



## Research paper



## Increasing the maturity of measurements of essential climate variables (ECVs) at Italian atmospheric WMO/GAW observatories by implementing automated data elaboration chains

Luca Naitza <sup>a</sup>, Paolo Cristofanelli <sup>a,\*</sup>, Angela Marinoni <sup>a</sup>, Francescopiero Calzolari <sup>a</sup>, Fabrizio Roccatò <sup>a</sup>, Maurizio Busetto <sup>a</sup>, Damiano Sferlazzo <sup>b</sup>, Eleonora Aruffo <sup>c</sup>, Piero Di Carlo <sup>c</sup>, Mariantonia Bencardino <sup>d</sup>, Francesco D'Amore <sup>d</sup>, Francesca Sprovieri <sup>d</sup>, Nicola Pirrone <sup>d</sup>, Federico Dallo <sup>e</sup>, Jacopo Gabrieli <sup>e</sup>, Massimiliano Vardè <sup>e,f</sup>, Giorgio Resci <sup>g</sup>, Carlo Barbante <sup>e</sup>, Paolo Bonasoni <sup>a</sup>, Davide Putero <sup>a</sup>

<sup>a</sup> CNR-ISAC, National Research Council – Institute of Atmospheric Sciences and Climate, Bologna, Italy

<sup>b</sup> ENEA, SSPT-PROTER-OAC, Lampedusa, Italy

<sup>c</sup> University “G. d’Annunzio” of Chieti-Pescara, Chieti, Italy

<sup>d</sup> CNR-IRA, National Research Council – Institute of Atmospheric Pollution, Rende, Italy

<sup>e</sup> CNR-ISP, National Research Council – Institute of Polar Sciences, Venezia-Mestre, Italy

<sup>f</sup> Department of Chemical and Pharmaceutical Sciences, Ferrara University, Ferrara, Italy

<sup>g</sup> Inkode Soc. Coop., Bologna, Italy

## ARTICLE INFO

## Keywords:

Data analysis

Data flagging

Atmosphere

Essential climate variables

Automatic processing

## ABSTRACT

In the framework of the National Project of Interest NextData, we developed automatic procedures for the flagging and formatting of trace gases, atmospheric aerosols and meteorological data to be submitted to the World Data Centers (WDCs) of the Global Atmosphere Watch program of the World Meteorological Organization (WMO/GAW). In particular, the atmospheric Essential Climate Variables (ECVs) covered in this work are observations of near-surface trace gas concentrations, aerosol properties and meteorological variables, which are under the umbrella of the World Data Center for Greenhouse Gases (WDCGG), the World Data Center for Reactive Gases, and the World Data Center for Aerosol (WDCRG and WDCA). We developed an overarching processing chain to create a number of data products (data files and reports) starting from the raw data, finally contributing to increase the maturity of these measurements. To this aim, we implemented specific routines for data filtering, flagging, format harmonization, and creation of data products, useful for detecting instrumental problems, particular atmospheric events and quick data dissemination towards stakeholders or citizens. Currently, the automatic data processing is active for a subset of ECVs at 5 measurement sites in Italy. The system represents a valuable tool to facilitate data originators towards a more efficient data production. Our effort is expected to accelerate the process of data submission to WMO/GAW or to other reference data centers or repositories. Moreover, the adoption of automatic procedures for data flagging and data correction allows to keep track of the process that led to the final validated data, and makes data evaluation and revisions more efficient by improving the traceability of the data production process.

## 1. Introduction

The monitoring of trace atmospheric constituents in the lower troposphere still represents a fundamental activity to assess their long- and short-term variability, to investigate the underlying processes and to assess the impact of natural and anthropogenic sources into the atmosphere. The Global Atmosphere Watch program of the World

Meteorological Organization (WMO/GAW) coordinates a global network of surface stations to understand and control the increasing influence of human activity on the global atmosphere (WMO, 2017). Atmospheric stations belonging to the WMO/GAW network are requested to adopt standard operating procedures (SOPs), to perform quality assurance/quality check (QA/QC) actions and routinely submit data to specific World Data Centers (WDCs) covering 6 focal areas (i.e.,

\* Corresponding author.

E-mail address: [P.Cristofanelli@isac.cnr.it](mailto:P.Cristofanelli@isac.cnr.it) (P. Cristofanelli).

<https://doi.org/10.1016/j.cageo.2020.104432>

Received 2 May 2019; Received in revised form 21 January 2020; Accepted 31 January 2020

Available online 6 February 2020

0098-3004/© 2020 Elsevier Ltd. All rights reserved.



Fig. 1. Geographical location and pictures of the monitoring stations considered in this work.

atmospheric aerosols, greenhouse gases, selected reactive gases, ozone, UV radiation, precipitation, and chemistry). In Italy, the NextData project was aimed at creating a network in mountain and remote areas, based on atmospheric observatories for the monitoring of atmospheric composition and ancillary data (meteorological parameters and solar radiation). The main goal of this network is to investigate the processes which influence the variability of air pollutants and climate-altering compounds to contribute towards a better assessment of the related impacts to mountain ecosystems and climate in the Mediterranean basin. The network comprises five high-mountain atmospheric observatories (Fig. 1): Monte Cimone (CMN, northern Apennines, the only WMO/GAW global station in Italy; 2165 m a.s.l.), Plateau Rosa (PRS, western Alps, WMO/GAW regional station; 3480 m a.s.l.), Col Margherita (MRG, eastern Alps; 2550 m a.s.l.), Monte Portella–Campo Imperatore (CMP, central Apennines; 2401 m a.s.l.), and Monte Curcio (CUR, southern Apennines, WMO/GAW regional station; 1796 m a.s.l.). In addition to these observatories, the WMO/GAW regional stations Capo Granitola (CGR, south-western Sicily) and Lampedusa (LMP, central Mediterranean Sea), provide complementary information on the background conditions of the Mediterranean basin marine boundary layer (Fig. 1).

Six of these observatories are already part of international projects/research programs for the monitoring of near-surface atmospheric Essential Climate Variables (ECVs). In agreement with WMO (2016), an ECV is “a physical, chemical or biological variable or a group of linked variables that critically contributes to the characterization of Earth’s climate”. The Global Climate Observing System (GCOS) program of the WMO currently specifies 54 ECVs. In the framework of NextData, “surface” (i.e., meteorological parameters) and “atmospheric composition” (i.e., trace gases and aerosol properties) ECVs are monitored at the atmospheric observatories. More specifically, in the framework of the WMO/GAW activities, observations of greenhouse and reactive gases are carried out at PRS, CMN, CUR, CGR and LMP (Table 1). Measurements of physical properties of atmospheric aerosols are performed at CMN, CUR, and CGR (Table 1).

To contribute to the implementation of this national network, one action carried out by NextData was to set up a suite of routines for

the automatic processing of the near-surface atmospheric data observed at these measurement sites. The final goal of this activity is to increase the overall maturity of the measurements. As defined in Thorne et al. (2017), with the term “maturity” we mean the capacity of a measurement to meet different best practices. Among the different categories identified in Thorne et al. (2017), the system we developed can contribute to increase the maturity in terms of metadata, public access and update, usage, sustainability and software. The measurement sites will be supported towards an optimization and harmonization of the data creation process, which is a pre-requisite for a fast and efficient data publication. Indeed, due to the large amount of data recorded at the measurement sites, it is not efficient to perform the data validation/flagging by the so-called “visual inspection” and by manual manipulation of data files. Moreover, there is a widely recognized need for the provision of ECVs in near-real time or real-time mode for a number of applications (data assimilation, atmospheric model verification, early warning systems, see, e.g., Wagner et al., 2015), which imply the delivery of quality-assessed data with specific common data models.

Reference examples of systems for the automatic execution of QA/QC activities in the framework of atmospheric composition observations can be found within the Integrated Carbon Observation System (ICOS) and Global Mercury Observation System (GMOS) initiatives. As described by Hazan et al. (2016), the former is related to the evaluation and quality control (EQC) of a relevant subset of atmospheric parameters related to the carbon cycle investigation (i.e., CO, CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, and meteorological parameters), while the latter implemented a web-based system for QA/QC to check raw data related to atmospheric mercury (D’Amore et al., 2015). In some ways, our system got inspiration from some concepts already developed in these previous initiatives. In particular, as following the ICOS-RI approach, a suite of standard “data products” (i.e., a suite of plots able to provide control and report about the different monitoring programs) are automatically produced and available to the data providers through a web-based interface and a dedicated SFTP server. However, differently from those initiatives (which only consider data from a specific subset of

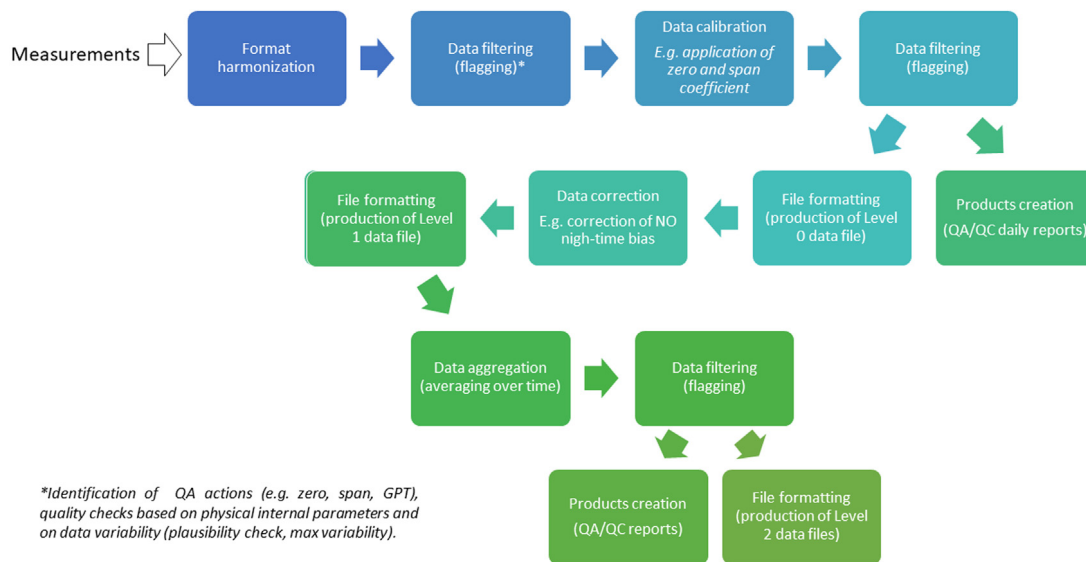


Fig. 2. Workflow of automatic data processing for Thermo 42i-TL.

Table 1

List of Essential Climate Variables (ECVs) collected at the atmospheric observatories part of NextData. Bold characters indicate the ECVs for which the automatic processing chain is already active.

ECV	Reference programs	Monitoring stations						
		PRS	MRG	CMN	CMP	CUR	CGR	LMP
CO <sub>2</sub>	WMO/GAW	X		X		X	X	
CH <sub>4</sub>		X		X		X	X	
CO				X		X	X	
O <sub>3</sub>		X	X	X	X	X	X	X
SO <sub>2</sub>				X			X	X
NO	WMO/GAW, ACTRIS			X		X	X	
NO <sub>2</sub>				X		X	X	
Particle scattering				X		X	X	
Particle absorption				X		X	X	
Particle size distribution (by SMPS)				X		X	X	
Particle number concentration				X		X	X	
Coarse particle size distribution (by OPC)				X		X	X	
Aerosol Optical Depth (AOD) <sup>a</sup>				X				
Meteorological parameters	WMO/GAW	X	X	X	X	X	X	X
Solar radiation			X	X	X	X	X	X

<sup>a</sup>AOD measurements are processed in the framework of AERONET and GAW-NRT programs.

instruments and measurement sites), our system is based on a suite of “R” (R Development Core Team, 2008) routines that can be adapted to different experimental setup and that can be run independently by each single data provider, still ensuring traceability of the data creation process and interoperability as well as reusability of the developed routines.

For these reasons, we hereby developed automatic procedures for ECVs data flagging, averaging, and formatting. Currently, the automatic processing of data is active for a subset of ECVs and measurement sites (see Table 2). In particular, the CMN observatory (Cristofanelli et al., 2018) was selected as a “proof-of-concept site”, due to the large number of ECVs observed and the large variety of data formats produced by the measurement systems. Then, the procedures developed for CMN have been adapted to other atmospheric observatories included in this exercise: currently, the automatic EQC process is fully active at five observatories (CMN, MRG, CMP, CGR, and LMP). The automatic data processing encompasses a preliminary harmonization of file formats, which is the pre-requisite for the subsequent data flagging, data aggregation (to common temporal frames: 1 and 60 min), and final harmonization of files according to WMO/GAW WDCs common data models. The automatic data processing also encompasses the creation of many data reports (updated daily), which provide 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 events of interest, or to perform preliminary data analysis).

## 2. Description of the automatic data processing

Specific procedures have been developed for each ECV and for specific instruments (i.e., one processing chain for each instrument type) used at the considered stations: Table 2 summarizes the list of these specific instruments, together with the formats used for file creation. All the procedures have been implemented in “R” Language and Environment for Statistical Computing (R Development Core Team, 2008). It is important to note that the NextData routines, even if implemented for a specific suite of instruments, can be virtually adapted to other instruments using the same principle of operation/detection.

To increase the inter-operability of the automatic procedures, the sequential steps of the workflow are the same for all the different ECVs, and specific modules or functions have been developed to be inserted in the programming code to increase flexibility of usage (see also Fig. 2):

1. data file collection from observatories;

**Table 2**

Summary of automated procedures developed for the processing of ECVs as a function of specific instruments and data file format for submission to WMO/GAW WDCs. For explanation of “data file formats” see Section 2.5.

Class	ECV	Instruments	Data file format
Trace gases (near-surface)	O <sub>3</sub>	Thermo 49i, Thermo49c	NASA-AMES
	NO, NO <sub>2</sub>	Thermo 42i-TL	NASA-AMES
	SO <sub>2</sub>	Thermo 42i-TL	NASA-AMES
	<sup>a</sup> CO	Picarro G2401	ASCII (WDCGG)
	<sup>a</sup> CO <sub>2</sub> , CH <sub>4</sub>	Picarro G2401	ASCII (WDCGG)
Aerosol (near-surface)	Absorption coefficient	MAAP 5012	NASA-AMES
	Scattering coefficient	TSI 3563	NASA-AMES
	Particle number concentration	TSI 3775	NASA-AMES
	Size distribution	Grimm 1.108	NASA-AMES
Meteorology	T, RH, P, WD, WS, Global and UV radiation	Various	NASA-AMES

<sup>a</sup>Automatic processing not operative yet.

2. data formatting of raw files coming from stations;
3. data check and flagging;
4. data correction (if needed);
5. data aggregation (time averaging) and flagging;
6. data formatting;
7. creation of data reports, including the “health status report”.

By this 7-steps process, three different data levels are produced according to WMO/GAW WDCRG and WDCA data reporting guidelines (see also <https://ebas-submit.nilu.no/Submit-Data/Data-Reporting>):

- Level-0: annotated raw data; contains all parameters and variables provided by the instrument; contains all parameters/info needed for processing to final value; “native” time resolution;
- Level-1: data processed to final variable (calibration and correction implemented to data series), invalid data and calibration episodes removed, “native” time resolution, correction to standard temperature and pressure (i.e., 273.15 K, 1013.25 hPa) if necessary;
- Level-2: data aggregated to hourly averages, atmospheric variability quantified by standard deviation or percentiles.

In the current prototype system, the automatic processing is executed by a set of four scripts specifically designed for each instrument/ECV and for each measurement site:

- “P10” is the script devoted to the formatting/homogenization of raw data files (step 2 of the aforementioned workflow);
- “P20” is the script devoted to the production of Level-0 data files (step 3);
- “P21” is the script devoted to the production of Level-1 and Level-2 data files (steps 4, 5, and 6);
- “P22” is the script devoted to the generation of data reports (step 7).

These scripts are run sequentially based on the above order list, starting at 02:00 UTC (Coordinated Universal Time) of each calendar day. For each observatory and ECV, the elaboration is started by executing the “P10” scripts. Then, after an appropriate time lag, “P20”, “P21” and, finally, “P22” scripts are run. Each day, all data since the beginning of the current year (i.e., January, 1st) are processed. In the current configuration, a total of 28 near-surface atmospheric ECVs from 5 observatories, the processing chain is completed in 4 h (most of the time is used as sufficient lag periods between the scripts). To provide an explanatory example, we will describe the chain and the routines developed for the Thermo 42i-TL, a state-of-art instrument for the continuous determination of nitrogen monoxide (NO) and nitrogen dioxide (NO<sub>2</sub>) mixing ratios, based on chemiluminescence detection (CLD) and equipped with a blue light converter (BLC). This instrument belongs to a class of instruments (Thermo “i-series”) widely diffused among the NextData measurement stations for trace gas measurements and, due to the complexity of steps necessary to obtain the final data, it represents an effective case study for implementing QA/QC routines. Fig. 2 reports the different steps of the data processing for this analyzer.

### 2.1. Automatic processing of ECVs: prototype of a centralized system

All the developed routines are virtually stand-alone and any hypothetical user, after installing “R”, which is a free software environment for computing and graphics, can use them on his/her own PC or server, for both automatic and on-demand applications. Thus, our software is characterized by a good level of portability, useful, as an instance, for migrating or installing to different computer systems. Each script is accompanied with a header where basic instructions for installation, usage and modification, along with code update history, are provided.

While, as detailed below, the delivery of data files from the atmospheric observatories must be performed by observatory personnel/institutions, the automatic operation of the routines involves a small group of people at the CNR-ISAC HQs in Bologna, including an IT expert and two routine developers. On a short to medium time scale this would assure the scientific and expert operational actions, as well as the programmatic support to underpin the system. In case the raw data formats are not changed, the system is expected to be robust and frequent technical intervention is not needed. The interrupted or delayed data flow from the observatories does not represent an issue, since the routines are designed to run on each calendar day, and to process all the synchronized available data files in the current year.

To check the correct execution of the routines, a specific product (called “health status report”) was designed to be accessed by measurement PIs or CNR-ISAC personnel. This product is generated on a daily basis and the plots indicate, for each observatory, the correct execution of elaboration routines (see Fig. 3 for an example). For each observatory and for each instrument, a block containing the name of the routine that worked successfully is drawn. In case a routine encountered some issues that prevented its complete execution (as for “P22” routine for SO<sub>2</sub> at CMN in Fig. 3), the related block is not plotted.

We rely on the strong assumption that the responsibility for the production of “final” consolidated data files (to be submitted to reference WMO/GAW data centers or simply to be published for external usage) totally relies on the measurement PIs. At the moment of data submission or publication, PIs are expected to review the result of the automatic flagging produced by the routines at the lowest data level (i.e., Level-0) and they can accept or modify the produced file. In the latter case, it is recommended that a new file version is created (the NASA-AMES format requires to declare the revision number as well as the date creation of each data file) and that, based on the revised (consolidated) Level-0 file, new versions of Level-1 and Level-2 files are created using “P20” and “P21” routines.

### 2.2. Automatic processing of ECVs: data collection and formatting of raw files

The instrumental raw data are transferred (at least) once a day from the measurement stations to a server located at CNR-ISAC HQs in Bologna. The files stored in the CNR-ISAC server already contain information of the measured quantities in geophysical units, as well

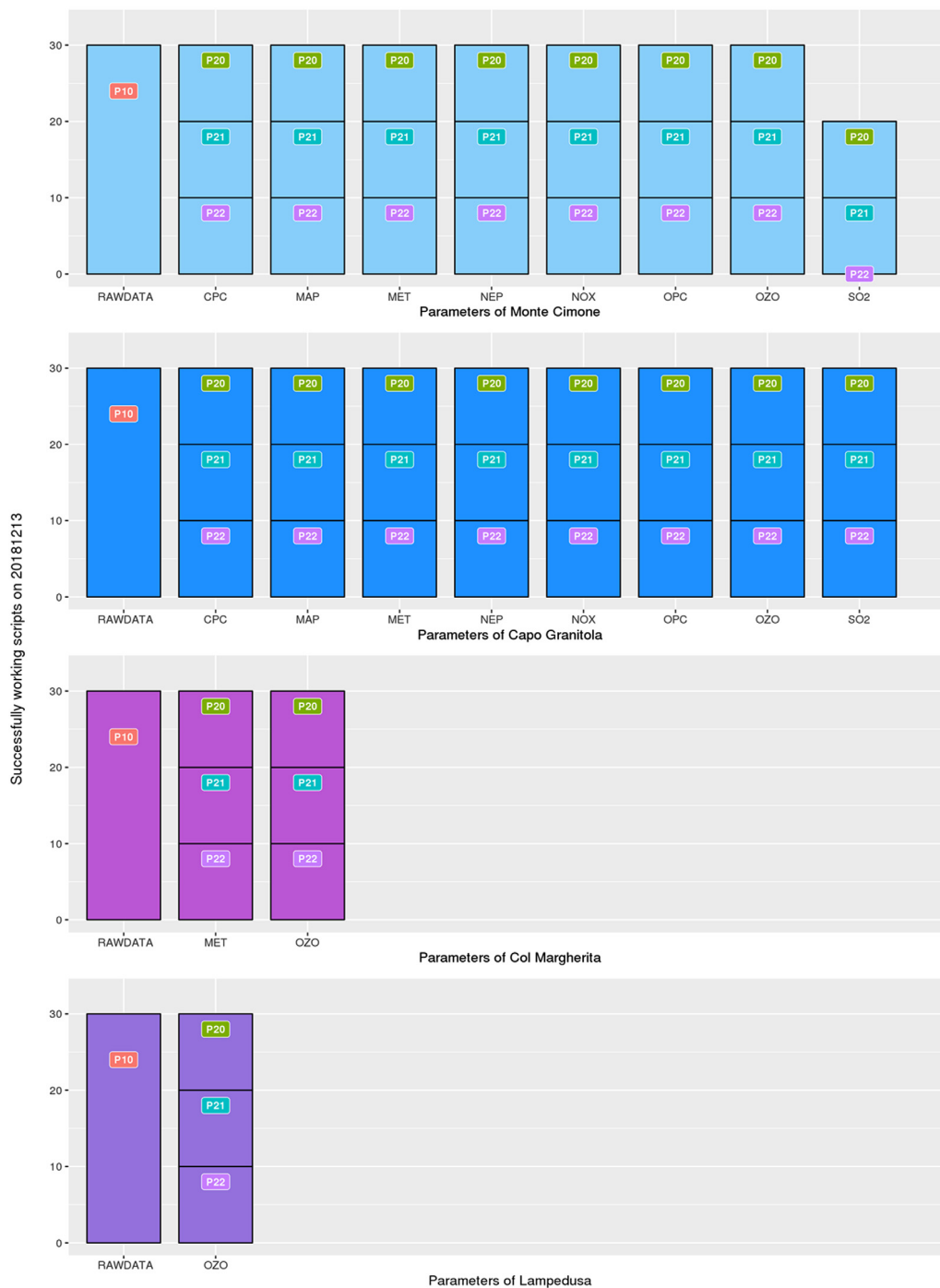


Fig. 3. Example of “health status report” (see Section 2.1) describing the execution status of the EQC routines for 2018, December 13th (the date is reported on the y-axis). For the abbreviations of the routines see Section 2.

as internal diagnostic parameters used for automatic QA/QC. During the development phase, different transfer strategies were applied, as a function of the specific requirements of the measurement stations. For instance, CMN data files are downloaded from the station server, while MRG data files are uploaded by the station server to the CNR-ISAC server: to facilitate the participation of the stations we let each station decide which strategy to adopt for data transfer during this development phase.

All the raw data files are centralized to the CNR-ISAC server as a function of their origin station, ECV class (trace gases, aerosol properties, meteorology) and specific ECV (see Table 3). Then, they are processed by the “P10” scripts to obtain a homogeneous file system

in terms of nomenclature and format, which is read by the processing chain. For the file name, we adopted the following coding:

$$SSS\_PPP\_YYYYMMDD\_TTT.dat \tag{1}$$

where SSS is the station code, PPP is the ECV code (see Table 3 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-min: “01M”, 30-min: “30M”, 60-min: “60M”). These homogenized data files refer to UTC, and they have a common structure for what concerns date and time: year (YYYY), month (MM), day (DD), hour (HH), minute

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

ECV class	ECV	Code
Trace gases	Ozone	OZO
	Carbon monoxide	CO
	Sulfur dioxide	SO2
	Nitrogen oxides	NO
Aerosol	Scattering coefficient	NEPH
	Absorption coefficient	MAAP
	Size distribution (by OPC)	OPC
	Total particle number concentration	CPC
Meteorology	T, RH, P, WD, WS	METEO
	Solar radiation	RAD_SOL

**Table 4**  
Example of descriptive flags (following the NASA-AMES format).

Flag	Description
0.000	Valid measurements
0.147	Under detection limit (but considered valid)
0.390	Less than 50% of data used for data averaging (used for Level-2 data)
0.440	Derived value (corrected for night-time offset, valid)
0.456	Invalidated by data originator (invalid)
0.459	Extreme value, unspecified error (invalid)
0.640	Instrument internal relative humidity above 40% (valid)
0.664	Low/high sampling flow, chamber pressure, PMT cooling (invalid)
0.682	Calibration or zero/span check (invalid)
0.699	Mechanical problem, unspecified error (invalid)
0.999	Missing data (invalid)

(MIN) and the decimal date for each measurement (DEC\_DATE), considered as fraction of time from the reference point (i.e., January 1st of each year, at 00:00 UTC).

### 2.3. Automatic processing of ECVs: data check and flagging

The data inspection consists of a series of checks that are automatically performed on data with native time resolution (usually 1-min), with the aim of referring each data record to a set of codified flags (that indicate whether a measurement is valid or not). Table 4 reports the list of flags adopted for NO<sub>x</sub> measurements, as defined in the framework of ACTRIS-2 project (see <https://ebas-submit.nilu.no/Submit-Data/Data-Reporting/Templates/Category/Trace-Gases/NOx>).

The first step of the automatic data check consists in the identification of the different measurement modes: “sample” (when ambient air is measured), “calibration” (when air from one or more laboratory standards is measured), “zero” (when the molecules to be quantified are scrubbed from the gas mixture, 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 point out changes in the instrumental sensitivity). Systematic variations with time of the zero offset 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, and to verify measurement linearity. Depending on the ECV and the instrument, the zero source can be either tanks with a dry gas mixture or a generator producing clean dry air, while span sources can be a tank with assigned mixing ratios for a specific chemical, a permeation tube with a precise and stable emitting rate for a specific chemical, or an internal generator (like a UV-lamp in case of O<sub>3</sub>). The identification of the measurement mode, which leads to the attribution of a specific flag to the concurrent data record, is fundamental for two reasons: (i) data affected by calibrations or quality checks must be discarded during time averaging processes, and (ii) data recorded during calibrations or quality checks are used to obtain correction factors or quality control metrics. The measurement periods affected by “calibrations” or “zero/span” checks are selected

considering two general cases: (i) by analyzing the variability of internal diagnostic parameters of analyzers or external “calibration units” (e.g., some instruments provide the information related to their current “mode” to the acquisition system), or (ii) by searching the existence of a log-file that indicates the start and end times of QC exercises.

The second step is the analysis of the variability of instrumental data (both related to internal diagnostic parameters and to the measured ECV). 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 compared with their typical ranges, which are reported on the instrument manuals. For each measurement, if at least one parameter fails a check, the data is flagged as invalid (see, e.g., Table 5 for NO and NO<sub>2</sub> threshold values at CMN). Moreover, also mechanical problems of different nature (see, e.g., Ni et al., 2009, for a list of possible sensor failures) are flagged at this stage;
- Plausibility checks: the measurement periods with values exceeding the expected variability are selected and identified as “outliers”. The allowed variability ranges are defined as a function of the measurement stations (e.g., the plausible range for atmospheric pressure at a surface station like CGR is different from that of a mountain station like MRG). Currently, two different processes for the definition of these ranges are considered. The first one is the adoption of fixed 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 based on statistical data analysis over specific time periods (1 h, 1 day, 1 month, or a full year), e.g., percentiles of the data-set or confidence intervals like  $n$ -times the standard deviation above or below the average values. In this context, specific calculations have been implemented to detect data outliers for the aerosol ECVs (i.e., NEPH, MAAP, OPC, CPC) and for the NO<sub>x</sub>, following the approaches suggested by Huang et al. (2016) and El Yazidi et al. (2018), see the Supplementary Material;
- 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 characteristics of each site, a range for maximum ECV variability is defined (typically on hourly basis, see, e.g., Table 5 for NO and NO<sub>2</sub> threshold values at CMN).

### 2.4. Automatic processing of ECVs: data calibration and correction (Level-1 production)

As specified in Section 2, Level-1 refers to the data set containing only “valid” records with calibration and corrections applied (i.e., data with “not-valid” or “calibration” numflags are removed).

For NO<sub>x</sub>, the major difficulty in this task is related to the use of the information provided by automatic calibration in the processing chain (zero and span). CLD instruments directly detect and quantify only NO; therefore, it is necessary to convert NO<sub>2</sub> to NO to quantify NO<sub>x</sub> (and finally NO<sub>2</sub>). The 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) in rural/remote locations, 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, the CMN instrument is equipped with a photolytic converter (Blue Light Converter, Teledyne, USA), which uses an UV light source to selectively convert NO<sub>2</sub> to NO. Since for some instruments BLC conversion efficiency (Sc) can be significantly lower than 100%, it is paramount to derive the actual value of Sc. The Sc obtained by a gas phase titration carried

**Table 5**

List of parameters used for data control of Thermo 42-iTL instrument at CMN, with defined plausible threshold values. Threshold values for plausibility and variability checks vary as a function of the measurement site. The complete list of threshold values for different instruments and measurement sites are provided in the Supplementary Material.

Parameter	Variability range	Application	Numflag or action
Sample flow	0.5–1.5 l/min	Diagnostic/instrumental checks	0.664 (for data outside the range)
Chamber pressure	200–400 mmHg	Diagnostic/instrumental checks	0.664 (for data outside the range)
PMT cooler	–40–10 °C	Diagnostic/instrumental checks	0.664 (for data outside the range)
Zero offset	–0.5–0.5 ppb	Diagnostic/instrumental checks	Calibration fault → the previous good calibration is used
Span coeff.	0.90–1.10	Diagnostic/instrumental checks	Calibration fault → the previous good calibration is used
Sc	0.1–1.0	Diagnostic/instrumental checks	Calibration fault → the previous good calibration is used
NO variability	NO(i)–NO(i+1) <0.5 ppb	Variability checks	0.456
NO <sub>2</sub> variability	NO <sub>2</sub> (i)–NO <sub>2</sub> (i+1) < 0.5 ppb	Variability checks	0.456
NO	<–20 ppb	Plausibility checks	0.459
NO <sub>2</sub>	<–20 ppb	Plausibility checks	0.459
NO	<0.05 ppb	Detection limit	0.147
NO <sub>2</sub>	<0.05 ppb	Detection limit	0.147

out during the calibration is then used to correct the NO<sub>2</sub> reading and to obtain the actual NO<sub>2</sub>. The calibration process is composed of the following steps:

1. Sampling of zero air: the NO reading is used to calculate instrumental zero-offset (bkg\_NO, bkg\_NO<sub>x</sub>);
2. Sampling of span air (obtained by diluting 5 ppm of NO in N<sub>2</sub> to about 100 ppb): the NO and NO<sub>x</sub> readings are used to calculate the NO and NO<sub>x</sub> span factors (coeff\_NO, coeff\_NO<sub>x</sub>) for obtaining corrected NO (NO<sub>x</sub>\_elab1) and NO<sub>x</sub> (NO<sub>x</sub>\_elab1);
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 about 80% of the NO is titrated (gas phase titration — GPT);
4. After stabilization, the data for NO (NO<sub>x</sub>\_elab1\_gpt) and NO<sub>x</sub> (NO<sub>x</sub>\_elab1\_gpt) are recorded and used to calculate converter efficiency (Sc, see below);
5. Sampling of zero air to purge the instrument after calibration.

Then, Sc is calculated as follows. The effective NO<sub>2</sub> amount produced by GPT results from:

$$\text{NO}_2 = \text{NO}_{\text{elab1\_span}} - \text{NO}_{\text{elab1\_gpt}} \quad (2)$$

The NO<sub>2</sub> amount converted by BLC is calculated by:

$$(\text{NO}_{\text{x\_elab1\_gpt}} - \text{NO}_{\text{elab1\_gpt}}) - (\text{NO}_{\text{x\_elab1\_span}} - \text{NO}_{\text{elab1\_span}}) \quad (3)$$

Accordingly, Sc is calculated by:

$$\text{Sc} = \frac{(\text{NO}_{\text{x\_elab1\_gpt}} - \text{NO}_{\text{elab1\_gpt}}) - (\text{NO}_{\text{x\_elab1\_span}} - \text{NO}_{\text{elab1\_span}})}{\text{NO}_{\text{elab1\_span}} - \text{NO}_{\text{elab1\_gpt}}} \quad (4)$$

For each step of the calibration event, only the last 10 min of data are considered to allow system stabilization. The values of NO, NO<sub>2</sub>, and NO<sub>x</sub> during the different steps of calibration, together with calculated zero offsets, span coefficients, and Sc values are stored in an internal table and plotted to allow verification of instrumental performance. For the first four months of 2018, Fig. 4 reports the different calibration coefficients and NO–NO<sub>2</sub>–NO<sub>x</sub> values measured at CMN during the calibration steps. The “calibrated” NO<sub>2</sub> is obtained by subtracting the “calibrated” NO from the “calibrated” NO<sub>x</sub>. Several threshold values are applied to the calculated calibration factors (Table 5): if the check of calibration factors (“Zero offset”, “Span coeff”, “Sc”) against defined threshold values fails, the last calibration factors successfully calculated are retained for data correction.

Concerning the data correction, at the current stage only the correction of the NO night-time bias is implemented. As concerning NO, for each date we calculate the night-time (00:00–04:00) 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:

**Table 6**

Description of data products.

Class	Description
GRAPH	Time series
TIMEVARIATION	Average diurnal/weekly/seasonal cycles
CALENDAR	Daily ECV average values laid out in a calendar format

In these conditions, NO is expected to decrease below the instrumental detection limit. If this does not happen, a “night-time” zero offset correction is calculated by smoothing the “night-time” NO reading over 2 days, and by subtracting the “night-time” offset from the Level-1 data series.

#### 2.5. Automatic processing of ECVs: data aggregation (Level-2 production)

Basing on the data screening and corrections applied in the previous steps (i.e., coding of Level-0 and Level-1 data), 1-min data are aggregated to hourly (60-min) 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. Besides average values, as a function of the different ECV, other statistical parameters are computed, as required by WDC data formats. In case less than 50% of data are used for the calculation of the hourly mean value, the hourly data is properly flagged (i.e., 0.390). Moreover, a set of three functions for creating, checking and aggregating numflags was developed. A brief description of each of them is provided in the Supplementary Material.

Fig. 5 shows an overview of NO<sub>2</sub> mole fraction measured at CMN on January 2018 (this is an extract of the operational product CMN\_NO2\_2018\_01\_MONTHLY\_GRAPH\_20180730.png, see the Supplementary Material). Here we reported the time series of Level-2 data (i.e., 1-h calibrated and corrected average values), the concurrent data flags and a comparison between “final” Level-2 data and raw data from the instrument. When looking at this comparison, one can see which data were discarded before averaging, leading to the appearance of missing values in the Level-2 data set (i.e., numflag equal to 0.999).

#### 2.6. Automatic processing of ECVs: 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 WDCGG, as reported in GAW Report N. 188. “Near-surface” reactive gases, aerosol properties and meteorological parameters are formatted following the NASA-AMES standard, as indicated by WMO/GAW WDCRG and WDCA. This format is based on the textual format ASCII NASA-AMES 1001 with additional metadata (as a function of the different ECVs). Among other metadata, these file headers include the provision of specific metadata about traceability to metrological standards, adoption of QA measures (e.g., on-site



Fig. 4. Time series of calibration factors (bkg\_NO, coeff\_NO, bkg\_NOx, coeff\_NOx, Sc) and NO-NO<sub>2</sub>-NO<sub>x</sub> values during the different steps of calibration events, from January to April 2018 at CMN (NO\_elab1\_span, NO\_elab1\_gpt, NO<sub>2</sub>\_elab1\_span, NO<sub>2</sub>\_elab1\_gpt, NO<sub>2</sub>\_elab2\_span, NO<sub>2</sub>\_elab2\_zero, NO<sub>x</sub>\_elab2\_span, NO<sub>x</sub>\_elab2\_gpt).

or off-site intercomparison, round-robin or on-site audit), uncertainty quantification and methods used for the calculation of the measurement uncertainty. 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>. 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.

### 2.7. Automatic processing of ECVs: data products

To support measurement stations in carrying out the QA/QC checks, a suite of products (i.e., data plots) is produced by the automatic data processing chain. These are updated on a daily basis, by using “P22” routines. To this aim, some specific functions of the “OpenAir” package (Carslaw and Ropkins, 2012) are also used. The data products are arranged as a function of different time windows: daily, monthly, seasonal and yearly. In total, 9 data reports are operationally produced for each ECV. Data reports are “.png” files identified by the following name code:

$$\text{SSS\_PPP\_yyyy\_mm\_PERIOD\_TYPE\_YYYYMMDD.png} \quad (6)$$

where SSS is the station code, PPP is the ECV code (see Table 3), yyyy\_mm identifies the time validity of the product (for data reports related to a full calendar year the code yyyy\_01 is conventionally

adopted), PERIOD is the period of time spanned by the data reports (i.e., “DAILY”, “MONTHLY”, “SEASONAL”, “SEMESTER”, “ANNUAL”), TYPE denotes the class of data product (i.e., “GRAPH”, “TIMEVARIATION”, “CALENDAR”, see Table 6), and YYYYMMDD is the file production date. For example, the code CMN\_NO2\_2018\_01\_MONTHLY\_GRAPH\_20180730.png identifies the monthly data product n. 1 (see below) created on July 30th, 2018, for NO<sub>2</sub> measured at CMN on January 2018 (CMN\_NO2\_2018\_01). In the Supplementary Material, we will provide a brief overview of the data products along with specific examples.

### 3. Self-evaluation of the methodology

In this section, we provide a self-assessment of the maturity of the presented system. To this aim, we use the tool presented by Thorne et al. (2017), but originally developed within the CORE-CLIMAX project (Su et al., 2018; Zeng et al., 2019), to assess the maturity of an environmental measurement network. In particular, a major strand of assessment related to the “software readiness” was introduced, which shall apply to measurement networks for which “routine-automated and substantive processing occurs from the raw measured data to the provided geophysical parameters of the measurement series” (Thorne et al., 2017; Su et al., 2018; Zeng et al., 2019). The sub-categories to be assessed are: “Coding standards”, “Software documentation”, “Portability and numerical reproducibility”, and “Security”. For each





Fig. 5. Upper: NO<sub>2</sub> (Level-2) at CMN in January 2018. The average value (red flat line) is also reported, together with the minimum (Min) and the maximum (Max) values for each month. Middle: numflag for NO<sub>2</sub> 1-min data (Level-1). Bottom: comparison between Level-2 data (red line) and raw 1-min readings from the instrument (blue area).

of these sub-categories, the assessment will assign a score from 1 to 6, reflecting the maturity of each specific aspect (i.e., 1: low maturity, 6: high maturity). The self-assessment was carried out by applying the guidelines provided in GAIA-CLIM (2015).

For what concerns the “Coding standards”, a score of 1 was given. Even if some ISO standards exist for software coding, they were not identified.

Concerning the “Software documentation”, a score of 3 was assessed. Information about “Configuration instructions”, “Installation instructions”, “Operating instructions”, “Copyright and licensing”, “Contact information” are provided within the script file headers and in the script repository. Moreover, information about software concept and design are provided by specific documentation (NextData, 2018).

The scripts generate numerically reproducible results on different computing platforms (e.g., Linux and Windows). Moreover, the scripts have been already used by third parties (other research groups from our Institute and other research Institutions), in operational environments. Thus, a score of 4 was given for “Portability and numerical reproducibility”.

For what concerns “Security” (i.e., the existence of software contents that either have the potential to destroy files), a score of 3 was assessed: specific tests have been carried out during the development and operational applications and threats have not been found, especially for what concerns the original raw data files.

Based on this assessment, the current maturity of the software readiness ranges from low to medium, as a function of different aspects.

#### 4. Specific additional services

The implementation of the automatic routines for the flagging and formatting of ECV data files provided the opportunity to implement a near-real time data delivery service. Two different data streams were activated for CMN: ACTRIS and GAW NRT for near-surface aerosol (i.e., eqBC and total particle number concentration), and CAMS-Copernicus for reactive gases (i.e., O<sub>3</sub>, NO, NO<sub>2</sub>, SO<sub>2</sub>). The two data streams are based on different routines and delivery

strategies that depend on the specific requirements from the two receiving programs. For ACTRIS/GAW NRT, the routine elaborates (as an instance a specific data header is created) the raw data files produced from the instruments (i.e., NEPH, CPC, and MAAP) and manages the automatic delivery to a SFTP server every 30 min. A data delivery service was activated towards CAMS-Copernicus for on-line model verification purpose (Wagner et al., 2015). In this case, a “bash” script was used to extract the requested information from the Level-2 files of reactive gases and to organize them in the formats requested from CAMS-Copernicus for submission. In particular, CAMS-Copernicus requests that daily files containing hourly averages and related standard deviation of the ECVs of interest are delivered with (at least) a daily frequency. The same script manages both the appending of the data header as well as the daily delivery of data files to a dedicated SFTP server.

In addition to the near-real time data delivery, a website (<http://nextdata.bo.isac.cnr.it>) for the on-line visualization of data products was specifically created. The website catalogs the ECVs and allows to search them by specifying the observatory and the class. All of the products presented in Section 2.7 are listed in the website; moreover, also the “health status report” (Fig. 3), for checking the correct execution of the elaboration routines, is shown. The website is updated daily, after the whole processing chain is completed.

#### 5. Summary and perspectives

To favor the integration of an atmospheric background observatory network in Italy, we implemented a first prototype of a centralized system to support measurement stations in data production and subsequent submission to reference data centers (WMO/GAW, or other research programs). Indeed, there is an increasing need for ECV data with well-assessed data quality (WMO, 2016), but the increasing data quality demand and the request of specific data format often faces against limited human resources at the atmospheric observatories. Leading research infrastructures (e.g., ICOS) or pan-European projects (e.g., ACTRIS-2) already implemented centralized facilities to process

raw data, to perform quality controls and to provide standardized data outputs, but these efforts are limited to a set of selected participating stations and to a list of “core” ECVs. It should be clearly stated that our system would not overcome station manager or instrument operators in their responsibility for the final validation and data submission to data centers. The NextData system does not generate “consolidated” data to be directly submitted; however, it represents a valuable tool to facilitate data originators towards a more efficient data production for those data streams which are not covered by already-existing services.

By using the “R” language (R Development Core Team, 2008) we developed, for each of the ECVs listed in Table 2, specific routines for data filtering, flagging, formatting, and creation of graphical data products. Our effort is expected to improve data quality and to accelerate the process of data submission to WMO/GAW WDCs (in this perspective, we will encourage the adoption of these procedures also by other measurement stations, not directly related to NextData). As assessed by the operational application of the proposed procedures in the framework of annual submission of data files to WMO/GAW and ACTRIS-2, if a user with relatively low skills in software programming is considered (e.g., only able to use spreadsheet applications like Microsoft Excel), the application of the proposed procedures would allow to save several days of manpower (depending on the specific ECV) throughout the different steps of datafile creation and submission processes.

Further improvements must be implemented in the current prototype version of the system. In this current implementation phase, a total of 68 scripts were coded to process 28 ECVs from 5 atmospheric observatories. If further resources will be available, an important step will be the optimization of the processing to generate for each script a single version able to manage each observatory. Further data correction procedures must be realized for some specific ECVs (e.g., water vapor and O<sub>3</sub> interferences to NO and NO<sub>2</sub>, see Gilge et al., 2014). Concerning the definition of the threshold values to be adopted for the automatic flagging, it should be clearly stated that the current set of values does not have the ambition to be “universal” (which is rather unfeasible considering the different span of atmospheric conditions that can affect measurements even in the same region), but they are specific for the ECVs and sites considered in this prototype application. However, they can be rather easily changed and adapted to specific conditions and measurement sites: each single PI will modify them and adapt them to his/her specific needs. Automatic functions for outlier selection, which rely on data variability, were implemented for the aerosol parameters. A sensitivity test was carried out to assess the impact of adopting different outlier selection methods on a subset of ECVs (NO, eqBC, and total particle number concentration): the results concerning their impact on data coverage fraction and on the correct representation of the diurnal variability are reported in the Supplementary Material.

Two additional services were also developed in the framework of this activity, as presented in Section 4. For a subset of ECVs (i.e., eqBC, total particle number concentration, O<sub>3</sub>, NO, NO<sub>2</sub>, and SO<sub>2</sub>), near-real time data delivery services were activated, to support the activities of the ACTRIS and GAW NRT and CAMS-Copernicus programs. A dedicated web site (<https://nextdata.bo.isac.cnr.it/>) was implemented for supporting measurement PIs in easily accessing the whole suite of generated data products (as a function of the different measurement sites and ECVs).

## 6. Routines and data availability

All the routines implemented for the different steps involved in the automatic processing of data are written by using the “R” language (R Development Core Team, 2008). All the routines are freely accessible by the NextData Geonetwork system (<http://nextdata.igg.cnr.it/>), under the category “Software”. The data for the different levels (i.e., Level-0, Level-1, and Level-2) shown in this paper are directly available

from the WDCs websites (see <http://ebas.nilu.no>). A subset of historical data about ECVs reported in this work are freely available on the platform MOVIDA-Multistat (<http://shiny.bo.isac.cnr.it:3838/plot-multistats-en/>), implemented in the framework of NextData. Metadata and data obtained thanks to the direct support of NextData can be found at the aforementioned website.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Luca Naitza:** Conceptualization, Data curation, Methodology, Software, Writing - original draft, Writing - review & editing. **Paolo Cristofanelli:** Project administration, Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing - original draft, Writing - review & editing. **Angela Marinoni:** Data curation. **Francescopiero Calzolari:** Data curation. **Fabrizio Roccatò:** Data curation. **Maurizio Busetto:** Data curation. **Damiano Sferlazzo:** Data curation. **Eleonora Aruffo:** Data curation. **Piero Di Carlo:** Data curation, Writing - review & editing. **Mariantonia Bencardino:** Data curation, Writing - review & editing. **Francesco D’Amore:** Data curation, Writing - review & editing. **Francesca Sprovieri:** Writing - review & editing. **Nicola Pirrone:** Funding acquisition, Writing - review & editing. **Federico Dallo:** Data curation, Writing - review & editing. **Jacopo Gabrieli:** Writing - review & editing. **Massimiliano Vardè:** Data curation, Writing - review & editing. **Giorgio Resci:** Software. **Carlo Barbante:** Funding acquisition, Writing - review & editing. **Paolo Bonasoni:** Funding acquisition, Writing - review & editing. **Davide Putero:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing - original draft, Writing - review & editing.

## Acknowledgments

This work has been supported by the National Project of Interest NextData by the Italian Ministry for Education, University and Research (MIUR) and by ACTRIS-2 H2020 project (G.A. 739530). Davide Putero and Luca Naitza grants have been supported by NextData. The “OpenAir” analysis package for R was obtained from <http://www.openair-project.org>.

## Appendix A. List of abbreviations and acronyms used in the paper

ACTRIS — European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases  
 BLC — Blue Light Converter  
 CAMS — Copernicus Atmosphere Monitoring Service  
 CGR — Capo Granitola observatory  
 CLD — Chemiluminescence Detection  
 CMN — Mt. Cimone observatory  
 CMP — Mt. Portella–Campo Imperatore observatory  
 CUR — Mt. Curcio observatory  
 ECV — Essential Climate Variable  
 EQC — Evaluation and Quality Control  
 GAW — Global Atmosphere Watch  
 GCOS — Global Climate Observing System  
 GMOS — Global Mercury Observation System  
 GPT — Gas Phase Titration  
 ICOS — Integrated Carbon Observation System  
 LMP — Lampedusa observatory  
 MRG — Col Margherita observatory  
 NRT — Near-Real Time  
 PI — Principal Investigator

PRS — Plateau Rosa observatory  
 QA/QC — Quality Assurance/Quality Check  
 SFTP — Secure File Transfer Protocol  
 SOP — Standard Operating Procedure  
 WDC — World Data Center  
 WDCA — World Data Center for Aerosol  
 WDCGG — World Data Center for Greenhouse Gases  
 WDCRG — World Data Center for Reactive Gases  
 WMO — World Meteorological Organization

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cageo.2020.104432>.

## References

- Carslaw, D.C., Ropkins, K., 2012. openair — An R package for air quality data analysis. *Environ. Model. Softw.* 27–28, 52–61.
- Cristofanelli, P., Brattich, E., Decesari, S., Landi, T., Maione, M., Putero, D., Tositti, L., Bonasoni, P., 2018. High-Mountain Atmospheric Research. The Italian Mt. Cimone WMO/GAW Global Station (2165 m a.s.l.). Springer International Publishing.
- D'Amore, F., Bencardino, M., Cinnirella, S., Sprovieri, F., Pirrone, N., 2015. Data quality through a web-based QA/QC system: implementation for atmospheric mercury data from the global mercury observation system. *Environ. Sci. Process. Impacts* 17, 1482–1491.
- El Yazidi, A., Ramonet, M., Ciais, P., Broquet, G., Pison, I., Abbaris, A., Brunner, D., Conil, S., Delmotte, M., Gheusi, F., Guerin, F., Hazan, L., Kachroudi, N., Kouvarakis, G., Mihalopoulos, N., Rivier, L., Serça, D., 2018. Identification of spikes associated with local sources in continuous time series of atmospheric CO, CO<sub>2</sub> and CH<sub>4</sub>. *Atmos. Meas. Tech.* 11 (3), 1599–1614.
- GAIA-CLIM, 2015. Report on system of systems approach adopted and rationale. URL [http://www.gaia-clim.eu/system/files/workpkg\\_files/640276\\_Report%20on%20system%20of%20systems%20approach%20adopted%20and%20rationale.pdf](http://www.gaia-clim.eu/system/files/workpkg_files/640276_Report%20on%20system%20of%20systems%20approach%20adopted%20and%20rationale.pdf).
- GAW, 2011. WMO/GAW expert workshop on global long-term measurements of nitrogen oxides and recommendations for GAW nitrogen oxides network. URL [https://www.wmo.int/pages/prog/arep/gaw/documents/Final\\_GAW\\_195\\_TD\\_No\\_1570\\_web.pdf](https://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_195_TD_No_1570_web.pdf).
- Gilge, S., Pass-Duelmer, C., Roher, F., Steinbacher, M., Fjaeraa, A.M., Lageler, F., Walden, J., 2014. Deliverable D4.10: Standardized operating procedures (SOPs) for NO<sub>x</sub> measurements. URL [http://fp7.actris.eu/Portals/97/deliverables/PU/WP4\\_D4.10\\_M42.pdf](http://fp7.actris.eu/Portals/97/deliverables/PU/WP4_D4.10_M42.pdf).
- Hazan, L., Tarniewicz, J., Ramonet, M., Laurent, O., Abbaris, A., 2016. Automatic processing of atmospheric CO<sub>2</sub> and CH<sub>4</sub> mole fractions at the ICOS atmosphere thematic centre. *Atmos. Meas. Tech.* 9 (9), 4719–4736.
- Huang, S., Wang, Y., Xie, Y., Zhao, P., Lüers, J., 2016. OutlierFlag: A tool for scientific data quality control by outlier data flagging. *J. Open Res. Softw.* 4.
- NextData, 2018. Launch of experimental services for data quality check and near-real time data delivery/early warning at the Mt. Cimone GAW global station. URL [http://www.nextdataproject.it/sites/default/files/docs/WP1.1\\_deliverable\\_1.1B.pdf](http://www.nextdataproject.it/sites/default/files/docs/WP1.1_deliverable_1.1B.pdf).
- Ni, K., Ramanathan, N., Chehade, M.N.H., Balzano, L., Nair, S., Zahedi, S., Kohler, E., Pottier, G., Hansen, M., Srivastava, M., 2009. Sensor network data fault types. *ACM Trans. Sensor Netw.* 5 (3), 1–29, URL <http://doi.acm.org/10.1145/1525856.1525863>.
- R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, URL <http://www.R-project.org>.
- Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevo, A.S.H., Hueglin, C., 2007. Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques. *J. Geophys. Res.* 112 (D11).
- Su, Z., Timmermans, W., Zeng, Y., Schulz, J., John, V.O., Roebeling, R.A., Poli, P., Tan, D., Kaspar, F., Kaiser-Weiss, A.K., Swinnen, E., Toté, C., Gregow, H., Manninen, T., Riihelä, A., Calvet, J.-C., Ma, Y., Wen, J., 2018. An overview of European efforts in generating climate data records. *Bull. Am. Meteorol. Soc.* 99 (2), 349–359.
- Thorne, P.W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., Tramutola, E., Arola, A., Buschmann, M., Mikalsen, A.C., Davy, R., Voces, C., Kreher, K., De Maziere, M., Pappalardo, G., 2017. Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment. *Geosci. Instrum. Methods Data Syst.* 6 (2), 453–472, URL <https://www.geosci-instrum-method-data-syst.net/6/453/2017/>.
- Wagner, A., Blechschmidt, A.-M., Bouarar, I., Brunke, E.-G., Clerbaux, C., Cupeiro, M., Cristofanelli, P., Eskes, H., Flemming, J., Flentje, H., George, M., Gilge, S., Hilboll, A., Inness, A., Kapsomenakis, J., Richter, A., Ries, L., Spangl, W., Stein, O., Weller, R., Zerefos, C., 2015. Evaluation of the MACC operational forecast system – potential and challenges of global near-real-time modelling with respect to reactive gases in the troposphere. *Atmos. Chem. Phys.* 15 (24), 14005–14030.
- WMO, 2016. Global Climate Observing System Implementation Plan: 2016. World Meteorological Organization, Geneva, Switzerland, URL [https://library.wmo.int/doc\\_num.php?explnum\\_id=3417](https://library.wmo.int/doc_num.php?explnum_id=3417).
- WMO, 2017. Global Atmosphere Watch Implementation Plan: 2016 - 2023. GAW Report No. 228, World Meteorological Organization, Geneva, Switzerland, URL [https://library.wmo.int/doc\\_num.php?explnum\\_id=3395](https://library.wmo.int/doc_num.php?explnum_id=3395).
- Zeng, Y., Su, Z., Barnpadimos, I., Perrels, A., Poli, P., Boersma, K.F., Frey, A., Ma, X., de Bruin, K., Goosen, H., John, V.O., Roebeling, R., Schulz, J., Timmermans, W., 2019. Towards a traceable climate service: Assessment of quality and usability of essential climate variables. *Remote Sens.* 11 (10), 1186, URL <http://dx.doi.org/10.3390/rs11101186>.