (5.8%)	15 (9.7%)
Cui2009_RH+O3	770hPa	25 (7.1%)	1 (4%)	2 (8%)

Basing on these results, we can summarize that STEFLUX showed satisfactory performances for PRS, which are somewhat comparable with the results obtained by Cui et al. (2009) by applying Lagrangian models for detecting SI at the high alpine site Jungfraujoch (Switzerland). STEFLUX performances appeared to decrease moving towards southern and lower measurement sites. However, in the latter case, it must be considered that relatively short time series were available for the assessment. Moreover, it must be noted that we adopted a more stringent metric for a positive SI detection than Cui et al. (2009) or Putero et al. (2016). Indeed, to have a positive detection, STEFLUX and measurement criteria must identify SI on the very same days of the record (both Cui et al., 2009 and Putero et al., 2016 adopted a less stringent approach). Basing on the “forward” approach, we defined the following settings as the best for detecting SI with STEFLUX:

- PRS: SF\_2deg\_660hPa
- CMP: SF\_2deg\_700hPa
- CUR: SF\_3deg\_820hPa

By considering STEFLUX as the tool to discriminate between SI and no-SI events, we evaluated the seasonal box-and-whisker plot for O<sub>3</sub> and RH for days affected or not by SI for the longest available time series, i.e. PRS (Fig. 8). In particular, we considered the STEFLUX outputs obtained by considering the “optimal” configuration for PRS presented above (SF\_2deg\_660hPa). Since both O<sub>3</sub> and RH are characterized by a different seasonality, their seasonality cycles were removed from the two datasets. For O<sub>3</sub>, statistically significant differences between SI and no-SI concentrations (computed by a Mann-Whitney “U test”) were found in spring, autumn and considering the entire dataset. It is interesting noting that not all differences imply higher O<sub>3</sub> mixing ratios during SI days; despite the limited measurements available, this feature needs to be investigated in deeper detail, taking also into account other factors that may drive the O<sub>3</sub> variation in conjunction with SI events. On the other hand, SI events are likely to decrease the RH levels on average, except for autumn (in which the increase during SI days is statistically significant). Indeed, statistically significant decreases are observed in winter, spring and considering all of the data, which is a strong hint towards actual occurrence of SI. It should be considered that RH observations at PRS on 2006-2007 present a different seasonality with respect to the following years, suggesting possible problems in the observations.



**Figure 9.** Seasonal box-and-whisker plot for  $O_3$  (panel a) and RH (b), for days affected or not by SI at PRS. The red boxes indicate those seasons in which the differences are statistically significant.

## References

- Collaud Coen, M., Weingartner, E., Schaub, D., Hueglin, C., Corrigan, C., Henning, S., Schwikowski, M., and Baltensperger, U., 2004. Saharan dust events at the Jungfraujoch: detection by wavelength dependence of the single scattering albedo and first climatology analysis. *Atmos. Chem. Phys.*, 4, 2465–2480.
- Cui, J., Sprenger, M., Staehelin, J., Siegrist, A., Kunz, M., Henne, S., and Steinbacher, M., 2009. Impact of stratospheric intrusions and intercontinental transport on ozone at Jungfraujoch in 2005: comparison and validation of two Lagrangian approaches, *Atmos. Chem. Phys.*, 9, 3371–3383, doi:10.5194/acp-9-3371-2009.
- Duchi, R., Cristofanelli, P., Landi, T. C., Arduini, J., Bonafè, U., Bourcier, L., Busetto, M., Calzolari, F., Marinoni, A., Putero, D., and Bonasoni, P., 2016. Long-term (2002–2012) investigation of Saharan dust transport events at Mt. Cimone GAW global station, Italy (2165 m a.s.l.). *Elementa*, 4, 000085, doi:10.12952/journal.elementa.000085.
- GAW, 2011. Expert Workshop on Global Long-term Measurements of Nitrogen Oxides and Recommendations for GAW Nitrogen Oxides Network, Hohenpeissenberg Meteorological Observatory (Germany, October 2009).
- Putero, D., Cristofanelli, P., Sprenger, M., Škerlak, B., Tositti, L., and Bonasoni, P., 2016. STEFLUX, a tool for investigating stratospheric intrusions: application to two WMO/GAW global stations. *Atmos. Chem. Phys.*, 16, 14203–14217.
- Škerlak, B., Sprenger, M., and Wernli, H., 2014. A global climatology of stratosphere-to-troposphere exchange using the ERA-Interim data set from 1979 to 2011. *Atmos. Chem. Phys.*, 14.
- Stein, A.F., and Ngna, F., 2015. NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. doi.org/10.1175/BAMS-D-14-00110.1
- Steinbacher, M., C. Zellweger, B. Schwarzenbach, S. Bugmann, B. Buchmann, C. Ordóñez, A. S. H. Prevot, and C. Hueglin, 2007. Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, *J. Geophys. Res.*, 112, D11307, doi:10.1029/2006JD007971.
- Thornes, J. E. and Stephenson, D. B., 2001. How to judge the quality and value of weather forecast products. *Meteorol. Appl.*, 8, 307–314.
- Trickl, T., Feldmann, H., Kanter, H.-J., Scheel, H.-E., Sprenger, M., Stohl, A., and Wernli, H., 2010. Forecasted deep stratospheric intrusions over Central Europe: case studies and climatologies, *Atmos. Chem. Phys.*, 10, 499–524, doi:10.5194/acp-10-499-2010.

Wilks, D. S., 2006. Statistical methods in the atmospheric sciences: an introduction, 2nd Edn., Academic press.