

Project Title	Prototype of HPC/Data Infrastructure for On-demand Services
Project Acronym	PHIDIAS
Grant Agreement No.	INEA/CEF/ICT/A2018/1810854
Start Date of Project	01.09.2019
Duration of Project	36 Months
Project Website	<a href="http://www.phidias-hpc.eu">www.phidias-hpc.eu</a>

## D4.1.2 – WP4 Work flow specification and product specification document

Work Package	<b>WP 4: Intelligent screening of a large amount of satellite data for detection and identification of atmospheric composition events</b>
Lead Author (Org)	<b>Pascal Prunet (SPASCIA)</b>
Contributing Author(s) (Org)	<b>Hervé Thevenon (SPASCIA), Olivier Lezeaux (SPASCIA), Claude Camy-Peyret (SPASCIA, IPSL), Dominique Jolivet (HYGEOS), Nicolas Pascal (ICARE)</b>
Due Date	<b>01.09.2020</b>
Date	<b>08.01.2021</b>
Version	<b>0.3</b>

### Dissemination Level

<input type="checkbox"/>	PU: Public
<input checked="" type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)



## Versioning and contribution history

Version	Date	Author	Notes
0.0	07.09.2020	Pascal Prunet (SPASCIA), Hervé Thevenon (SPASCIA), Olivier Lezeaux (SPASCIA), Claude Camy-Peyret (SPASCIA, IPSL), Dominique Jolivet (HYGEOS)	Draft version v0.0
0.1	10.10.2020	Nicolas Pascal (ICARE), Pascal Prunet (SPASCIA)	Version reviewed by ICARE
0.2	30.11.2020	Pascal Prunet (SPASCIA), Florian Piffet (CINES), Thierry Bidot (NEOVIA)	Version reviewed based on feedbacks from CINES and NEOVIA
0.3	08.01.21	Florian Piffet (CINES)	Final edition

### Disclaimer

This document contains information which is proprietary to the PHIDIAS Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the PHIDIAS Consortium.

## Table of Contents

Executive Summary .....	5
1 Objectives of WP4 Use Case .....	6
1.1 PCA-based screening of L1 data.....	6
1.2 Detection of gas plumes and sources from L2 products.....	7
2 Work Flow and product specification : first iteration .....	8
2.1 Input data .....	8
2.2 Processing.....	12
2.3 Output product.....	16
3 New AI methods for automatic detection of sources plumes from L2 products : Processing development status .....	17
3.1 Cluster.....	17
3.2 Data sets .....	21
3.3 Pseudo code .....	24
3.4 Next steps .....	25
4 Detection of gas plumes and sources : Preliminary tests on a dedicated use case : monitoring atmospheric pollution level at city scale during covid 19 event .....	26
4.1 Introduction .....	26
4.2 Data and methods .....	29
4.3 Results for Paris .....	31
4.4 Results for Milan.....	44
4.5 Results for Madrid .....	52
4.6 Results for Athens.....	57
4.7 Conclusion .....	63
5 PCA-based screening of L1 data : Processing development status (to be completed : HYGEOS, SPA/IPSL) .....	66
5.1 Implementation of the processing.....	66
5.2 Interfacing and pre-processing of S5P L1b spectra.....	67

## TERMINOLOGY

---

Terminology/Acronym	Description
AI	Artificial Intelligence
IASI	Intreferomètre Avancé de Sondage dans l'Infrarouge
COV	Covariance Matrix
L1	Level 1
L2	Level 2
PCA	Principal Component Analysis
PCC	Principal Component Compression
S5P	Sentinel 5 Precursor
WP	WorkPackage

## Executive Summary

---

*This document provides first iteration on the Workflow and product as defined and developed in WP4 of the PHIDIAS project. It provides the development status for the two WP4 proposed processings, as well as an example of the application of the S5P level 2 processing to a dedicated use case.*

# 1 Objectives of WP4 Use Case

---

WP4 use case addresses a well identified, key requirement for the exploitation of Earth observations: to provide the capacity of intelligent screening of large amounts of satellite data for targeting scenes of interest (e.g., extreme atmospheric events), in view of their dedicated processing or exploitation. The potential main users are: the scientific community (for the efficient specific/thematic analyses of the data); the operational services at national/European levels (for the monitoring and validation of their forecast systems); the service providers, for implementing dedicated applications for end users (e.g., alert or decision support services).

Objective of the WP4 Use Case: using HPC (high-performance computing) and HPDA (high-performance data analytics) capacities for developing intelligent screening approaches for the exploitation of large amounts of satellite atmospheric data in an operational context. It is proposed to implement these approaches as a prototype service on the already available S5P European atmospheric sounding mission.

The objective is to implement, demonstrate and made available to the users a service offering smart data filtering, in order to:




- detect the relevant data within the huge volume of measurements,
- identify and assess the pertinent information
- Qualify the targeted data for dedicated exploitation by users.

and provide the users with an access to the filtered data and the corresponding metadata.




Two different processing, applied to two different S5P datasets and addressing different, will be implemented:

1. PCA-based screening of L1 data (SWIR) for detection of extreme events. Based on experience and methods developed for IASI, implementation and consolidation of algorithms and tools for generic processing of atmospheric spectra recorded by S5P.
2. New AI methods for objective/automatic detection of pollutant plumes and sources from L2 products (CO, others?: CH<sub>4</sub> NO<sub>2</sub>, SO<sub>2</sub>) : Based on innovative algorithms and on SPASCIa experience with physical methods analyzing specific signal enhancements.

## ***1.1 PCA-based screening of L1 data***

-  Input : Applied on S5P Level 1 (L1, correspond to calibrated spectra) data.
-  Objective : On the flow, real time processing of the data, for detection of extreme events : Output : selection of characterised (through metadata) and stored dataset of interest, which will be available to the users. No storage of input data is needed.
-  Prototyping : during the development phase, a prototype will be tested on 1 year of S5P L1 dataset. Need for storage of this 1 year dataset at AERIS/ICARE during the development phase (i.e., duration of the PHIDIAS project).

## ***1.2 Detection of gas plumes and sources from L2 products***


-  Input : Applied on S5P Level 2 (L2, correspond to retrieved gas concentrations) data.
-  Objective : On demand re-processing of the data, for detection of plumes and sources over a region/period specified by the user : Output : characterisation of this region/period dataset (through metadata) flagging the data associated to the plumes, and localising the corresponding sources, which will be available to the users. Storage of the input data and output metadata is envisaged.
-  Prototyping : during the development phase, a prototype will be tested on 1 year of S5P L2 dataset (chosen period : 1st May 2018 to 30th April 2019). AERIS/ICARE will also store the continuous flow of all the L2 data from 1st May 2018.

## 2 Work Flow and product specification : first iteration

### 2.1 Input data

#### 2.1.1 Description

The TROPospheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor satellite, which was successfully launched in October 2017, is a spaceborne nadir viewing imaging spectrometer measuring solar radiation reflected by the Earth in a push-broom configuration. It has a wide swath on the terrestrial surface and covers wavelength bands between the ultraviolet (UV) and the shortwave infrared (SWIR) (See Table 1) combining a high spatial resolution with daily global coverage. These characteristics enable the determination of gases with unprecedented level of detail on a global scale introducing new areas of application. Abundances of the atmospheric column-averaged dry air mole fractions of several gases (Level 2 products, see Table 2) are retrieved from TROPOMI's radiance measurements (Level 1 products, Table 1).

<div>TROPOMI</div>								
	UV		UVIS		NIR		SWIR	
Band	1	2	3	4	5	6	7	8
Spectral coverage [nm]	270 – 320		320 – 495		675 - 775		2305 – 2385	
Full spectral coverage [nm]	267 - 332		303 - 499		660 - 784		2299 - 2390	
Spectral resolution [nm]	0.49		0.54		0.38		0.25	
Spectral sampling ratio	6.7		2.5		2.8		2.5	
Spatial sampling [km²]	7 x 28	7 x 3.5				7 x 3.5	7 x 7	

**Table 1 : TROPOMI Level 1B products**




Product	Spectrometer	Application
Ozone	UV, UVIS	Ozone layer monitoring, UV-index forecast, Climate monitoring
NO <sub>2</sub>	UVIS	Air quality forecast and monitoring
CO	SWIR	Air quality forecast and monitoring
CH <sub>2</sub> O	UVIS	Air quality forecast and monitoring
CH <sub>4</sub>	SWIR	Climate monitoring
SO <sub>2</sub>	UVIS	Air quality forecast and monitoring, Climate monitoring, Volcanic plume detection
Aerosol	UVIS, NIR	Air quality forecast and monitoring, Climate monitoring, Volcanic plume detection
Clouds	UVIS, NIR	Climate monitoring
UV-Index	UVIS	UV index forecast

**Table 2: TROPOMI Level 2 geophysical products**

As for all Sentinel missions, the Sentinel-5P products are freely available to users from Copernicus. There is a specific data hub for S5P, called the SP5 pre-ops, which is different than the SENTINEL Open Hub: <https://s5phub.copernicus.eu>. Data are available operationally from 30<sup>th</sup> April 2018. Detailed information are available on <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-5p/products-algorithms>, and is reproduced below.

### Level 1B – User technical documentation

The information needed to properly use the Level 1B data can be found in Table 3:





-  IODS (Input Output Data Specification): description of the products that are the result from the Level 0 to Level 1b processing
-  ATBD (Algorithm Theoretical Basis Document): high level description of the algorithms used in the Level-0 to 1b data processing
-  PRF (Product Readme File): description of changes between different product versions and overall quality information (available a few months after launch)

**Table 3: TROPOMI Level 1B radiance/irradiance products and user documentation**

File type	Spectrometer	Spectral range [nm]	Comment	User Documentation
L1B_RA_BD1	UV	270 - 300	Radiance product band 1	<a href="#">IODS</a>
L1B_RA_BD2		300 - 320	Radiance product band 2	
L1B_RA_BD3	UVIS	320 - 405	Radiance product band 3	
L1B_RA_BD4		405 - 500	Radiance product band 4	
L1B_RA_BD5	NIR	675 - 725	Radiance product band 5	<a href="#">ATBD</a>
L1B_RA_BD6		725 - 775	Radiance product band 6	
L1B_RA_BD7	SWIR	2305-2345	Radiance product band 7	<a href="#">PRF</a>
L1B_RA_BD8		2345-2385	Radiance product band 8	
IR_UVN	UVN	270-775	Irradiance product UVN module	<a href="#">Other documents</a>
IR_SIR	SWIR	2305-2385	Irradiance product SWIR module	

## Level 2 – User technical documentation

The information needed to properly use the Level 2 data can be found in Table 4:

-  PUM (Product User Information): information on the technical characteristics of the S5P/TROPOMI Level 2 products
-  ATBD (Algorithm Theoretical Basis Document): detailed information on the retrieval algorithms
-  IODD (Input Output Data definition): description of the input and output data of the S5P/TROPOMI Level 2 processing
-  PRF (Product Readme File): description of changes between different product versions and overall quality information (available few months after launch)

**Table 4: TROPOMI Level 2 geophysical products and user documentation**

Product type	Parameter	User Documents
L2__O3__	Ozone (O <sub>3</sub> ) total column	<a href="#">PRF-O3-NRTI</a> , <a href="#">PRF-O3-OFFL</a> , <a href="#">PUM-O3</a> , <a href="#">ATBD-O3</a> , <a href="#">IODD-UPAS</a>
L2__O3_TCL	Ozone (O <sub>3</sub> ) tropospheric column	<a href="#">PRF-O3-T</a> , <a href="#">PUM-O3_T</a> , <a href="#">ATBD-O3_T</a> , <a href="#">IODD-UPAS</a>
L2__O3__PR	Ozone (O <sub>3</sub> ) profile	<a href="#">PUM-PR</a> , <a href="#">ATBD-O3 PR</a> , <a href="#">IODD-NL</a>
L2__O3_TPR	Ozone (O <sub>3</sub> ) tropospheric profile	<a href="#">PUM-PR</a> , <a href="#">ATBD-O3 PR</a> , <a href="#">IODD-NL</a>
L2__NO2__	Nitrogen Dioxide (NO <sub>2</sub> ), total and tropospheric columns	<a href="#">PRF-NO2</a> , <a href="#">PUM-NO2</a> , <a href="#">ATBD-NO2</a> , <a href="#">IODD-NL</a>
L2__SO2__	Sulfur Dioxide (SO <sub>2</sub> ) total column	<a href="#">PRF-SO2</a> , <a href="#">PUM-SO2</a> , <a href="#">ATBD-SO2</a> , <a href="#">IODD-UPAS</a>
L2__CO__	Carbon Monoxide (CO) total column	<a href="#">PRF-CO</a> , <a href="#">PUM-CO</a> , <a href="#">ATBD-CO</a> , <a href="#">IODD-NL</a>
L2__CH4__	Methane (CH <sub>4</sub> ) total column	<a href="#">PRF-CH4</a> , <a href="#">PUM-CH4</a> , <a href="#">ATBD-CH4</a> , <a href="#">IODD-NL</a>
L2__HCHO__	Formaldehyde (HCHO) total column	<a href="#">PRF-HCHO</a> , <a href="#">PUM-HCHO</a> , <a href="#">ATBD-HCHO</a> , <a href="#">IODD-UPAS</a>

L2__CLOUD_	Cloud fraction, albedo, top pressure	<a href="#">PRF-CL</a> , <a href="#">PUM-CL</a> , <a href="#">ATBD-CL</a> , <a href="#">IODD-UPAS</a>
L2__AER_AI	UV Aerosol Index	<a href="#">PRF-AI</a> , <a href="#">PUM-AI</a> , <a href="#">ATBD-AI</a> , <a href="#">IODD-NL</a>
L2__AER_LH	Aerosol Layer Height (mid-level pressure)	<a href="#">PRF-LH</a> , <a href="#">PUM-LH</a> , <a href="#">ATBD-LH</a> , <a href="#">IODD-NL</a>
UV product	Surface Irradiance/erythemal dose	-
L2__NP_BDx, x=3, 6, 7	Suomi-NPP VIIRS Clouds	<a href="#">PRF-NPP</a> , <a href="#">PUM-NPP</a> , <a href="#">ATBD-NPP</a>

## Data Volume

Product type	Volume/day (GB)
L1B	475 GB
L2	35 GB
Total	510 GB

### 2.1.2 Access for PHIDIAS processing chain through ICARE facilities

## Data Access

The input data will be available through the usual AERIS/ICARE distribution channels:

- HTTP : <https://www.icare.univ-lille.fr/data-access/data-archive-access/?dir=S5P>
- FTP : <ftp://ftp.icare.univ-lille1.fr/SPACEBORNE/S5P>

The data are publically available, but the users need to be registered to access to these services. The request form is available here: <https://www.icare.univ-lille.fr/register>

To end with, the users accredited by project PIs can have a direct access to the AERIS/ICARE computational resources through SSH. The service is described here: <https://www.icare.univ-lille.fr/services-2/computing-resources>.

The AERIS/ICARE S5P data archive will also be shared with CINES through iRODS, for systematic processing.

### 2.1.3 Use in the PHIDIAS processing chain

### 2.1.4 PCA-based screening of L1 data

**Operational use case** (at the end of PHIDIAS project): On the flow, real time processing of the data, for detection of extreme events. This will produce a selection of characterised (through metadata) and stored dataset of interest (preliminary estimate: less than 1% of the data), available to the users.

**Development phase** (during the PHIDIAS project): a prototype will be tested on 1 year of S5P L1 dataset (chosen period : 1<sup>st</sup> May 2018 to 30<sup>th</sup> April 2019). There is a need for storage of this 1 year dataset at ICARE during the development phase (i.e., duration of the PHIDIAS project). Specification, volume and definition of output products will be done during this development phase.

### 2.1.5 Detection of plumes and sources from L2 products

**Operational use case** (during phase 2 of PHIDIAS project (pilot use cases)) : On demand re-processing of the data, for detection of plumes and sources over a region/period specified by the user : Output : characterisation of this region/period dataset (through metadata) flagging the data associated to the plumes, and localising the corresponding sources, which will be available to the users. Storage of the input data and output metadata is required.

## 2.2 Processing

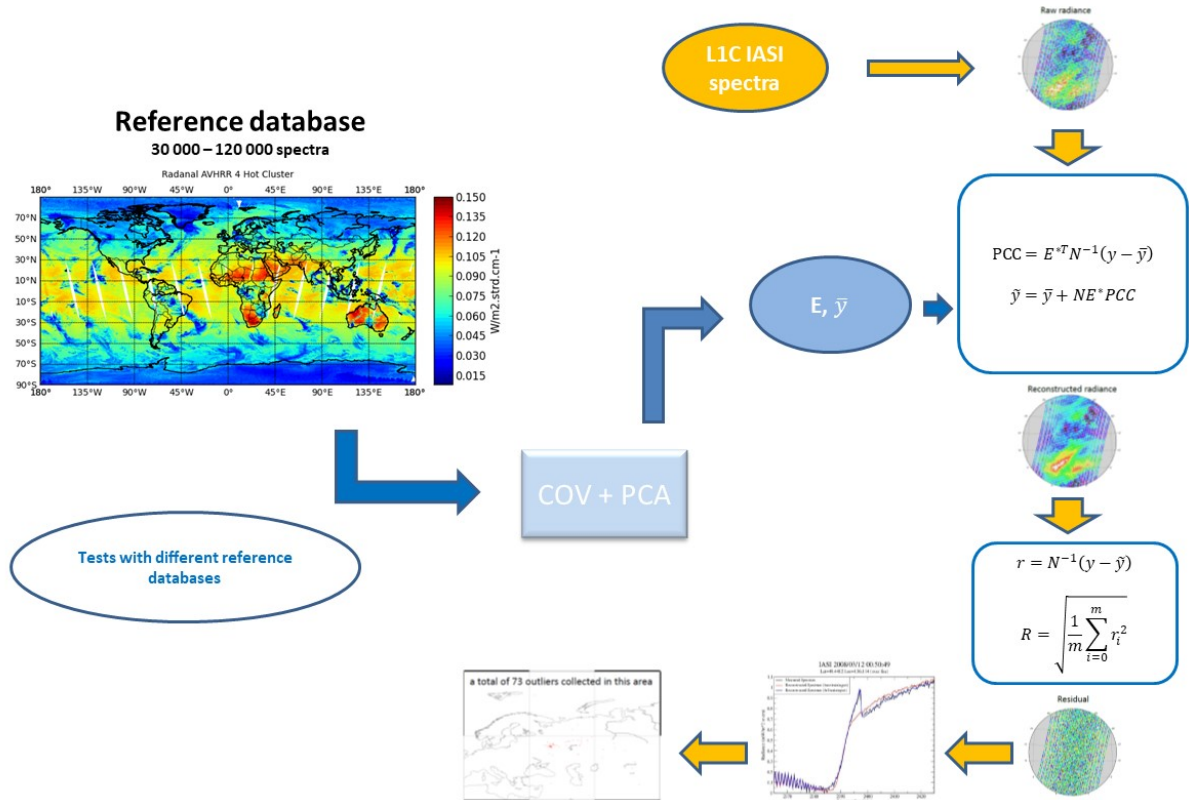
### 2.2.1 PCA-based screening of L1 data

### 2.2.2 High level description

The method is based on the analysis of the variability of a set of  $n$  S5P spectra  $y$  (S5P L1 product) of dimension  $m$  (the number of spectral channels) : this is the learning dataset, or reference dataset. This dataset shall include observed data with different atmospheric and surface conditions, different viewing angles ... It shall contains a large number of data (typically 100 000) in order to avoid non representative correlations. The behaviour of the process and its property of noise filtering and signal screening are strongly dependent on the definition of the reference dataset.

The principle is to compute from this reference dataset the covariance matrix of the measurements variability, then to apply principal component analysis (PCA) in order to extract the main eigen structures (eigenvectors and eigenvalues). These eigenstructures will serve as a reference basis (or PC basis) for the analysis of each measurement : projection of the measurement on the reference basis; truncated reconstruction, computation of the reconstruction residuals and of the reconstruction score.

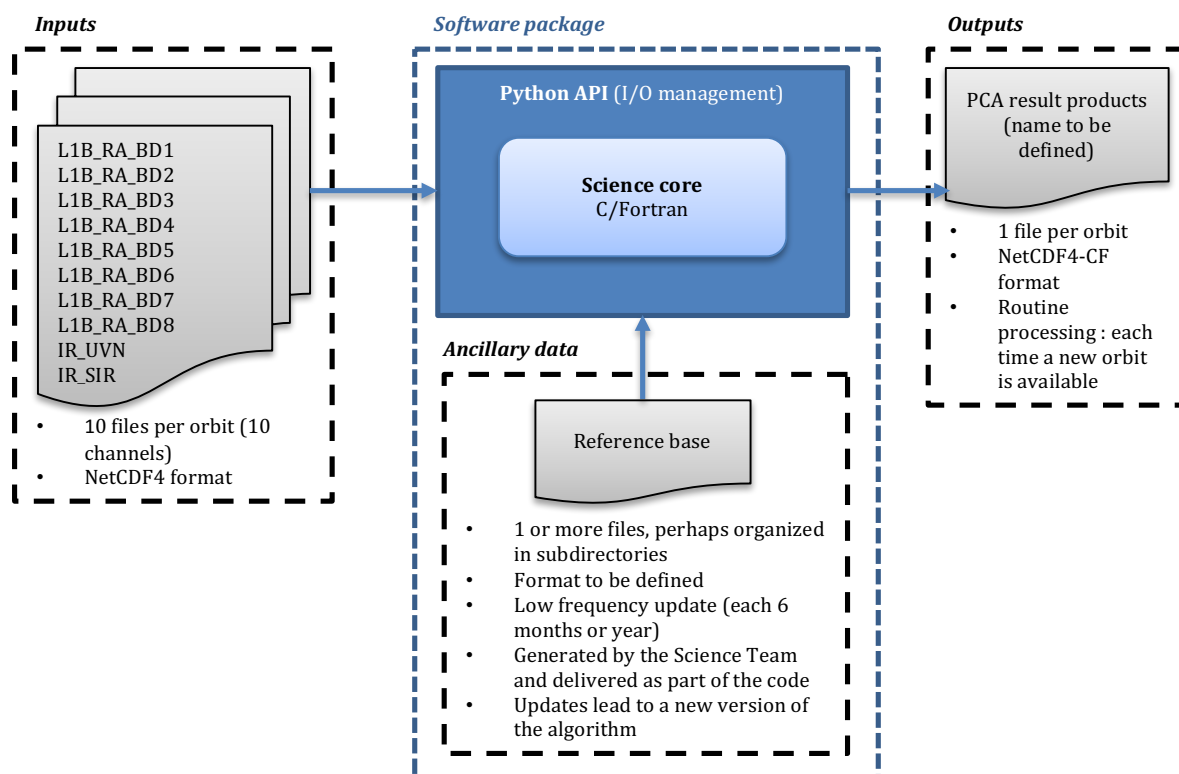
The proposed processing is illustrated on the Figure below for the analysis of IASI data. A similar processing will be implemented for S5P.



PCA based screening, illustration for IASI data : input are in orange, ancilliary data in blue. The notations used are the following : COV (covariance matrix computed from the reference database); PCA (Principal Component Analysis of the covariance matrix); E (basis of eigenvectors),  $\bar{y}$  (average spectrum from the reference database);  $y$  (current spectrum); PCC (Principal Component Compression); N (instrument noise covariance matrix, used for normalisation);  $\hat{y}$  (reconstructed spectrum);  $r$  (spectral residual vector);  $r_i$  (spectral residual sample  $i$ );  $m$  (number of spectral samples in the spectrum); R (reconstruction score)

This processing allows an efficient detection of extreme event which are not- or under-represented in the reference dataset, associated to a high reconstruction score. The identification and analysis of the specific detected event should require the exploitation of the spectral structure of the reconstruction residuals.

### 2.2.3 General Technical Scheme



### 2.2.4 Detection of plumes and sources from L2 products

### 2.2.5 High level description

The approach will use as input a regional or global map of level 2 product associated with a pollutant (e.g. : CO, CH<sub>4</sub>), consistently connected in space and close in time (typically for a given day). The processing will explore the map in order to :

- identify the sources points associated with point source emissions of the mapped pollutant, and flag the corresponding pixels.
- Identify all the pixels associated with the plume emitted by a given source point, and flag the corresponding pixels.
- For a pixels impacted by several sources, provide an estimation of the contribution of each sources to this pixel.

The processing will need, in addition to the input L2 product, the following ancillary information:

- 3D wind fields associated to each pixel map, interpolated from ECMWF forecast (NRT process) or analysis (non NRT process). The availability of ECMWF fields as input is a critical point that should be specified soon in the development phase, because the archive of data must be constructed, and the authorizations to use these data need to be validated by ECMWF.

- A priori estimate of the position of the sources, if available.

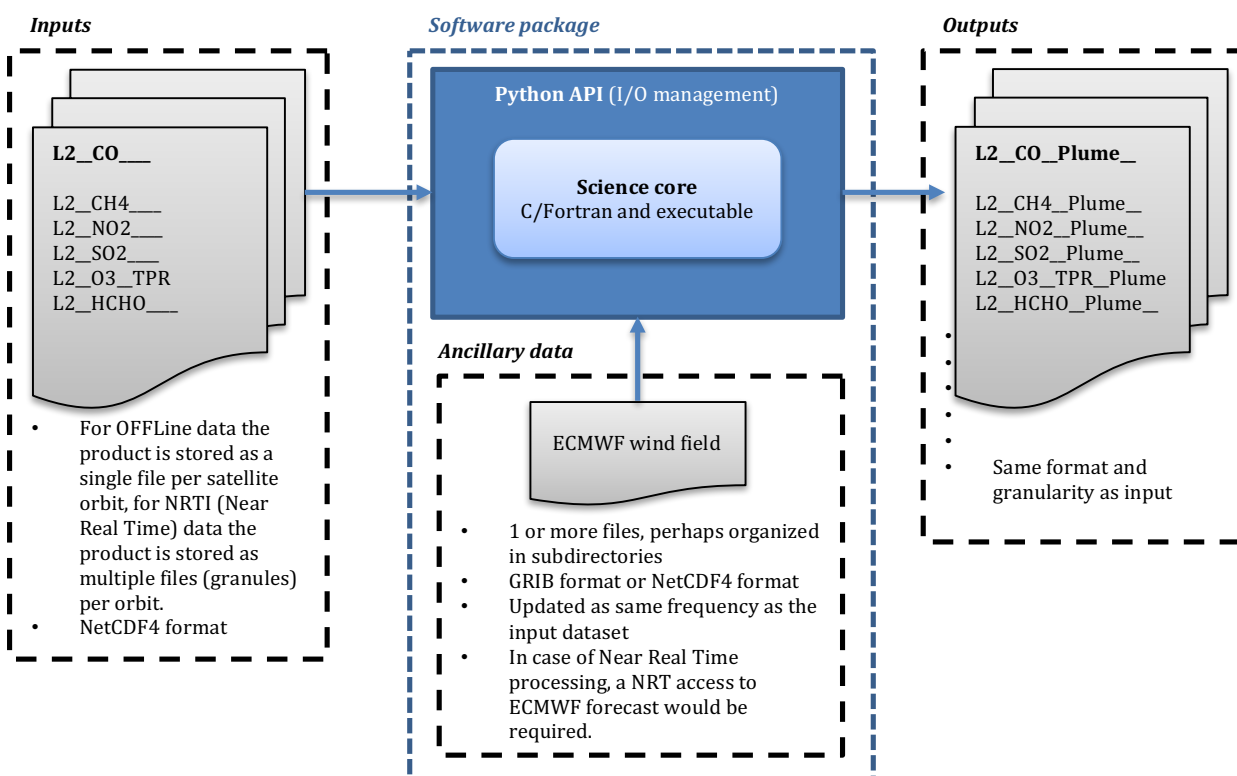
The output of the processing will be the input regional/global L2 product map with additional information, consisting in additional flags, typically :

- a flag associated to the qualification of the pixel (source/plume/none),
- a quality flag,
- additional information (number of associated sources, contribution of each source, distance to the source, ...).

Note that the above description is for the operational phase. During the development phase of this processing, will involve specific computation and storage resources for the computation and tests of different reference databases..

## 2.2.6 General Technical Scheme

The baseline processing will apply on off Line data. If Near Real Time (NRT) processing is required (to be discussed with the users) specific access to some data (in particular ECMWF ancillary dataset) shall be verified.



The technical scheme above is nominally applying on one input type (**L2\_CO\_**, in bold), providing one corresponding output (**L2\_CO\_Plume\_**, in bold). In a next step, it could be applied to other type of input, and/or to multiple inputs, providing the corresponding number of outputs.

## **2.3 Output product**

A complete product specification is not available at this development stage of the project. In particular, the technical specification describing format, name of the datasets, metadata, average volumes, etc ... is not provided. Final file formats and product characterisations are still under development and will be detailed in a next version of this document. It will be based on NetCDF and will follow the Climate Forecast convention” for instance.

### **2.3.1 PCA-based screening of L1 data**

Use Case I : systematic processing, for use in real time and/or storage for future use.  
Only selected input data and output metadata to be provided to the users

### **2.3.2 Detection of plumes and sources from L2 products**

Use Case I : systematic processing on selected regions of interest ?

Use Case II : On demand processing : product type, region and period of interest to be specified by the users.

Input products and output metaproducts to be provided to the users, over the specified region/period.

Implementations, still under discussion in WP2, could be:

- Setting up an access to the softwares, data and processing facility through a Jupyter notebook for on-demand runs.
- Providing an interactive web interface to trigger the processing,
- Providing a RestAPI,
- Providing a WPS system.

### **2.3.3 Expert specific usages**

Use Case III : interactive processing of both L1 PCA-based screening and L2 plume/source detection by experts. To be specified.

Implementation could be:

- Setting up an access to the softwares, data and processing facility through a Jupyter notebook for on-demand runs.

## 3 New AI methods for automatic detection of sources plumes from L2 products: Processing developement status

---

In this approach, the automatic detection of NO<sub>2</sub> plumes is based on an analogy with the animal kingdom, specifically with the moth pheromone system that allows males to find swarms of females from 50 meters away downwind, with two one millimeter long antennae. This is indeed the moth's detection and homing system that we'd like to mimic. We may not have the ability to reverse engineer the logic at work in the moth's nervous system, but a recent thesis in neuroscience (2018) provides us with the ability to try and evolve a connectome that achieves similar results. This connectome is a dynamic stateful system that belongs to the class of mathematical objects known as complex dynamic systems. The inputs of a connectome are sequential sparse and ambiguous events (coming from the real world or from a time series) and outputs are actions that attempt to modify the environment (including moving within the environment) for survival. Connectomes are not expected to perform specific tasks but to meet specific goals. In essence, this is what is being referred to by "new AI methods".

### 3.1 Cluster

#### 3.1.1 Computational requirements

At the moment, the only known process that builds connectomes suitable to achieve a goal is to evolve them through a stochastic optimisation process with objective function. In other words, this is a massive trial and error exercise. At the time of writing, we expect the search space of the problem at hand to be in the order of 10<sup>70</sup> to 10<sup>90</sup>. The challenge is to try as many solutions as possible in a reasonable amount of time, while knowing that any amount of human time will barely make a dent in the vastness of the search space.

In order to minimise the processing time, all the software has been optimised: the code that runs the connectomes is written in C and heavily uses pointers (we refer to it as the client), the stochastic optimisation process was made continuous in order to prevent the huge amount of NOOPs incurred by the standard epoch or generation based algorithms, and the data is prepared in such a way that clients can make use of it with the least overhead.

A database was also introduced as the blackboard of the massively parallel process. Each client pulls new candidates from the database and pushes back the results after each client is evaluated. Results include all the information required to allow for assessing and improving the objective function, including what was sensed and what actions were taken, when. Back-end processes are in charge of generating new candidates before the participating hosts get starved; they make use of the objective function to rank the evaluated candidates given the returned results.

It is a tightly integrated and optimised system that allows multiple hosts to perform a massively parallel task.

#### 3.1.2 Computational options

The optimisation choices that were made are difficult to integrate on a scientific cluster and commercial cloud computing facilities. Indeed, most scientific clusters rely on heterogeneous

nodes where code written for interpreted languages (e.g. Python) run seamlessly, but where compiled code is cumbersome to exploit. Also, scientific clusters don't let the users handle the scheduling of their work units. These characteristics defeat the optimisations described previously.

In an era where GPUs are being used for seemingly everything, the question whether CPUs are the best fit for purpose may arise. The truth is that GPUs are the least suited hardware that may run the connectome algorithm, because they are designed to execute lots of identical work items that carry out the same instructions without branching on data that has – from the perspective of the work units – no structure. For each connectome decides what data will be used next, we have here the elements of branching and random access to the local memory that are unsuited for GPUs. In addition, for each client interprets a data structure (the connectome) as a code to execute, connectomes or clients cannot be considered as work items of a GPU. We have here all the fundamental weaknesses of GPUs.

High-end FPGAs would probably be the most suited hardware but this would require to develop both a tool to convert the client and the connectomes into VHDL and a system to drive a small cluster of FPGAs. This is a project in itself, that would require very specific human and hardware resources. That is not an option at the moment.

Consequently, we decided to build a cluster dedicated to the specific requirements outlined above. The result is a prototype nicknamed Cleo, a 432-core cluster built around 108 single board computers within 4 months and for a cost of 6,000 Euros, manpower included.



*Figure 1. The cluster. Two stacks of Pis on the sides, network switches in the center, and a MacMini as the backend in the centre.*

### 3.1.3 Functional presentation of the cluster

In its most generic form (figure 2), the system required to evolve connectomes consists of three software stacks:

- the platform management stack (SBC update, client update, monitoring),

- the stochastic optimisation stack (generation of candidates, monitoring of the candidates pool, creation of mutants according to the objective function),
- the environmental interaction stack (bespoke server, blackboard) that allows for multiple connectomes to interact within the same environment.

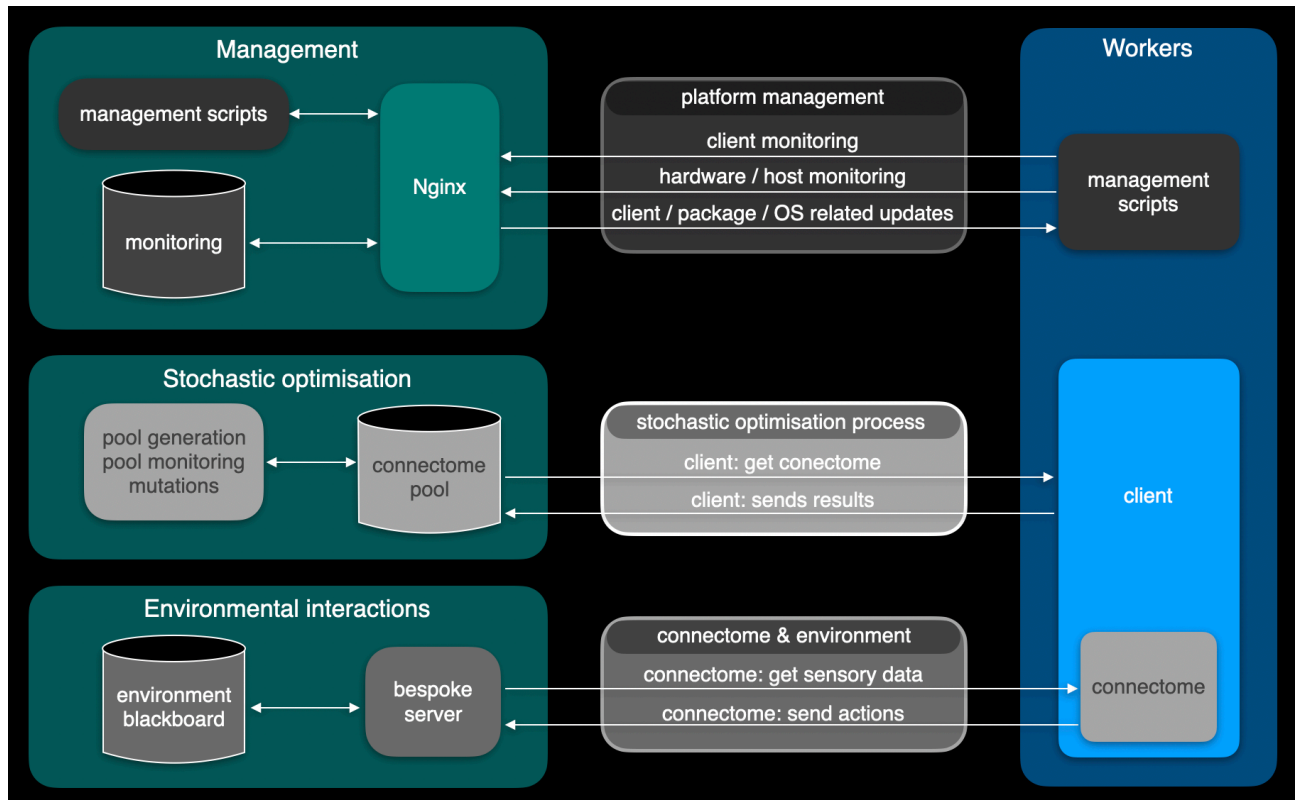


Figure 2. Functional representation of the cluster in the generic form.

In the present problem, there is no benefit in getting connectomes to interact with one another. Consequently, there is no need for an environment blackboard and its interface (the bespoke server) whose key jobs are (a) to provide sensory feedback after an action and (b) to arbitrate conflicting actions between connectomes.

However, each individual connectome still requires to get sensory feedback when performing an action on the dataset. We can't remove the environmental interactions all together.

For the problem at hand, the current thinking is to provide the connectome with an indication of the energy expanded to reach a new location (calculated on the basis of the wind speed available from ECMWF ancillary data or from S5P input files). Consequently, the environmental interactions stack will be split between a static repository of time series, and a context dependent environmental feedback embedded within the client. The result is depicted in Figure 3.

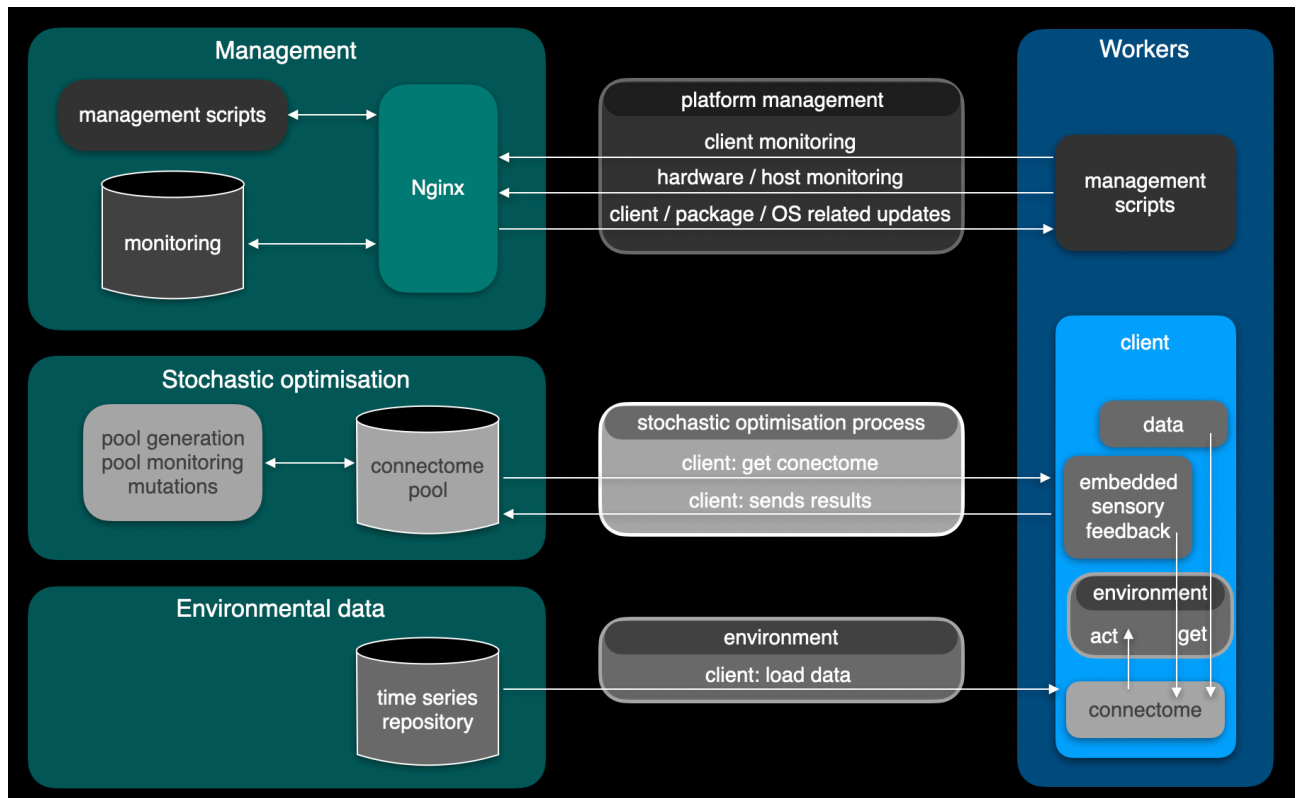


Figure 3. Functional representation of the cluster, simplified to suit a context where an environmental simulation is not needed.

### 3.1.4 Technical overview

In this section, we briefly present the technical organisation of the cluster in the latter form, i.e. the form tailored for the problem at hand (figure 3).

The platform management stack was designed in the MVC style. The code is written in PHP and BASH and leverages an http server (nginx) and a database in order to persist the existence of the workers, and maintain a picture of their transient characteristics (e.g. current IP, last ping received, online/offline) and states (update trigger, update status, a record of all the updates applied).

The stochastic optimisation works on a client-server basis, with a database as the server. There are two sides to this stack: the client and the backend. A client that has no connectome running fetches a connectome that is yet to evaluate. It does that within an atomic transaction that both SELECTs a single connectome that matches a number of criteria, and UPDATEs that record with the worker's hostname and the PID of the client onboard of the host. This allows to detect conditions (in relation to the connectome) that makes the client crash, and go and a bug hunt with some prior knowledge. The goal is indeed to remove inefficiencies in the process.

Regarding the environmental data, a database is ideally suited for the time series repository as the workers have no storage capability and must therefore load the data in RAM. This could be achieved with file sharing (e.g. NFS or SAMBA) but databases definitely allow for more flexibility in the retrieval of data (on the fly filtering and ordering). On the sensory feedback, we can note that embedding the function within the client is by construction more performant

than leaving it on another host for it saves on I/Os (mainly network latency). In the present work, the data is downloaded and imported in a database (Postgresql 11.6, PostGIS 3.0 3) in order to simplify data retrieval within time periods and geospatial constraints, as well as for performing geodetic distance and surface calculations (typically in order to obtain pixel size). The import is performed with a bespoke code written in C. The parameters stored in the database consists of: *[longitude]*, *[latitude]* *[nitrogen dioxide tropospheric column]*, *[nitrogen dioxide tropospheric column precision]*, *[cloud fraction crb nitrogen dioxide window]*, *[qa value]*, *[surface altitude]*, *[northward wind]* and *[eastward wind]* for files created with the processor 1.3.0, along with the source file's acquisition date and time. The database was queried to produce spatially constrained daily statistics as csv outputs (that could be provided as supplementary materials) for further treatment

All these scripts have been adapted from Thevenon's thesis. They were made available to Spascia to speed up the development of the cluster. These are not free to re-distribute. Copyright belongs to Imezio Ltd.

### **3.2 Data sets**

In this section we present the structure of the time series designed to meet the requirements of the connectomes (see environmental data in section 1, Figure 3).

#### **3.2.1 Primary data**

In this work we use TROPOMI Level 2 products of NO<sub>2</sub> tropospheric column, RPRO (reprocessed), OFFL (offline) and NRTI (near real time), distributed by the European Space Agency (ESA). The primary data source being used is Sentinel 5P's level 2 NO<sub>2</sub> Near-Real Time (NRTI) netcdf files, published by ESA and mirrored on ICARE. NRTI files are preferred over OFFL files as their smallest geographical footprint allows for a natural geographical indexing of the data within the database (figure 4). However, the NRTI shouldn't be used for analysis, because they have a 1 month lifetime only. OFFL data, and when available, RPRO data, must be preferred. NRTI will be used only for processing recent data or near real time data, when the OFFL products are not available yet.

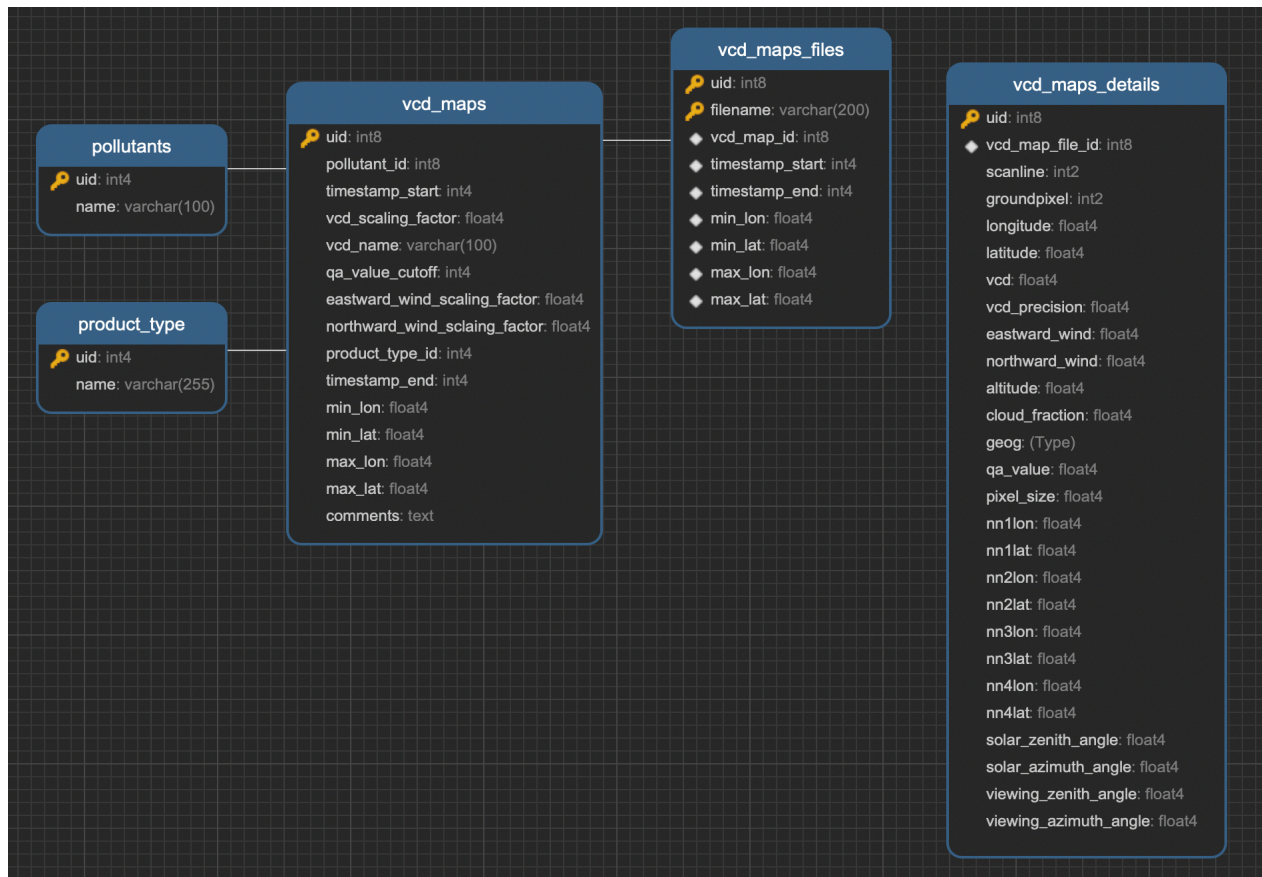


Figure 4. netcdf importation schema.

Only a tiny part of the information contained in the netcdf files is of operational interest to us at this stage of the project, namely the coordinates of each observation, the NO<sub>2</sub> concentration (vcd), wind. A number of additional fields were added for checks (zenith and azimuth angles, altitude, cloud fraction, qa\_value). The rest of the fields (nn\*lat, nn\*lon) were created in order to prepare the data structure used by the workers: the playgrounds.

The code that transforms the netcdf files into the data structure presented in figure 4 will be added to the code repository.

### 3.2.2 Playgrounds

The paradigm we use prompts us to provide the connectomes with data that is analog to the olfactory landscape a moth navigates in space and time.

We know that early connectomes will perform a random walk, however we can make no guess about what navigational patterns the best connectomes will develop. In both case, our best move is to provide the clients with a playground, i.e. a collection of daily maps centred around the same given location (preferably not a pollution source). Each map corresponds to a specific observation day. Playgrounds may be seen as cylinders, where the thinnest section is a day. Sections are ordered chronologically.

In order to allow for the clients to sense the NO<sub>2</sub> concentration of the next pixel at near zero-cost, each pixel of the cylinder is indexed, and the 8 nearest neighbours of each pixel are calculated and recorded. These nearest neighbours are in this order, the most Eastern

neighbour, the most Northern neighbour, the most Western neighbour and the most Southern neighbour, for both the next day and the previous day. We have decided to allow for time backward in order to test whether our artificial moths will make use of the ability to go backward in time. Beside these navigational indexes, each pixel comes with the observation day, the vertical column density, and the wind intensity along both the North-South and the East-West directions. The wind intensity will be used to calculate the effort incurred in moving from a pixel to another, which effort will be accrued on each move until the connectome dies of exhaustion. We will take this into account as an analog for survivability.

The result is the "compacted" table that is indexed like an array (index) and where the nearest neighbours in space and time are designated by their own index in the array. The code that transforms the data structure presented in figure 4 into the data structure presented in figure 5 will be added to the code repository.

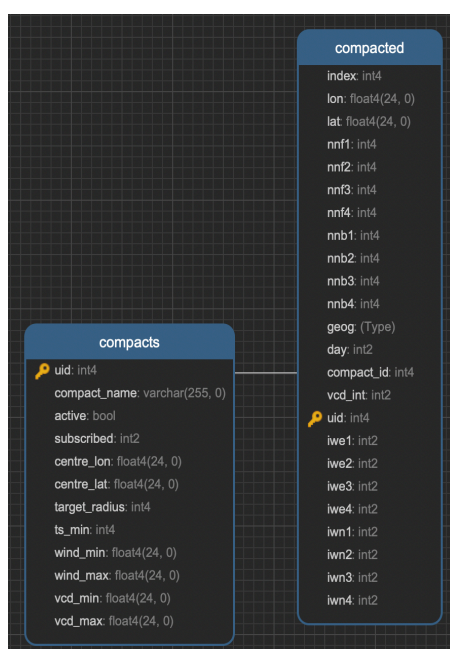


Figure 5. playgrounds schema.

A number of validation tests were performed on the playgrounds produced. The figures below are relative to a 250km radius cylinder centred on Toulouse.

	Rows: 2886382	
Global duplicates:	147258	non-dups: 2679732 (94.79%)
zeroes:	0	non-zs: 2826990 (100.00%)
Bearings potential issues:	14730	uncompromisingly ok: 2871652 (99.49%)

We find that the neighbour-construction algorithm succeeds in associating 99.49% of all the pixels to 8 neighbours whose bearings suit the definition given above. We also find that 5% of the pixels contained in the cylinder are duplicates. The cause is that several NRTI files were used to create each section and while they do overlap over Toulouse (the amount of overlapping depends on the latitude) we did not sanitise the duplicates. This may be a requirement for higher latitudes (e.g. nordic countries).

### 3.3 Pseudo code

As presented in section 1, the current instance of the cluster runs two software stacks: the platform management stack and the stochastic optimisation stack.

#### 3.3.1 Platform management stack

The code is written in PHP and BASH and leverages an http server (nginx) and a database in order to persist the existence of the workers, and maintain a picture of their transient characteristics (e.g. current IP, last ping received, online/offline) and states (update trigger, update status, a record of all the updates applied). The pattern is MVC.

Script	Purpose
alive.php	The workers call this script every 2 minutes (/etc/init.d/spsc_ping.sh /usr/bin/spsc_ping.sh). The first argument is the worker's MAC address; it serves as a unique identifier. The second argument is the worker's IP address, that may later be used to ssh to the worker (specifically in the development cycle). When called, the script updates the record that has the same MAC as the first argument, with both the IP address received and the epoch-based timestamp of the call. The timestamp is necessary in order to detect PIs that have stopped communicating with the hub for an abnormally long period of time. If the MAC doesn't exist, a new record is created with the provided data. Returns UPDATE when an update is scheduled for this worker, LED when the worker needs to be physically identified (a led is turned on), or OK when no action is required from the worker.
update.php	The workers call this script when UPDATE is received (see alive.php). This is a general purpose bulk update system. The expected sequence is for the worker to install, execute, and report back to the cluster on the outcome of the update. This script provides the required information to the workers given their progression in the sequence, and the possible errors that may occur.

Future development includes to divide all the functions performed by the workers' /usr/bin/spsc\_ping.sh in order for them to be performed asynchronously, and to build more flexibility in the server-side update system.

#### 3.3.2 Stochastic optimisation stack

The stochastic optimisation works on a client-server basis, with a database as the server. There are two sides to this stack: the client and the backend.

##### 3.3.3 Client

Clients are first and foremost the embodiment of a connectome within a given environment. Therefore, the internal configuration of a client depends on the connectome it embodies.

Nonetheless, the algorithm performed around a connectome is generic and consists in two sequential tasks:

- (i) to sense the environment: the connectome's sensory nodes are fed with the environmental data in their (location, time) context,

- (ii) to carry out the impacts of any firing node on the whole connectome: when the potential of a node reaches the firing threshold, this node releases that potential to the downstream nodes it is connected to.

Clients perform ancillary tasks including:

- environmental arbitrations when the task is not centralised on a server (see 1. Functional presentation of the cluster),
- logging evaluation results in the back-end when an evaluation is complete,
- fetching a new connectome from the back-end results are logged. This is performed within an atomic transaction that both SELECTs a single connectome that matches a number of criteria, and UPDATEs that record with the worker's hostname and the PID of the client onboard of the host.

### 3.3.4 Back-end

The back-end handles the connectomes pool, the ranking of the connectomes according to the results returned by the clients, and the mutation of the connectomes with the best performances (according to the objective function).

Module	Function
generator.c	Creates a random connectome and writes it in the database given a number of sensors, actuators, and inter-neurons, as well as an average number of interconnections per node.
generator.sh	A shell script that produces a pool of random connectomes in bulk, using generator.c
monitor.php	A shell script that maximises the productivity of the cluster by triggering a bulk mutation when the number of connectomes that remain to be evaluated is getting low when compared to the cluster's processing speed.
mutator.c	Randomly mutates a given connectome, according to specifications passed as arguments, such as: average number of connections and firing patterns to be modified.
mutator.php	A shell script that performs the bulk mutation of a number of connectomes, when triggered by monitor.php (these functions must be performed asynchronously).

## 3.4 Next steps

- Adaptation and compilation of the stochastic optimization stack on the back-end.
- Installation of the stochastic optimization database
- Adaptation of the stochastic optimization database to the objective function(s).

## 4 Detection of gas plumes and sources: Preliminary tests on a dedicated use case : monitoring atmospheric pollution level at city scale during covid 19 event

---

The Sentinel 5P's level 2 NO<sub>2</sub> database presented in the previous section has been exploited in a preliminary testing phase with a simple, physically-based, method for detection and analysis of the plumes. The goal of this preliminary processing test was both to use and test the primary data in a classical way for evaluation and validation purpose in support of our New AI method implementation, and to propose a first analysis of these data in a dedicated use-case: the provision of quantitative product for monitoring atmospheric pollution level at city scale during covid 9 event.

Nitrogen dioxide (NO<sub>2</sub>) is a good indicator for air quality (AQ) in urban and industrialised areas. TROPOMI provides high quality daily measures of NO<sub>2</sub> tropospheric column from space. Covid-19 and associated lockdown measures by cities is an unprecedented example of human activities/emissions reduction, that significantly impacted AQ and NO<sub>2</sub> content in urban areas. This preliminary use case aims at assessing the impact of the lockdowns that occurred in Europe from February to May 2020 on the NO<sub>2</sub> tropospheric content from space.

We implemented a dedicated processing for selecting the appropriate data and we computed a daily “NO<sub>2</sub> city plume mass” from TROPOMI NO<sub>2</sub> columns, representative of the NO<sub>2</sub> plume emitted by the city. We built time series of the NO<sub>2</sub> plume mass related to four European cities based on TROPOMI's data for 2019 and 2020. We analyzed the difference of these NO<sub>2</sub> signals between 2019 and 2020 at different temporal scales (daily, weekly, fortnightly, monthly), and we derived “temporal averaged reduction rates” as quantitative estimates of the NO<sub>2</sub> plume mass decrease due to lockdown, together with a dedicated estimate of their uncertainties.

On that basis, human activity reduction on the NO<sub>2</sub> tropospheric level is quantified for four cities over Europe. It is shown that the impact of human activity reduction on city NO<sub>2</sub> tropospheric level is clearly seen from space for 3 cities over 4. Quantification of this impact and the associated uncertainty is preliminary proposed.

### 4.1 Introduction

The ambient level of nitrogen oxides (NO<sub>x</sub> = NO<sub>2</sub> + NO) is a good indicator for air quality (AQ) in urban and industrialised areas. NO<sub>x</sub> atmospheric concentrations over urban areas are directly linked to human activities (fuel combustion from road traffic, residential and tertiary sector, industrial activities, e.g., EPA, 2008). The species NO<sub>x</sub> are produced during combustion processes and thus may serve as a proxy for fossil fuel-based energy usage as well as co-emitted greenhouse gases and other pollutants (Duncan et al., 2016). NO<sub>x</sub> emissions are regulated in many countries (e.g., Vestreng et al., 2009; UNECE, 2007; <https://www.unece.org/env/lrtap/welcome.html.html>), as NO<sub>2</sub> is unhealthy to breathe, and NO<sub>x</sub> is a necessary ingredient for the formation of surface ozone.

NO<sub>x</sub> can be measured at ground level but measurement networks are sparse and measured values are quite sensitive to the distance from major traffic routes. Pollution maps based on ground measurements involve modelling in combination with ancillary data (elevation, meteorology, inventories of pollution sources...). The results are plagued with large

uncertainties. ESA funded Copernicus Atmosphere Monitoring Service (CAMS) project is one of the worldwide efforts to reduce these uncertainties.

The Copernicus European program is also preparing the operational capacity of measuring pollution markers from space with the Sentinel 5 and 4 fleets. Satellite observations of NO<sub>2</sub> atmospheric and tropospheric columns are already regularly and globally provided by the Sentinel 5 precursor (S5P) high resolution space-based sensor, launched in October 2017 (Veefkind et al., 2012). The Tropospheric Monitoring Instrument (TROPOMI) onboard S5P provides daily global coverage at 5.5 km x 3.5 km spatial resolution (7 km x 3.5 km before 6 August 2019) and at 13:40 local time at equatorial crossing.

Tropospheric NO<sub>x</sub> has a relatively short lifetime of several hours to 1 day (Leue et al., 2001; Beirle et al., 2003; Beirle et al., 2011). As a result, NO<sub>2</sub> has its largest concentrations in the boundary layer close to the emission sources, making measurements of NO<sub>2</sub> column amounts from space well suited to infer NO<sub>x</sub> emissions. However, the solution is not a silver bullet either for the link between tropospheric NO<sub>2</sub> concentrations observed from space and the actual emissions of pollutants is complex, as illustrated for example by ECMWF/CAMS, 2020-1. The key reasons are

- the tropospheric NO<sub>2</sub> column measured by TROPOMI is impacted by other factors than the emission, such as regional meteorological conditions and long-range atmospheric transport of NO<sub>2</sub> (e.g., Zhang et al., 2008, Duncan et al., 2016).
- There is a complex photochemical equilibrium between NO<sub>2</sub>/NO, volatile organic compounds (VOCs) and O<sub>3</sub> coupled with the related diurnal variations (Kimbrough et al., 1994, Atkinson, 2000).
- NO<sub>2</sub> measurements from space are also impacted by clouds which limit the availability of continuous and uniform sampling and coverage.
- Another effect potentially impacting specifically the measurement of NO<sub>2</sub> from space is the solar zenith angle (SZA) at the time/position of the measurement. This quantity is modulating the solar flux at ground level, which impacts the direct and diffuse actinic fluxes driving the photodissociation of NO<sub>2</sub>, and thus the equilibrium NO<sub>2</sub>/NO for a (relatively stable) overall NO<sub>x</sub> content (e.g., Kimbrough et al., 1994). The relevant SZA (at the observation point) is varying significantly with seasons, and to a lesser extent with the satellite viewing zenith angle (VZA).

However tropospheric NO<sub>2</sub> measurements have been successfully exploited in many studies for estimating trends and variations in atmospheric concentrations at global, regional, country and city scales (Beirle et al., 2003; Van der A et al., 2008; Castellanos and Boerma, 2012; Hilboll et al. 2013; Schneider et al, 2015; DeFoy et al., 2016; Miyazaki et al., 2020), or for monitoring emission changes in local sources (Lorente et al., 2019; Griffin et al., 2019; Ding et al, 2015; Duncan et al., 2013).

The tropospheric vertical column density (VCD) of NO<sub>2</sub>, as derived from satellite measurements, is an effective proxy for surface NO<sub>x</sub> in many air quality applications (e.g., Lamsal et al., 2015). The total VCD is defined as the number of molecules in the vertical atmospheric column (per unit area) above the Earth's surface. The tropospheric VCD is obtained by subtracting the stratospheric NO<sub>2</sub> column from the total column (van Geffen et al., 2018), where the stratospheric NO<sub>2</sub> amount is estimated using information from a chemistry transport model by way of data assimilation (Boersma et al., 2004). The global

pandemic of Covid-19 resulted in a massive reduction of human activities in most countries (Hale et al., 2020; Wikipedia, 2020; Sandford, 2020; Breteau, 2020), with a particular impact on NO<sub>x</sub> emissions that have been reduced in many industrial countries and cities around the world (e.g., Forster et al., 2020; Zhang et al., 2020; AirParif, 2020). Several communications have associated lockdowns resulting from Covid-19 with a decrease in NO<sub>2</sub> air pollution as observed from space by TROPOMI (e.g., ESA, 2020; Bauwens et al., 2020; MacGarth, 2020), suggesting that TROPOMI data could be exploited to monitor the NO<sub>2</sub> level associated to pollution plumes emitted by cities, and to quantify the impact of human activity reductions resulting from the spread of Covid-19.

This paper wants to figure out whether year on year variations of NO<sub>2</sub> tropospheric column as observed from space by TROPOMI, considered at the appropriate city scale and over representative periods of time, are undoubtedly impacted during the lockdown periods and can be used as a proxy to quantify city pollution reduction associated with variations of human activities at city scale.

In order to achieve this goal, the ESA L2 VCD data are used through a dedicated city-scale processing, that will include proper filtering and suitable spatio-temporal averaging methods for generating average maps over cities. We then computed daily “NO<sub>2</sub> city plume mass” from TROPOMI NO<sub>2</sub> columns, directly (i.e., without additional information than the TROPOMI product) representative of the NO<sub>2</sub> plume emitted by the city. Time series of the NO<sub>2</sub> plume mass related to four European cities based on TROPOMI data for 2019 and 2020 are used to analyse the year-to-year difference of these NO<sub>2</sub> signal at different temporal scales (daily, weekly, fortnightly, monthly), and to discuss the multiple causes of these temporal variations as listed above. Examined confounding variables are mainly the variability of the meteorological conditions (wind, temperature), the photochemical equilibrium, the impact of NO<sub>2</sub> sources located outside the main city perimeter. We derived “temporal averaged reduction rates” as quantitative estimates of the NO<sub>2</sub> plume mass decrease due to lockdown, together with a dedicated estimate of their uncertainties. We associated these results with the respective lockdown periods in the four cities within so-called confusion matrices.

The present analysis aims at providing objective evidence of direct correlation between space-based measured NO<sub>2</sub> tropospheric column and lockdown episode and a first quantification and monitoring of the impact of reduction of human activities on the NO<sub>2</sub> atmospheric levels over target cities from S5P/TROPOMI products, together with estimation of the uncertainties. In the present work we focus on Paris, Milan and the Po valley, Madrid, and Athens, that applied lockdown measures on 17 March 2020, 8 March 2020, 14 March 2020, and 23 March 2020, respectively (Wikipedia, 2020; Sandford, 2020). Note that these dates are indicative, because the actual implementation of governmental containment and lockdown policies at the local and national scales were complex and progressive. A detailed information on these measures, as well as a comprehensive work and synthesis effort for deriving usable criteria for objective analysis, is provided by Oxford University’s coronavirus government response tracker ([www.bsg.ox.ac.uk/covidtracker](http://www.bsg.ox.ac.uk/covidtracker)).

## 4.2 Data and methods

### 4.2.1 Satellite measurements database

TROPOMI is a single payload on board of the Copernicus S5P, one of the European Sentinel satellites dedicated to the monitoring of atmospheric composition, launched on the 13 October 2017. The instrument is a four spectrometers system that measures in the ultraviolet (UV), UV–visible (UV–VIS), near-infrared (NIR) and shortwave infrared (SWIR) spectral bands. The NO<sub>2</sub> columns retrieval algorithm developed by the Royal Netherlands Meteorological Institute (KNMI) exploits the UV–VIS TROPOMI measurements in the 405–465 nm wavelength range [33]. The TROPOMI NO<sub>2</sub> processing system is based on the DOMINO, an algorithm previously used for the OMI instrument [Van Geffen et al., 2019]. In this work, the offline Level 2 NO<sub>2</sub> tropospheric column data for the periods January–April 2019 and January–April 2020 have been used. Two other versions of data are also available: the near-real-time (NRTI) product, available within 3 h after data acquisition, and the reprocessed (RPRO) product (Eskes et al., 2019). This data is regularly downloaded from the Copernicus open data access hub (<https://s5phub.copernicus.eu>) by the ICARE/AERIS French data centre (<ftp.icare.univ-lille1.fr>), providing archive of historical data and on-the-fly access to recent data (typically a few days delay) to the French scientific community. The data is downloaded and imported in a database (Postgresql 11.6, PostGIS 3.0 3) in order to simplify data retrieval within time periods and geospatial constraints.

We get the data for 4 target areas centered on Paris, Milan and the Po valley, Madrid, and Athens. The selection of these areas is described in the next section.

For the analysis of our results, Level 2 NO<sub>2</sub> TROPOMI data are compared with NO<sub>2</sub> observations obtained by the air quality monitoring stations distributed on the Paris area. The stations are managed and provided by the Association Interdépartementale pour la gestion du Réseau automatique de surveillance de la Pollution Atmosphérique et d’alerte en Région Ile-de-France (AIRPARIF) (AIRPARIF data are available online : <https://www.airparif.asso.fr/telechargement/telechargement-station>). The network consists of 40 automatic monitoring stations for the Paris area. The surface concentration of NO<sub>2</sub>, measured in  $\mu\text{g}/\text{m}^3$ , is available on a time slot of 1 h. From these data, hourly values at 12h UTC have been extracted, and daily averages have also been computed, for comparison with consistent satellite observations.

### 4.2.2 Processing of the data at city scale

We generated data subsets for the 4 considered cities (Paris, Madrid, Athens, Milan region) and 2 periods of 4 months, from January to April 2019 and 2020. The data was processed according to the following four-step procedure:

**City scale area selection:** Datasets are prepared by selecting pixels within a 500 km x 500 km area centred on Paris, Madrid, or Athens and within a 500 km x 630 km area centred on Milan and the related Po valley. S5P L2 NO<sub>2</sub> pixels are collected each day from the overpass(es) (1 or 2 overpasses) within the corresponding areas.

**Data filtering:** Within each area and for each day, we filtered out the pixels whose quality assurance value is lower than 0.5 and the cloud fraction (within the NO<sub>2</sub> retrieval spectral

window) above 40%. This choice is consistent with the recommendations for data usage provided by S5P product documentation (Eskes et al., 2019), and allows to maximise the available pixels to derive the plume mass. As we keep pixels over clouds with good quality retrievals, we apply a supplementary filtering on the cloud fraction. For the remaining dataset we generated two (2019 and 2020) 3-month time series of daily data boxes of filtered NO<sub>2</sub> pixels for the 4 cities.

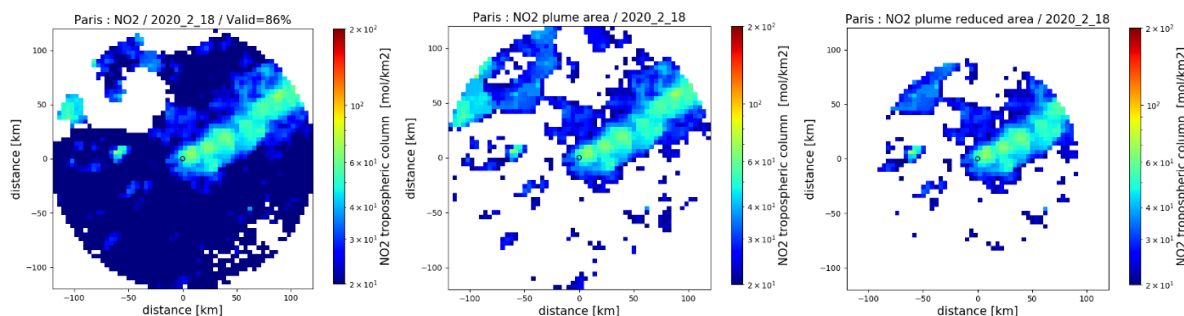
*Spatial gridding and interpolation:* The daily samples of the four areas were subsequently processed as follows. Figure 1 illustrates the result.

- For each day of the two 4-month periods, we consider the filtered pixels within a disk of 240 km diameter (Paris, Athens), 300 km diameter (Madrid), or 500 km diameter (Milan), for properly accounting for their respective extension. The NO<sub>2</sub> tropospheric column is interpolated on a regular 4 km x 4 km grid using the Inverse distance weight (IDW) method which computes the gridded values from the 3 nearest filtered pixels, weighted by the inverse of the distance of each pixel from the grid point. The 4km x 4km grid is chosen to keep close to the original typical pixel size of the TROPOMI measurements, since a regridding procedure at a better resolution (has done, e.g., by Cersosimo et al., 2020) is not judged necessary for our analysis of large city plumes. IDW method put more weight to measurement closest to the interpolation point, reducing interpolation errors (note that the remaining interpolation errors will be removed in the computation of the NO<sub>2</sub> mass of the plume).
- In order to properly interpolate spatially on the regular grid only where and when it is possible, we applied the following constraint: a gridded value is computed only if its position is surrounded by at least 3 filtered pixels located at a distance less or equal to a fixed threshold distance. Grid points that do not satisfy this criterion are left empty. The interpolated gridded pixels constitute the set of valid pixels for the given day (Figure 1a). For the rest of the analysis, we retain only the days whose percentage of valid pixels is higher than 80% of the total number of gridded pixels within the disk. This conservative value has been chosen empirically to limit the ratio of non-informative pixels for the computation of the plume mass.

*Spatial integration of the NO<sub>2</sub> mass and uncertainty estimate:*

- The total NO<sub>2</sub> mass of the plume for the selected days is calculated by integrating only the pixels whose NO<sub>2</sub> VCD (in mol/km<sup>2</sup>) is higher than a constant background value (Figure 1b). The background value was empirically set to 10 mol/km<sup>2</sup> for Madrid and Athens, and to 20 mol/km<sup>2</sup> for Paris and Milan (values adapted to the noise and the area covered by the corresponding images). The term “plume” is used here because the background is subtracted from the mass integration. The definition for this constant background is justified by our choice for a simple and systematic approach, but could be further improved for deriving absolute values or refine the analysis of the relative decrease/increase between 2019 and 2020 (which could be impacted by continent-wide implementation of lockdown measures or by other effects such as a warmer winter 2020 over Europe significantly limiting the use of heating power at continental scale (e.g., Cersosimo et al., 2020)).

- The uncertainty of the plume's total NO<sub>2</sub> mass is estimated for each day, by computing the difference between the plume's total NO<sub>2</sub> mass and the mass computed for a disk with a smaller diameter (75% of the nominal size) (Figure 1c). This method is providing a reasonable upper estimate of the uncertainty, since for a radius larger than the nominal size, the contribution from the target city is supposed to decrease rapidly.



**Figure 1 : Example of NO<sub>2</sub> tropospheric column maps (displayed in logarithmic colour scale) over Paris area for a given day (18/02/2020) for illustrating the data treatment: a) the valid pixels are the pixels of a regular 4 km x 4 km grid effectively interpolated from the filtered pixels (quality index higher than 0.5, cloud fraction lower than 40%) using the IDW interpolation criterium (all grid points framed by 3 filtered pixels closer than a threshold distance) within a 250 km diameter disk; the day is selected if the percentage of valid pixels in the disk is higher than 80%; b) only valid pixels with NO<sub>2</sub> column higher than 10 mol/km<sup>2</sup> are considered for calculating the total NO<sub>2</sub> plume mass within the disk; c) the NO<sub>2</sub> plume mass for a smaller disk (75% of the initial diameter) is also computed, and the difference with the total NO<sub>2</sub> plume mass provides the uncertainty estimate on the total NO<sub>2</sub> plume mass.**

The plume's daily total NO<sub>2</sub> masses for Milan, Madrid, Paris and Athens over the two periods, were subsequently processed in order to produce daily and monthly maps, as well as time series (7-day, 15-day and monthly averaging windows). The results are discussed in the next sections.

It is stressed that the analysis method proposed and used in this study is limited to a quantitative estimate of the pollution levels over target cities (using NO<sub>2</sub> tropospheric level as a proxy) and the analysis of its temporal evolutions for discussing pollution response to city emission regulations. A simple approach has been chosen (without characterisation of plume limits, and without exploitation of wind information) that allow dealing with all kind of plums and wind conditions. It does not pretend being as advanced as emission quantification methods from space such as that pioneering developed by Fioletov et al (2016) on OMI SO<sub>2</sub> data, or those implemented by Lorente et al (2019) or Goldberg et al. (2019) from TROPOMI NO<sub>2</sub> data. These published methods derive quantitative estimates of the emissions by exploiting wind information. We choose to define a simple method only using the available data from the TROPOMI NO<sub>2</sub> retrieval product. Thus, our method does not consider the wind variability that impact the NO<sub>2</sub> concentration and does not derive NO<sub>2</sub> fluxes. Nevertheless, it is expected that the careful analysis of temporal variabilities and the use of monthly mean values will tend to limit that impact by smoothing the high frequency temporal variability of the wind.

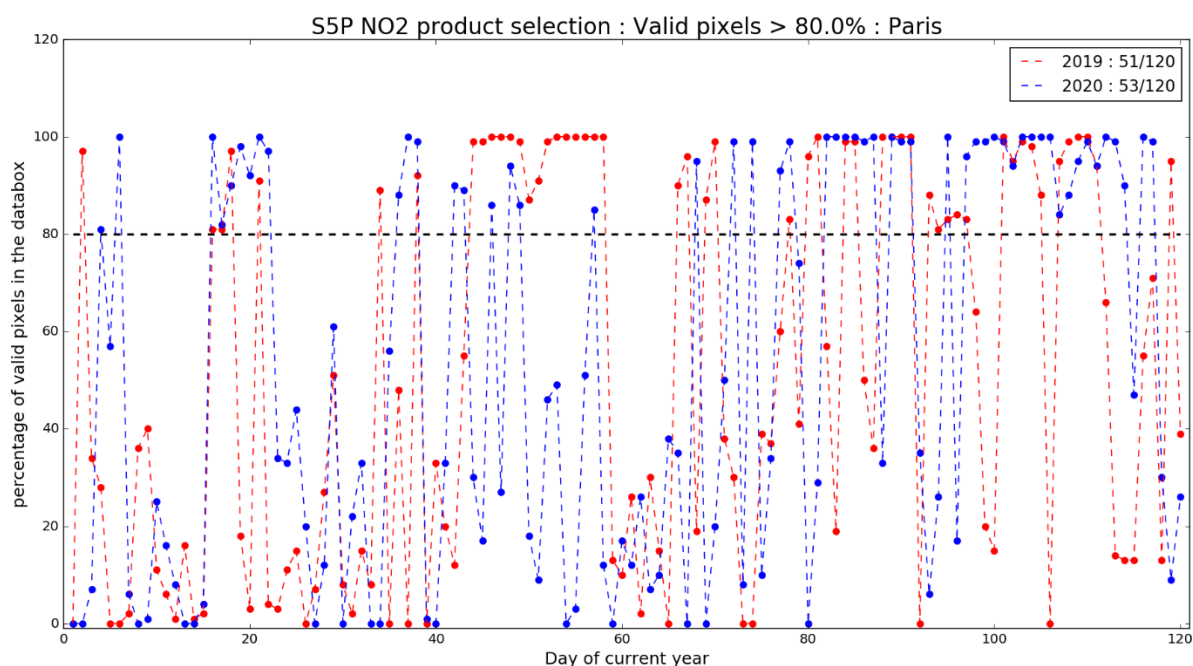
### 4.3 Results for Paris

#### 4.3.1 Qualitative analysis of daily data

Figure 2 presents the percentage of valid pixels for each day for the two (2019 and 2020) 4-month time series. The percentage of valid days over 4 months (120 days) is within 50-75% for Madrid, Milan and Athens, and approximately 40% for Paris (due to frequent cloud coverage over the period considered). Figure 2 shows that several periods are largely under-sampled: (i) 1 or 2 valid days for the first half of January 2019 and 2020, (ii) sparse days for first half of February 2019 and February 2020, (iii) no valid days for the first week of March, and (iv) sparse data for the first half of March (for both years). The averages calculated for these periods are therefore not as robust as we would wish for. The robustness improves for the second half of March and whole April.

Figure 3 presents typical maps of NO<sub>2</sub> tropospheric plumes in the Paris area for selected days in the January-February / 2019-2020 periods. The day-to-day variability is of the same order of magnitude as the inter annual day-to-day differences and can be essentially attributed to three effects: meteorology (wind variability), atmospheric transport of pollutant from sources outside the city, and Paris's own pollutant emissions.

Figure 4 presents similar maps for the end of March period, which in that case illustrates more systematic differences between 2019 and 2020 daily maps, despite a small number of valid days for this period. On the contrary, longer time series of valid daily NO<sub>2</sub> images for both April 2019 and 2020 are available. They are shown in Figure 5 and Figure 6 for first half and second half of April, respectively. The qualitative systematic differences between 2019 and 2020 clearly suggest a local emission reduction from 2019 to 2020. However, it is not possible to conclude based only on the examination of daily maps, and the analysis of averaged quantities is required. If the small number of valid days in March limits this analysis, the data of April allows to derive reliable estimates.



**Figure 2 : Daily selection criteria for Paris for 2019 (red) and 2020 (blue) periods (January-April). Percentage of valid pixels in the disk (the day is selected if the percentage of valid pixels is larger than 80%), number of selected days over total number of days are indicated in the legend.**

#### 4.3.2 Quantitative analysis of averaged NO<sub>2</sub>

#### 4.3.3 Discussion of time series of averaged NO<sub>2</sub> and of perturbing effects

Figure 7 compares weekly (7-days), fortnightly (15-days) and monthly averages of total NO<sub>2</sub> plume mass from January to April 2019 and 2020. The quantities shown are weighted means  $\langle M \rangle$  of the daily values weighted by their relative uncertainty, computed as follows:

$$\langle M \rangle = \frac{\sum w_i M_i}{\sum w_i} \quad \text{with} \quad w_i = \frac{1}{\left(\frac{eM_i}{M_i}\right)^2}$$

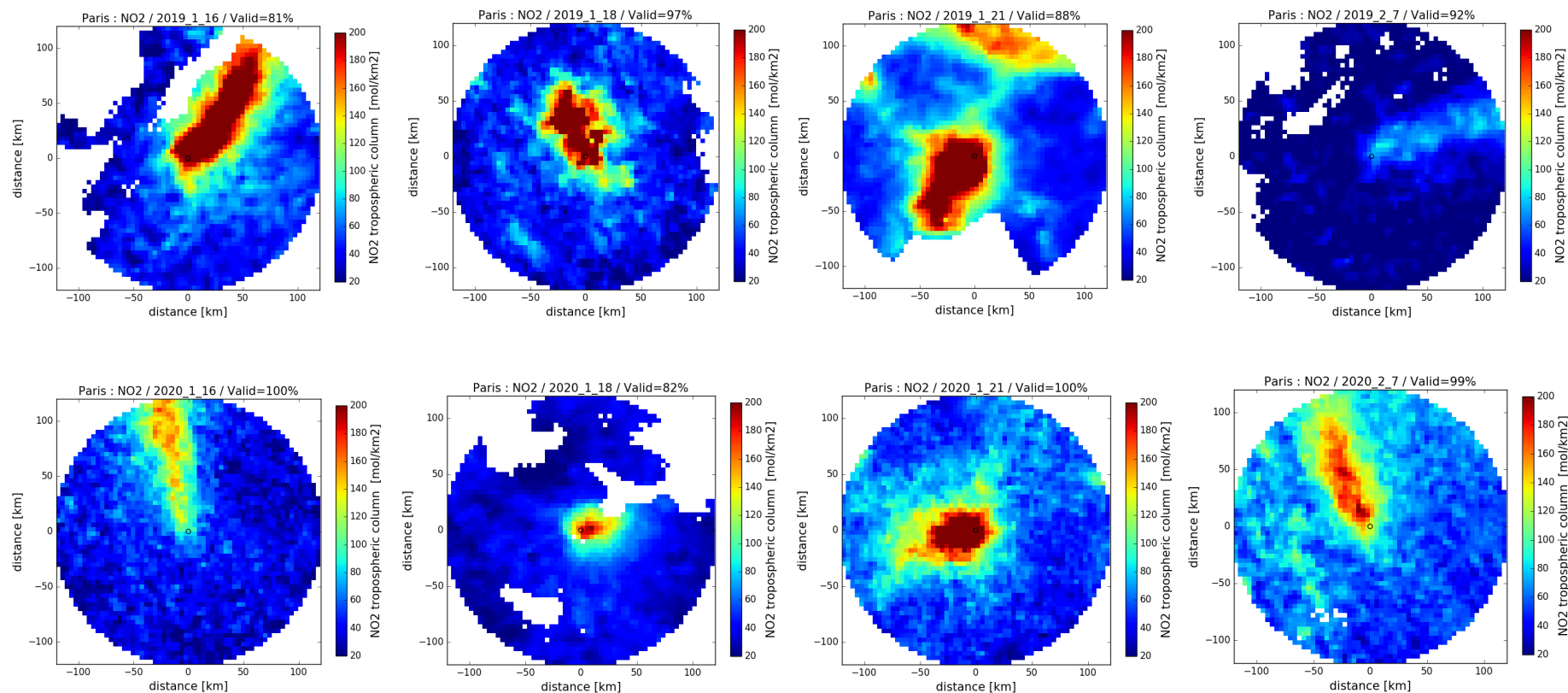
Where  $M_i$  is the NO<sub>2</sub> mass single daily retrievals,  $eM_i$  is the uncertainty on  $M_i$ , and  $w_i$  is the weight of daily retrieval. The associated bars represent the relative uncertainty on the weighted average, computed as:

$$e\langle M \rangle = \frac{\langle M \rangle}{\sqrt{\sum w_i}}$$

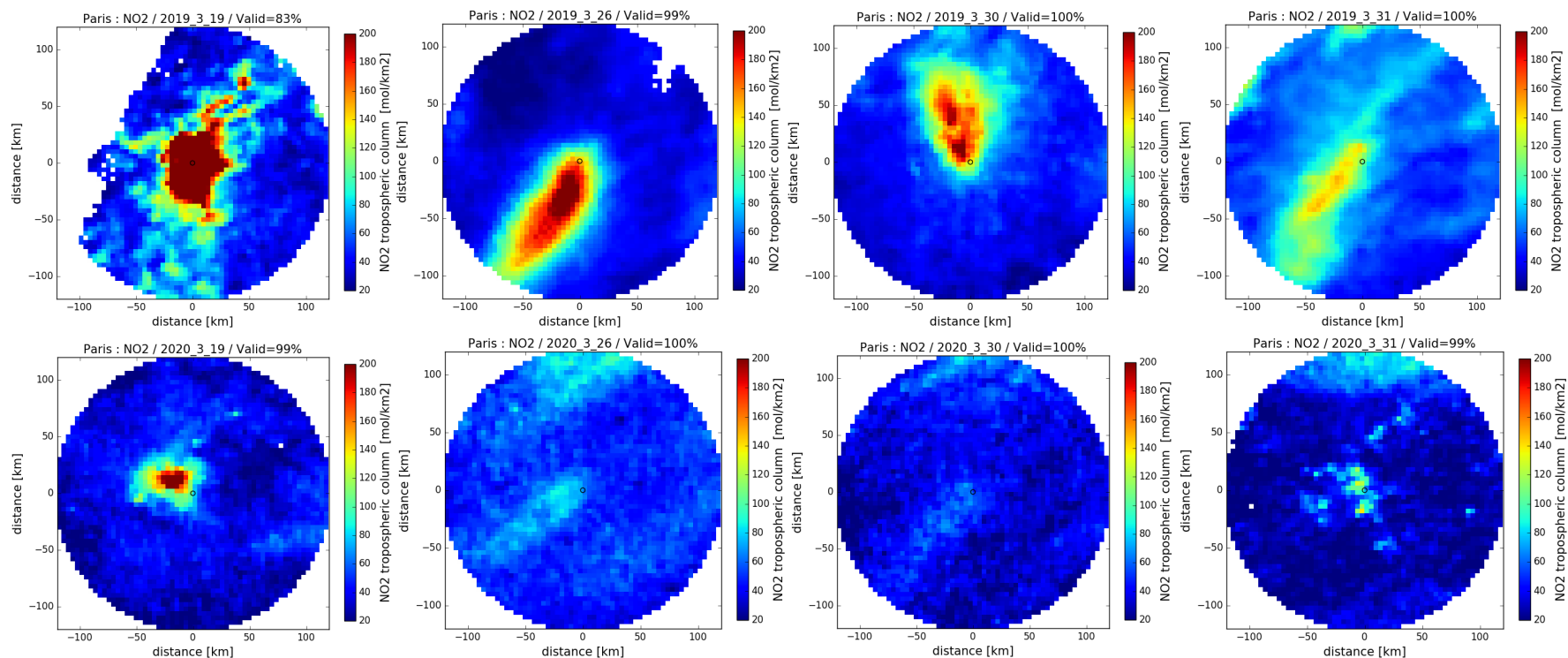
Missing values indicate the lack of valid data (i.e., less than 2 values within the considered temporal window) for building reliable averaged quantities.

The monthly average comparisons (Figure 7, top) demonstrate a clear and robust decrease of NO<sub>2</sub> in March and April 2020 when compared to the same months in 2019. The corresponding fortnightly values indicate a real start of the decrease from mid-March onwards. The analysis (disaggregation) of April values into fortnightly then into weekly averages suggest a significant decrease in the first half of April. However, the second half of April suffers from missing values in 2019, preventing a robust interpretation of the 2019/2020 relative values for this period.

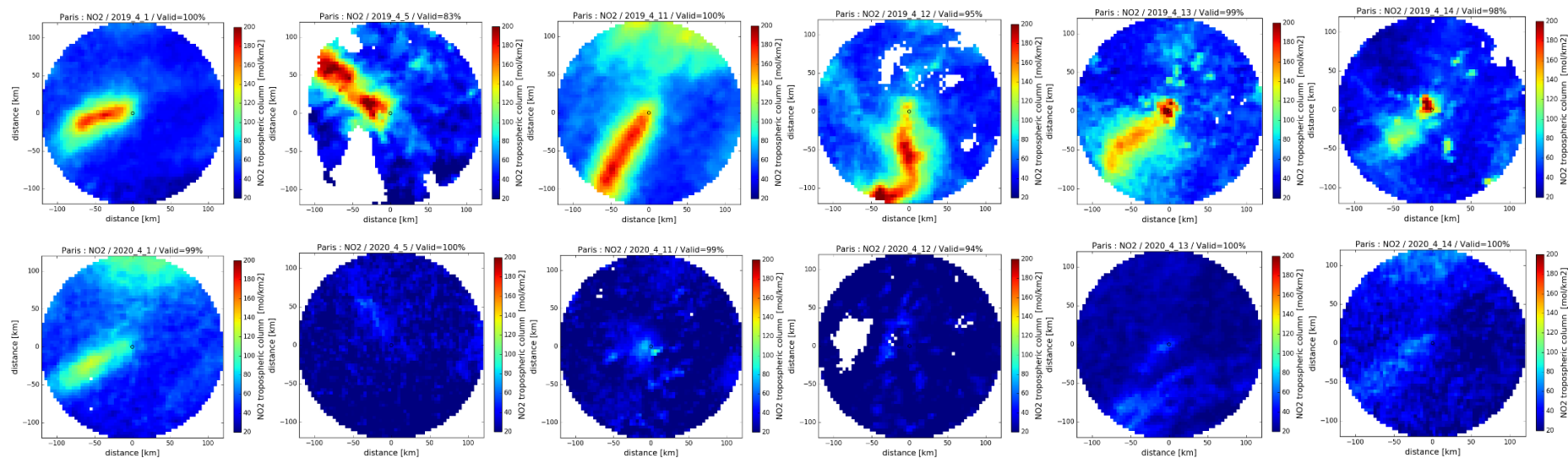
In some cases, we observe a large variability between consecutive 15-day mean values, that suggests a significant impact of meteorological conditions and/or atmospheric transport. This occurred in the first half of March (Figure 7, middle), which are noticeably lower than the values for end of February and end of March. The corresponding weekly averaged values (Figure 7, bottom) are also low and the mean value for the second week of March 2019 is missing. This behaviour can be explained by (a) the small number of valid daily data and (b) the low level of NO<sub>2</sub> tropospheric columns for the corresponding valid data, potentially due to high wind.



**Figure 3 : Daily maps of NO<sub>2</sub> tropospheric column over Paris area, for selected days in January/February. Top: 4 days in 2019. Bottom: same 4 days in 2020.**



**Figure 4 : Daily maps of NO<sub>2</sub> tropospheric column over Paris area, for selected days of the end of March. Top: 4 days in 2019. Bottom: same 4 days in 2020.**



**Figure 5 : Daily maps of NO<sub>2</sub> tropospheric column over Paris area, for selected days of first part of April. Top: 6 days in 2019. Bottom: same 6 days in 2020.**

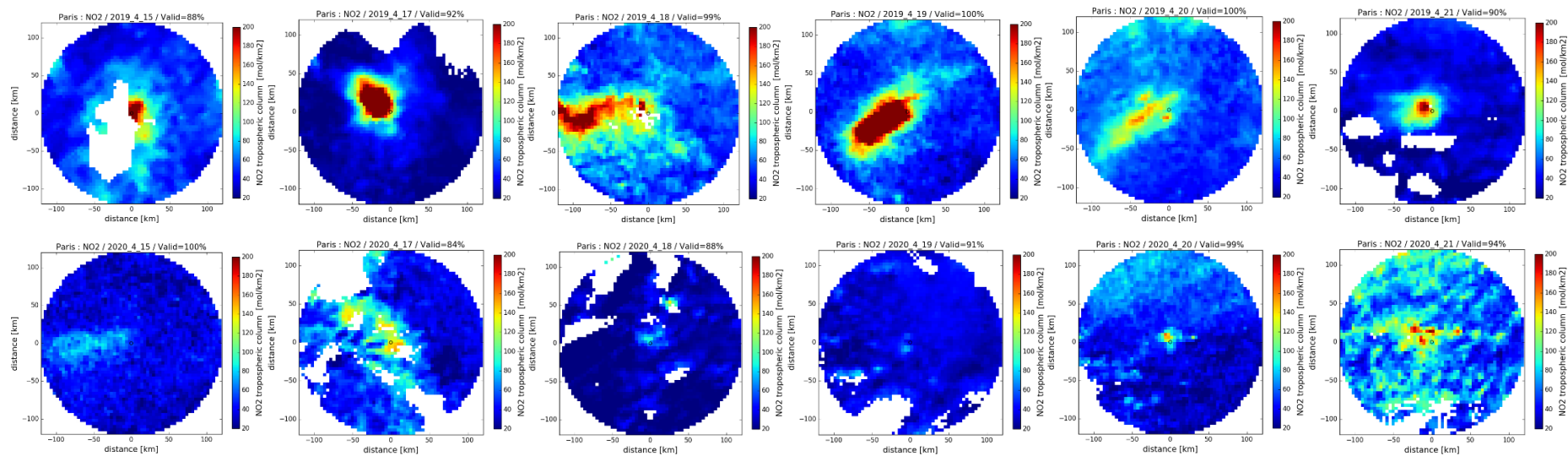
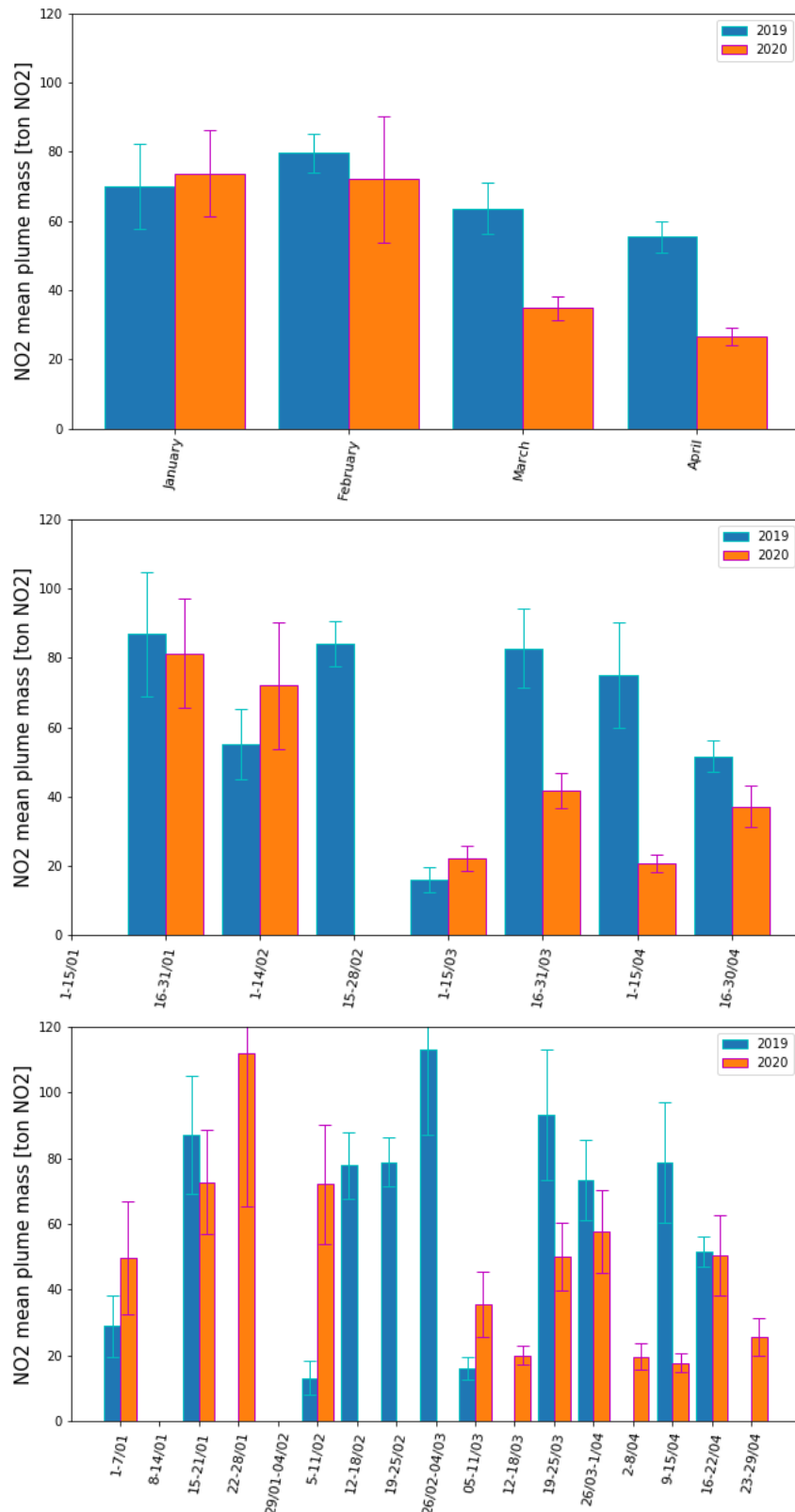
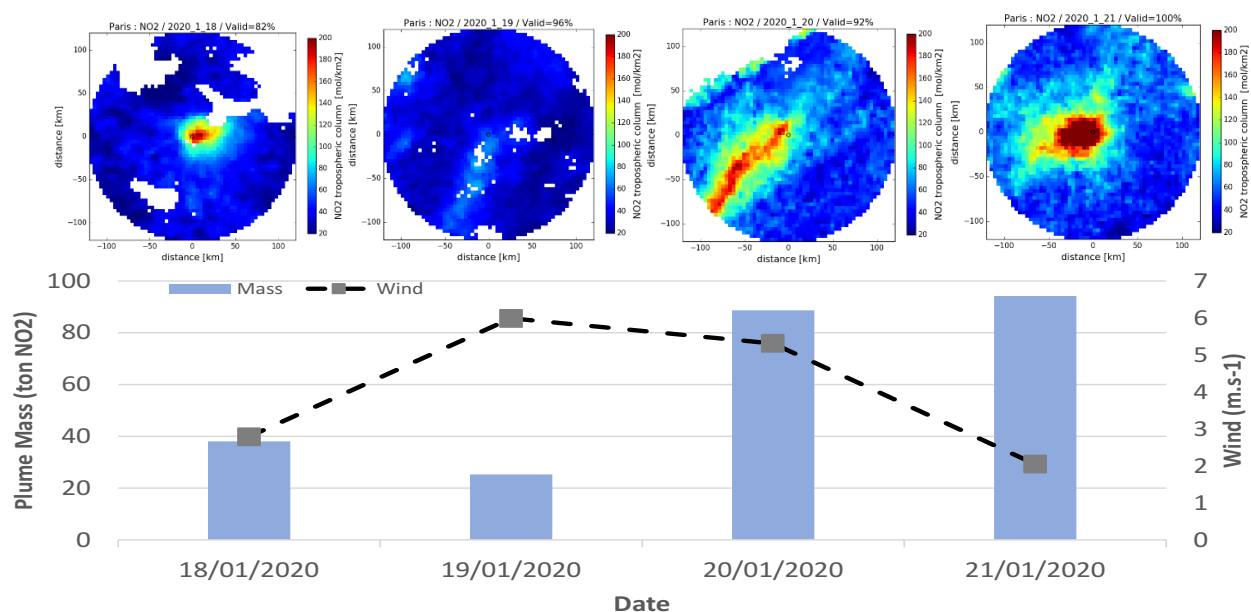


Figure 6 : Daily maps of NO<sub>2</sub> tropospheric column over Paris area, for selected days of second part of April. Top: 6 days in 2019. Bottom: same 6 days in 2020.



**Figure 7 : Temporal variations of total NO<sub>2</sub> mean plume mass over Paris region for the January-April period of 2019 (blue) and 2020 (orange). NO<sub>2</sub> plume mass is averaged over 1-month (top) 15-day (middle) and 7-day temporal windows (bottom). Missing values indicate that less than 2 daily values are available for the considered temporal window. Uncertainty bars on the mean values are derived from daily uncertainty estimates.**

As an example, an illustration of the impact of wind variability can be found by examining a series of consecutive daily images. Figure 8 shows the images from 18<sup>th</sup> to 21<sup>st</sup> January 2020, and correlates well with the corresponding values of NO<sub>2</sub> plume mass (estimated in this work through the procedure described in Section 4.2.2) and with a typical wind speed (the average of 10 meters winds for all the valid pixels, as provided by the S5P L2 product). The significant short-term reduction of the level of tropospheric NO<sub>2</sub> for the 19<sup>th</sup> of January (compared with the levels observed the days before and after) cannot be solely explained by a decrease of Paris's own emissions, however the relatively strong wind pushing the NO<sub>2</sub> out of the city may be an explanatory factor. The increasing NO<sub>2</sub> mass for the two days after (20<sup>th</sup> and 21<sup>st</sup> of January) is probably due to the combined effects of wind reduction/change and emission increase, illustrating the complexity of processes affecting air quality.



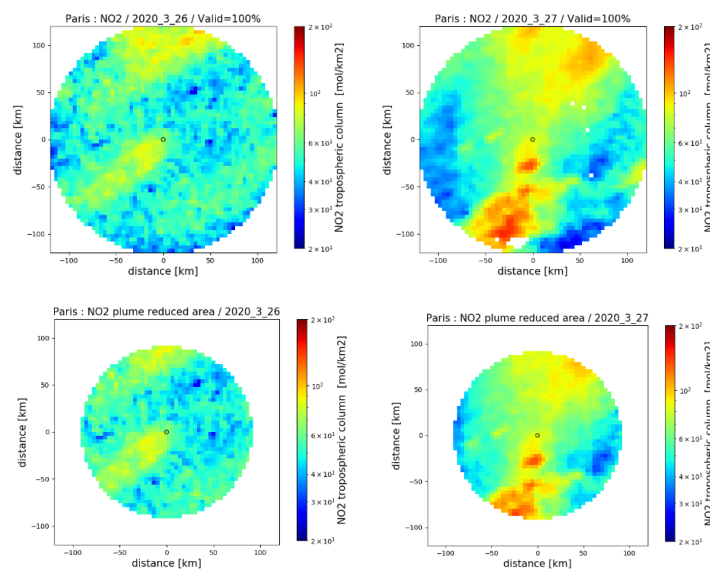
**Figure 8 :** Top: Daily maps of NO<sub>2</sub> tropospheric column over Paris region, for 4 consecutive days: 18<sup>th</sup>, 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup> January 2020. Bottom: corresponding value of estimated NO<sub>2</sub> plume mass (blue bar) compared with mean 10 meter wind (average over the domain of the ECMWF 10 meter wind provided in the S5P L2 product)

Conversely, Figure 9 illustrates the impact of NO<sub>2</sub> atmospheric transport from outside the domain on the total NO<sub>2</sub> plume mass. It shows the daily maps of NO<sub>2</sub> tropospheric column over Paris for two illustrative days at the end of March, presenting high concentrations of NO<sub>2</sub> in the north of the area due to atmospheric transport from outside the domain (probably coming from Benelux or Germany). This effect contributes to overestimate the total NO<sub>2</sub> plume mass that can be attributed to Paris. However, in such a configuration, the uncertainty value associated to the total NO<sub>2</sub> plume mass estimate, obtained from the difference with the plume mass over a smaller disk, provides a reasonable estimator of this impact (also illustrated in Figure 9).

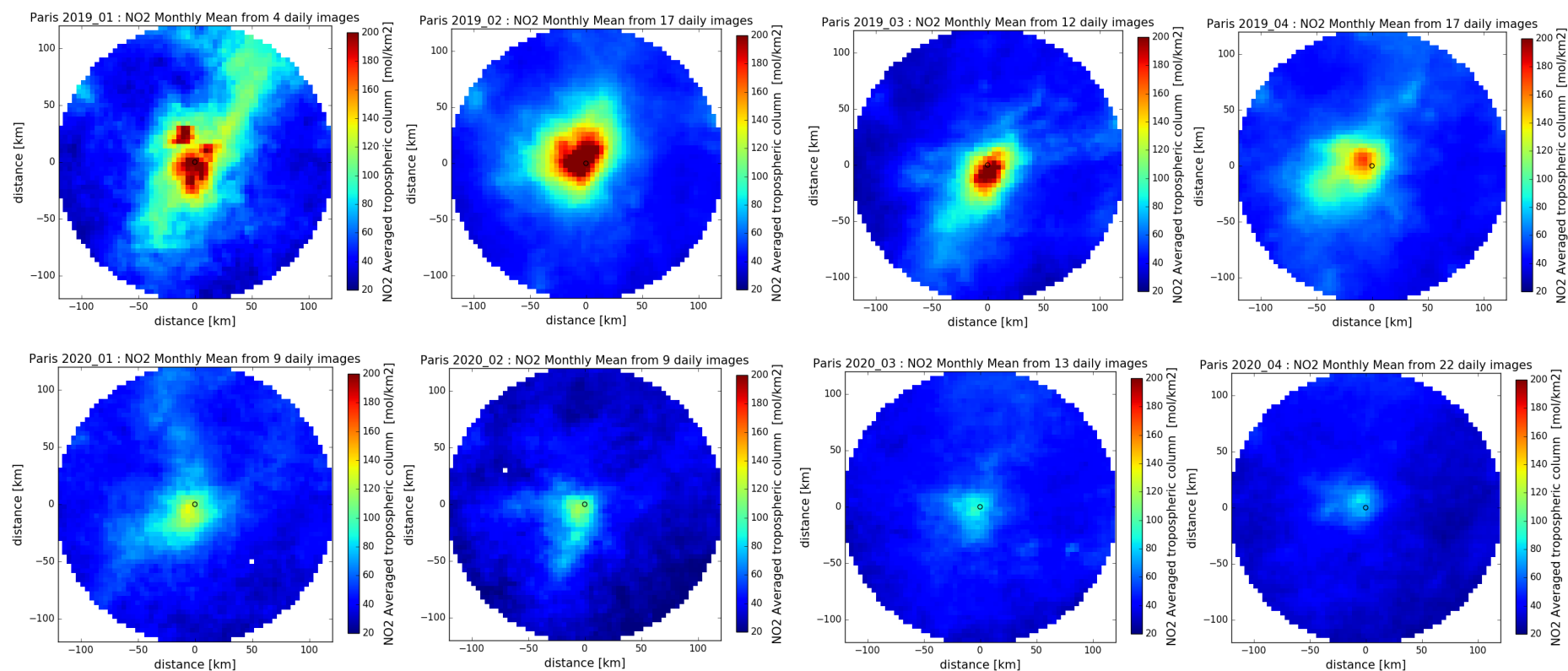
These two latter examples demonstrate that properly dealing with average quantities and exploiting the uncertainty values can help with mitigating the importance of meteorological perturbations such as those discussed above. Effectively, this self-sanitising process would allow to assess city emission reductions by leveraging the S5P data. These examples also confirm the critical importance of discussing emission changes in terms of *year-to-year relative values*, i.e., relative differences between 2019 and 2020 quantities.

#### 4.3.4 Maps of averaged NO<sub>2</sub>

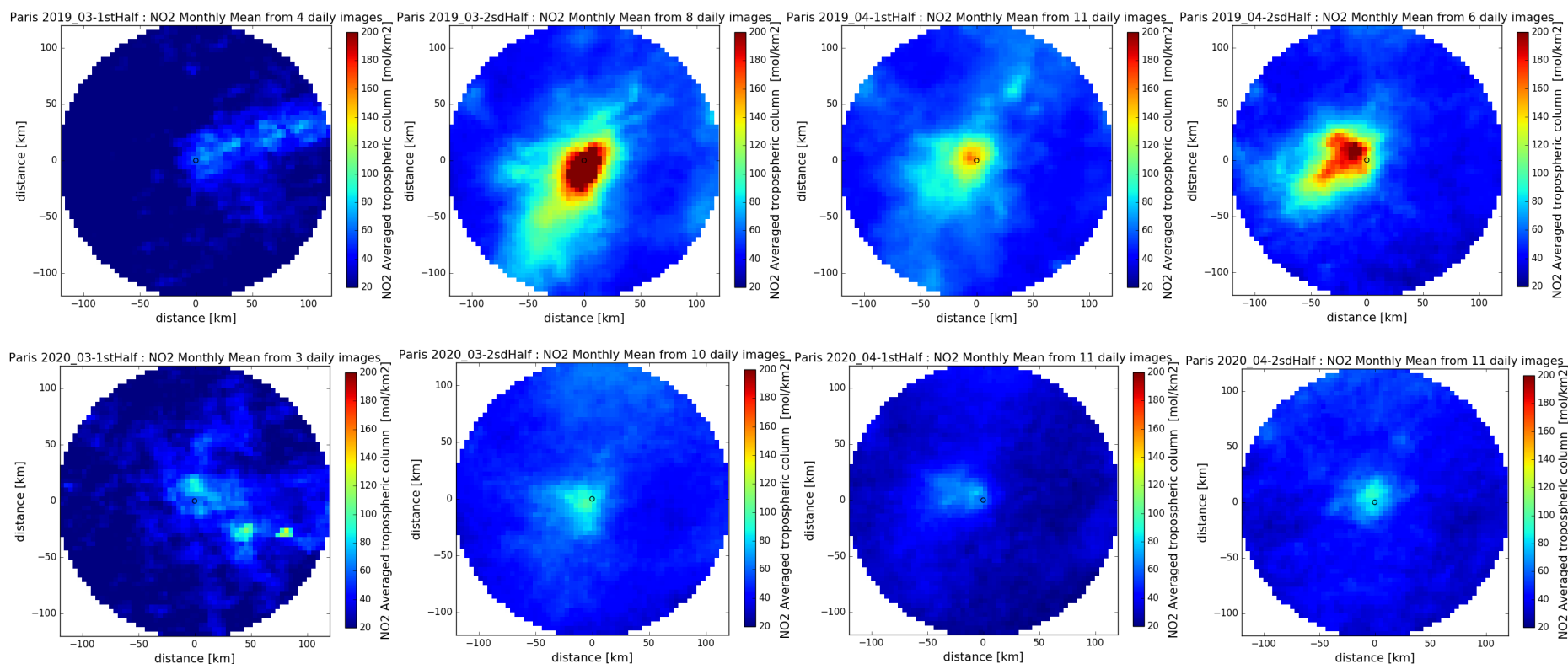
Figure 10 compares monthly mean maps for the two periods 2019 and 2020. These results can be compared to the monthly mean plume masses (Figure 7, top), albeit with some caution: the weighting applied to the daily plume masses used in the averaging process for obtaining monthly means of Figure 7 is not applied to produce Figure 10 averages. Apparently contradictory results arise when monthly means are obtained from a small number of valid days. This is the case for January and February: the monthly plume masses in Figure 7 indicate values in 2020 equivalent to those in 2019 (slightly larger in January, slightly lower in February). However, the monthly mean maps of NO<sub>2</sub> tropospheric columns suggest significantly lower values in 2020 than in 2019 for the two months. For January, this apparent contradiction is meaningless, because of the small number of daily values (only 4) available to derive the January 2019 value, preventing any robust estimate of the corresponding averages. The two average quantities in January 2019 (monthly plume mass and monthly map of NO<sub>2</sub> tropospheric column) are two interpretations of the same uncertain situation. For February, the 2019 quantities appear robust (based on 17 daily values), and the apparent difference in the analysis of 2020 (plume mass or map of NO<sub>2</sub> columns) reflects the impact of weighting daily values with uncertainties estimated with our specific approach. Roughly speaking, our uncertainty calculation estimates a larger uncertainty when the NO<sub>2</sub> content spreads out all over the domain (when the wind is strong) and a smaller uncertainty when the NO<sub>2</sub> content is concentrated in the centre of the domain (when the wind is weak). This basically leads to put more weight on the larger values in the weighted average of plume masses, acknowledging the risk that lower but non-negligible values close to the domain boundaries can be associated to some missing pollutant outside the considered domain.



**Figure 9 : Daily maps of NO<sub>2</sub> tropospheric column (displayed in logarithmic colour scale) over Paris region, for 2 days: 26<sup>th</sup> (top left and bottom left) and 27<sup>th</sup> (top right and bottom right) March 2020. The impact of NO<sub>2</sub> high concentrations coming from external sources outside the domain is clearly visible for the 2 days in the North of the domain of the nominal disk (top left and top right). The reduced disk (bottom left and bottom right, 75% of the diameter of the nominal disk) used to estimate the uncertainty value associated with the total NO<sub>2</sub> plume mass is less impacted, suggesting that the uncertainty provides a reasonable estimator to deal with this perturbation effect.**



**Figure 10 : Monthly mean maps of NO<sub>2</sub> tropospheric column over Paris region re shown from left to right for January, February, March and April, respectively: Maps for the year 2019 are shown on the top. Maps for the year 2020 are shown at the bottom. The number of valid days used for the monthly mean is indicated in the title of each map.**



**Figure 11 : March fortnightly (from 1 to 15 and from 16 to 31) maps of NO<sub>2</sub> tropospheric column over Paris region: Maps for March 2019 are shown on the top. Maps for March 2020 are shown at the bottom. The number of valid days used for the fortnightly mean is indicated in the title of each map. Left column shows the 2d half March maps in linear colour scale.**

This illustrates the difficulty of comparing and interpreting maps derived from a limited number of daily values (January) and the necessity to carefully consider quantities with large error bars (February). The comparisons for March and April averages are more meaningful. Monthly maps exhibit robust estimates, consistent with the corresponding plume mass time series. They show a significant reduction of the NO<sub>2</sub> pollution level, clearly suggesting an effective decrease of the Paris NO<sub>2</sub> emissions in March and April 2020 as compared to 2019. Fortnightly average maps for March and April 2019 and 2020 shown in Figure 11 confirm this trend with robust estimates from the second half of March. The interpretation of the relative impact of Paris emission is not completely robust from January to mid-March because of the lack of exploitable data over the considered area. From mid-March to end of April 2019 and 2020, the amount of valid data becomes sufficient, and the plume masses and NO<sub>2</sub> column maps are consistent, including for fortnights.

#### 4.3.5 Quantitative estimate of NO<sub>2</sub> reduction rates

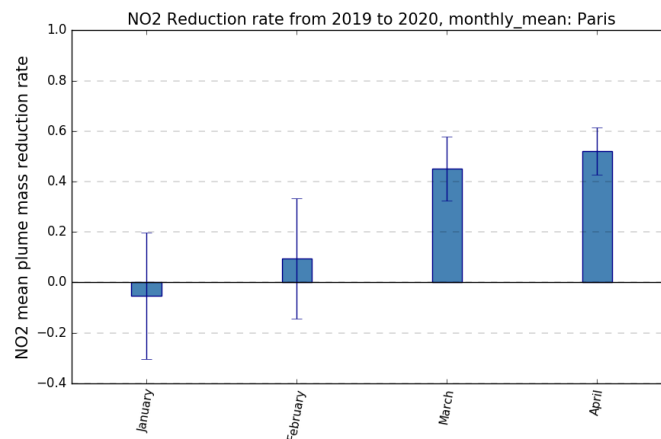
A first quantitative estimate of the NO<sub>2</sub> reduction linked to human activities can then be computed from the monthly and fortnightly values of NO<sub>2</sub> plume mass and their respective uncertainties. As discussed, the relevant quantity for this estimate is the relative variation from 2019 to 2020 for a given monthly or fortnightly averaging period. The corresponding ratio (or decreasing rate) is less sensitive to meteorological effects and is not affected by variations of SZA impacting the NO<sub>2</sub> content measurement from space. Thus, we have computed the reduction rate  $R = \Delta PM_{NO_2} / PM_{NO_2}$  where  $PM_{NO_2}$  is the plume mass of NO<sub>2</sub> for a given averaging period, and  $\Delta PM_{NO_2}$  is the difference between 2019 and 2020.

$$\Delta PM_{NO_2} = PM_{NO_2}(2019) - PM_{NO_2}(2020)$$

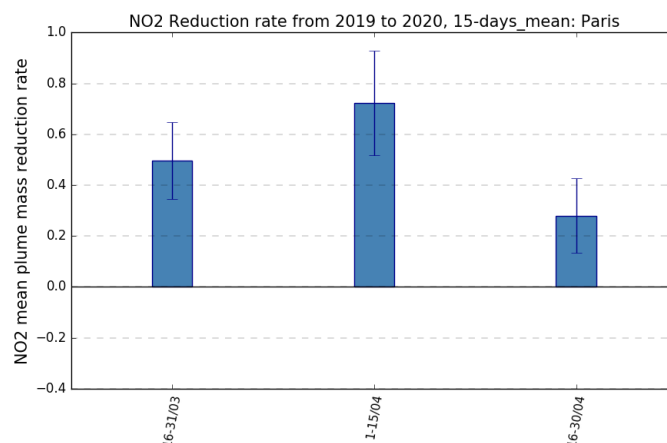
$$PM_{NO_2} = PM_{NO_2}(2019)$$

The uncertainty  $\delta R$  on  $R$  is derived from the uncertainties  $\delta PM_{NO_2}$  estimated on the plume masses.

$$\delta R = (\delta PM_{NO_2}(2019)^2 + \delta PM_{NO_2}(2020)^2)^{1/2} / PM_{NO_2}(2019)$$



**Figure 12 : Reduction rates of the NO<sub>2</sub> plume mass monthly mean between 2019 and 2020 over Paris city. A positive value indicates a reduction from 2019 to 2020. The corresponding uncertainty bars are also provided.**



**Figure 13 : Reduction rates of the NO<sub>2</sub> plume mass fortnightly mean between 2019 and 2020 over Paris city, from mid-March to end of April, and corresponding uncertainty.**

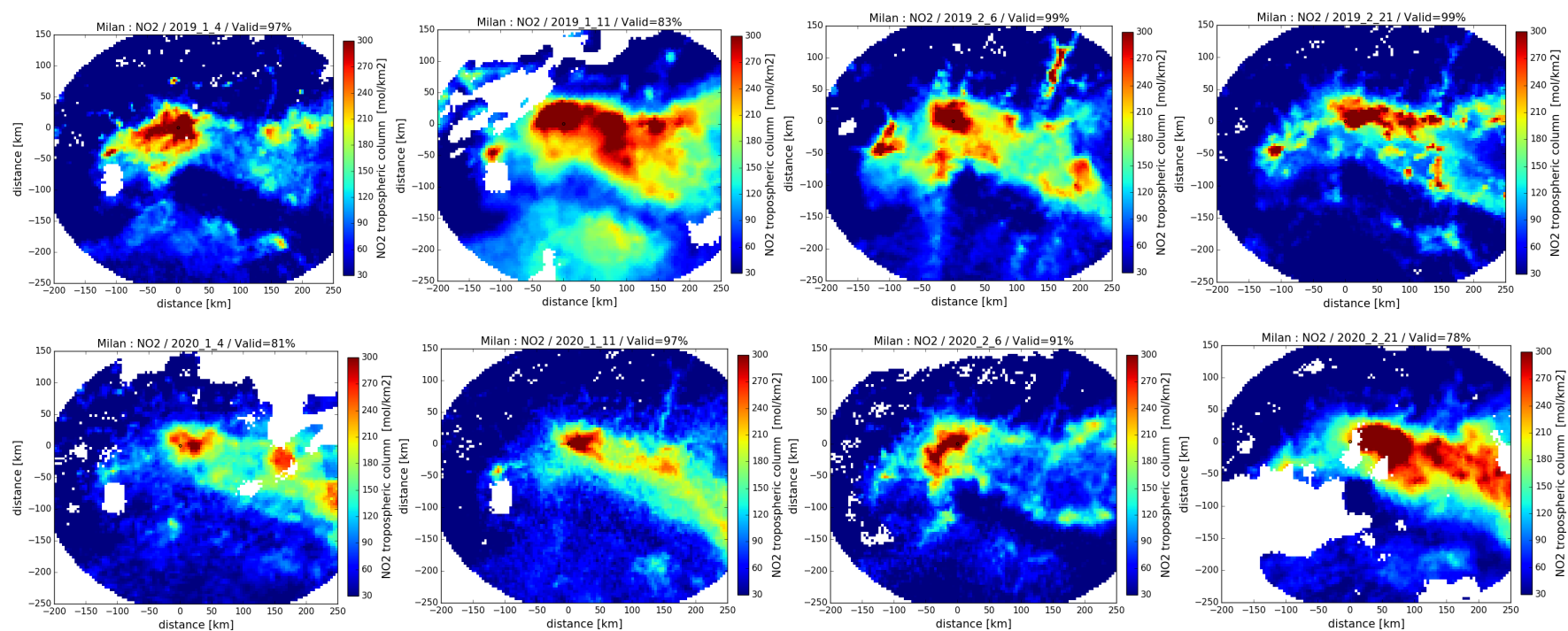
Reduction rates and the corresponding uncertainties are computed from plume mass monthly averages (Figure 12). Monthly reduction rates for January and February are non-significant. However, a reduction of the 2020 NO<sub>2</sub> pollution level is clearly observed in March (45%, +/- 13%) and April (52%, +/- 9%), as compared to 2019. The reduction rates computed for plume mass fortnightly averages, from Mid-March to end of April (Figure 13), indicate a maximum reduction in the period from 1<sup>st</sup> to 15<sup>th</sup> of April, reaching 72% +/- 20%.

#### **4.4 Results for Milan**

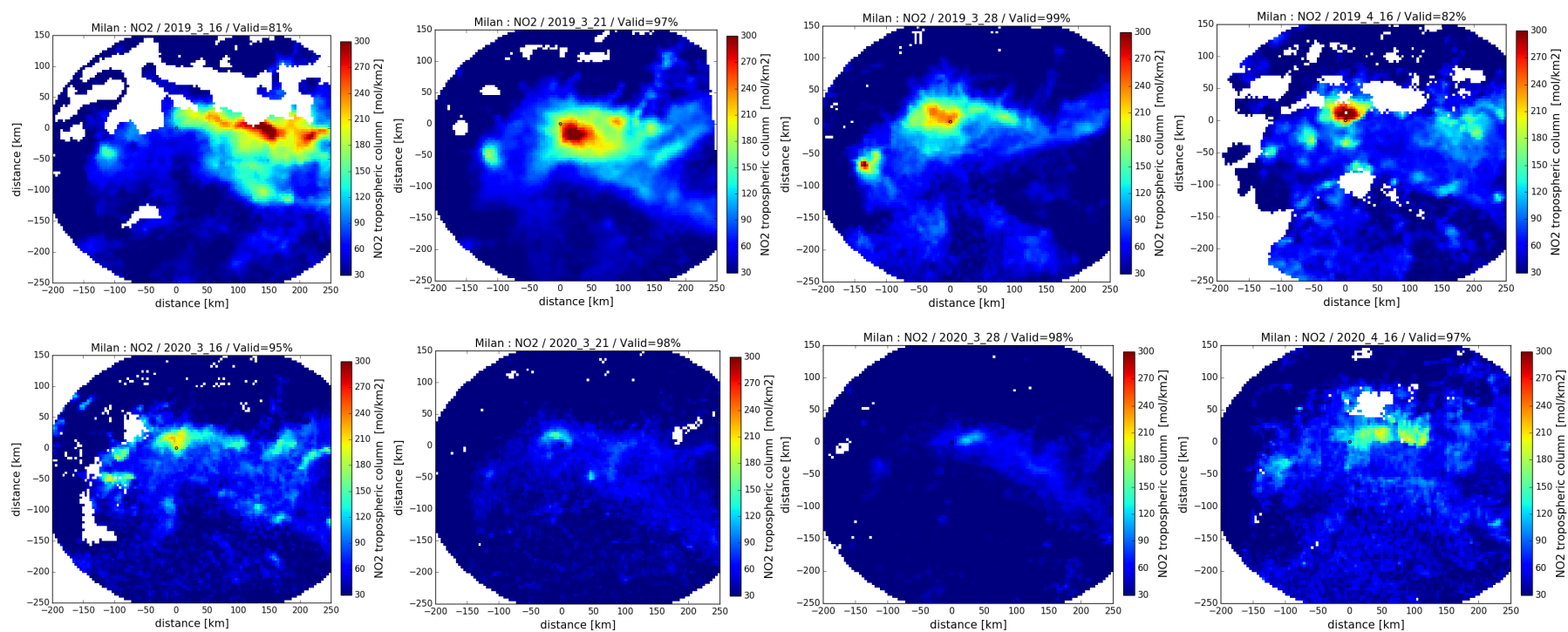
For Milan, it has been necessary to enlarge the area of analysis in order to cover all the Po valley where the NO<sub>2</sub> plume generally spreads. In that case the number of selected S5P data is higher than for Paris. The percentage of valid days over 4 months (120 days) is 55-62% for the Milan region. Most of the periods are well represented, except the beginning of March 2020 and the beginning of April 2019, which suffer from a lack of valid days. Apart these two periods, the computation of reliable averages and the analysis of the impact of human activities is robust in this region.

Typical daily maps are presented in Figure 14 comparing 2019 and 2020 NO<sub>2</sub> tropospheric columns for the January-February period. The day-to-day variability is of the same order of magnitude as the inter annual day-to-day differences.

Figure 15 presents similar comparisons for the mid-March to Mid-April period, which in this case exhibit clear systematic differences in daily maps between 2019 and 2020. Again, the analysis of averaged quantities is required to consolidate and quantify the reductions. However, the daily maps already provide an evidence of the impact of emission reduction. This is also due to the geographical situation of the Milan region, which maintains most of the pollution within the Po Valley (limited loss of NO<sub>2</sub> mass outside the domain) and which seems relatively well preserved from NO<sub>2</sub> atmospheric incursions from external sources outside the domain.



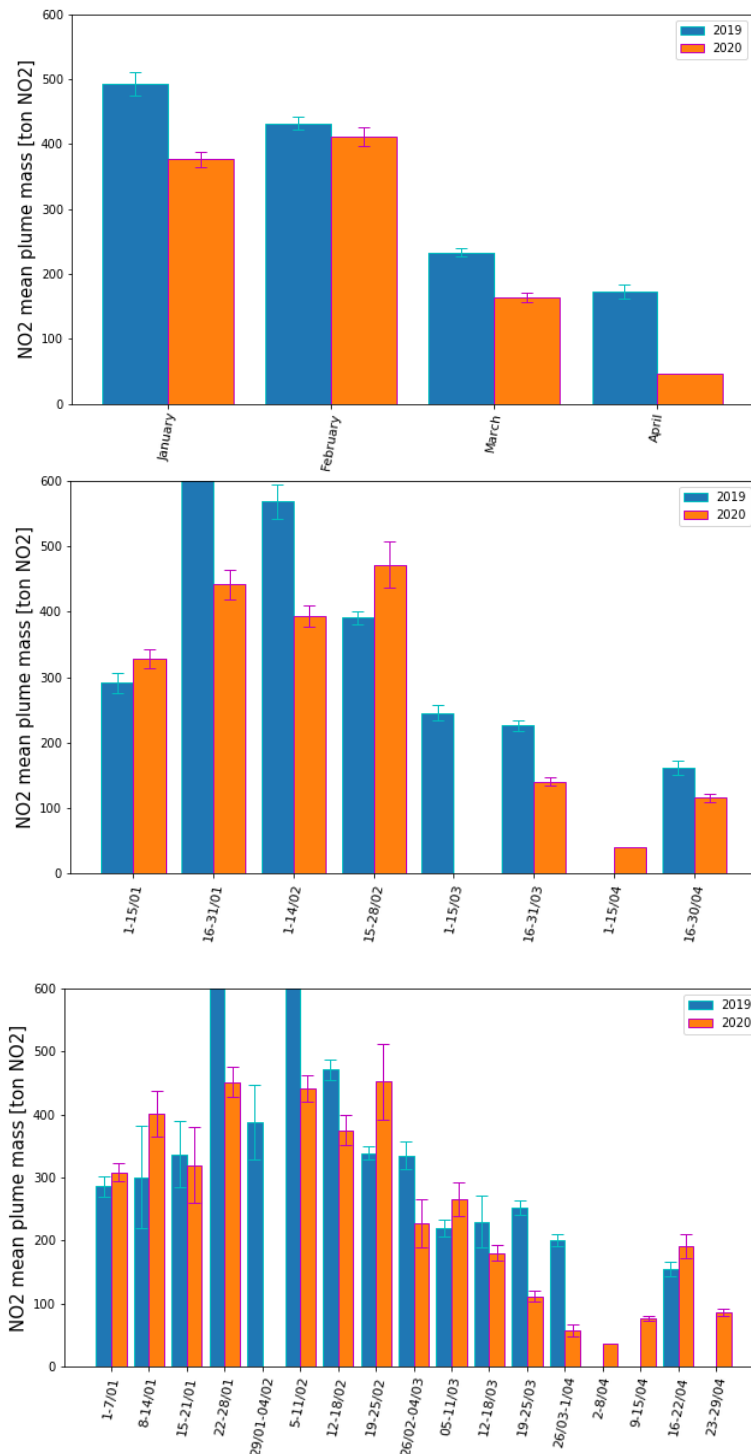
**Figure 14 : Daily maps of NO<sub>2</sub> tropospheric column over Milan region, for selected days in January/February. Top: 4 days in 2019. Bottom: same 4 days in 2020.**



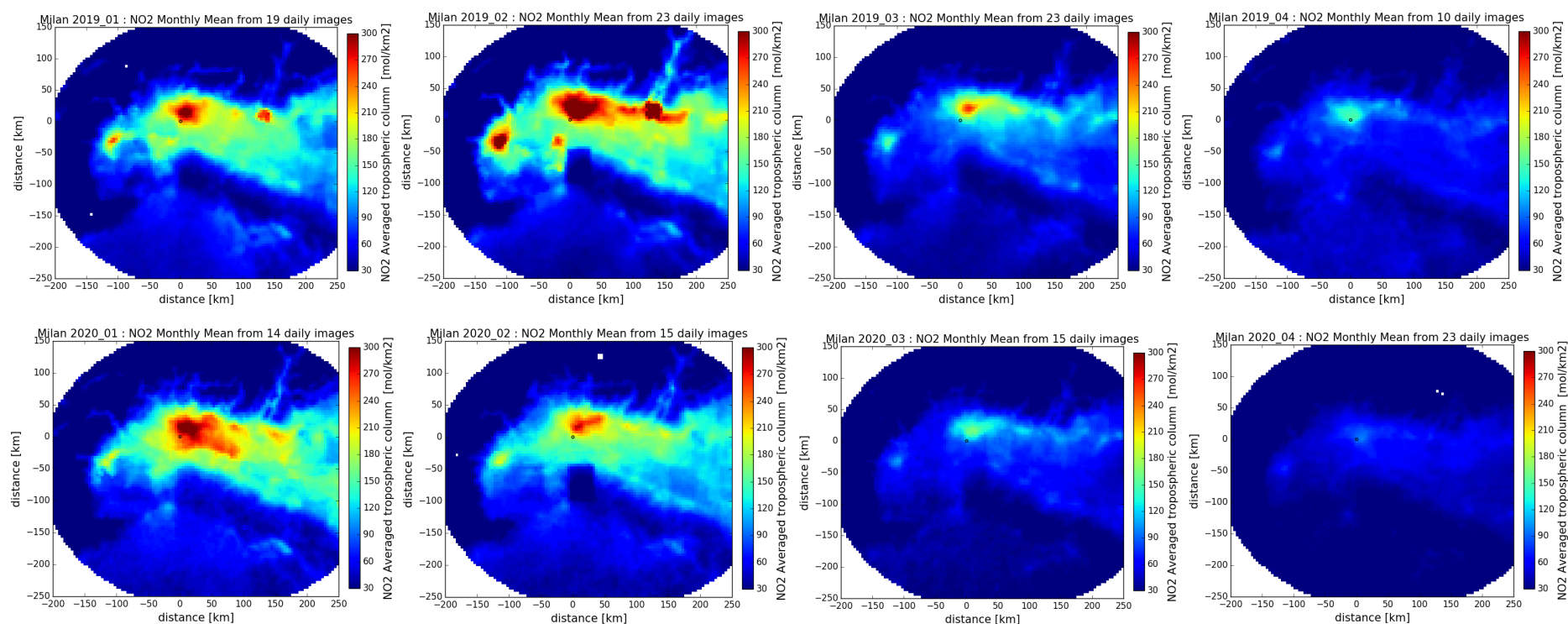
**Figure 15 : Daily maps of NO<sub>2</sub> tropospheric column over Milan region, for selected days of the March and April. Top: 4 days in 2019. Bottom: same 4 days in 2020.**



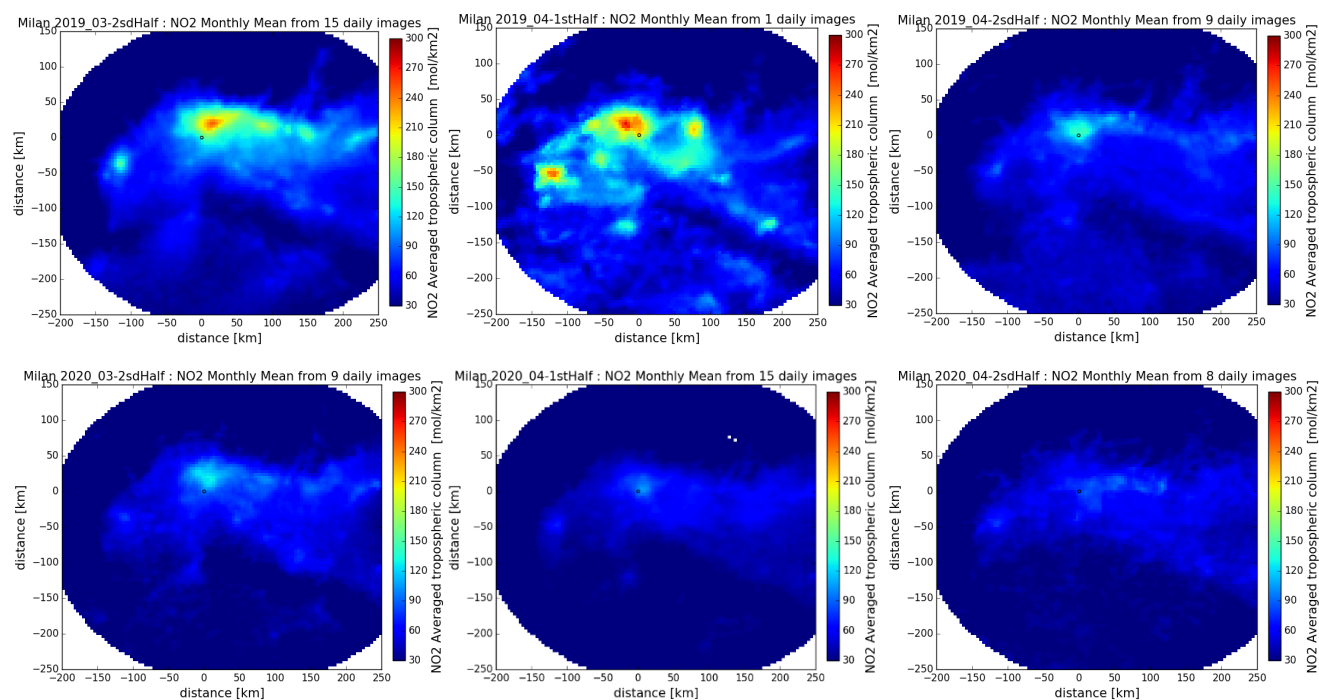
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.



**Figure 16 : Temporal variations of total NO<sub>2</sub> plume mass over Milan region for the January-April period of 2019 (blue) and 2020 (orange). Total NO<sub>2</sub> plume mass is averaged over 1 month (top), 15-day (middle) and 7-day (bottom) temporal windows. Missing values indicate that less than 2 daily values (or less than 1 daily value for 7-day periods) are available for the considered temporal window. Uncertainty bars on the averaged values are derived from daily uncertainty estimates.**



**Figure 17: Monthly mean maps of  $\text{NO}_2$  tropospheric column over Milan region, for January (left) February (middle) and March (right): Maps for the year 2019 are shown on the top. Maps for the year 2020 are shown at the bottom. The number of valid days used for the monthly mean is indicated in the title of each map**



**Figure 18: March fortnightly (from 16 to 31) and April fortnightly (from 1 to 15 and 16 to 31) maps of NO<sub>2</sub> tropospheric column over Milan region: Maps for 2019 are shown on the top. Maps for 2020 are shown at the bottom. The number of valid days used for the fortnightly mean is indicated in the title of each map.**

Figure 16 compares monthly, fortnightly (15-days) and weekly (7-days) averages of the total NO<sub>2</sub> plume mass time series from January to April 2019 and 2020. Despite the fact that each averaging period is largely sampled, some specific periods are impacted by the lack of data: this is the case for the period 22-28 January 2020 (only 2 valid days). We also observe large differences between 2019 and 2020 for the end of January and the beginning of March, and this should be further investigated. There is also a lack of representativeness for the 01-15 March 2020 and for the 01-15 April 2019 temporal windows. This is preventing a reliable relative year-to-year comparison for these two periods and could introduce some bias in the monthly averages for March 2020 and April 2019. However, the estimates for March/April 2019 and March/April 2020 appear to be robust at the monthly scale indicating a reduction of the total NO<sub>2</sub> plume mass over the Po region from 2019 to 2020, which can be interpreted as a reduced pollution level over Milan and the regional area. This reduction is well marked from mid- to end of March and from mid- to end of April. The average data for the three last weeks of March suggests an increasing reduction with time, from about 20% (Figure 16 bottom, 12-18/03 values) to about 65% (Figure 16 bottom, 19-25/03 and 26-31/03 values).

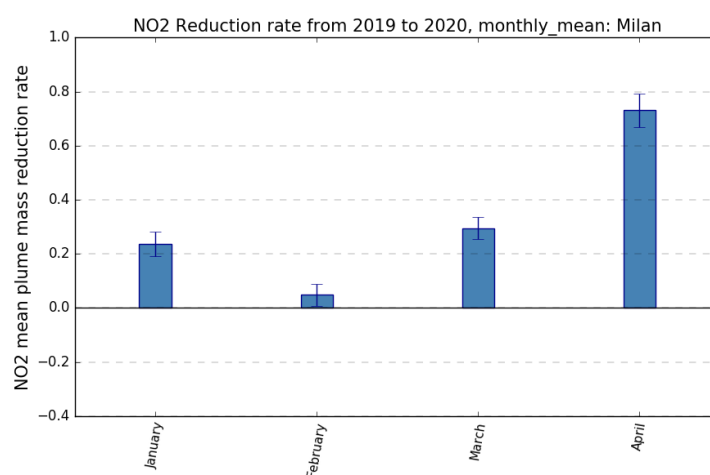
Figure 17 compares monthly maps for the two periods 2019 and 2020, which illustrates the robust analysis resulting from a large number of daily values at monthly scale. Fortnightly average maps for mid-March to end of April 2019 and 2020 (Figure 18) are based on a reliable statistics for the 2019/2020 comparison over these fortnightly periods except for first half of April : only 1 daily map is available in 2019 for this specific period and this prevents a firm interpretation of the corresponding reduction rate. However, the other fortnightly maps confirm the reduction.

The reduction rates and corresponding uncertainties are computed from plume mass monthly averages (Figure 19). The monthly reduction rate for February is non-significant, but for January it is representing a reduction from 2019 to 2020, which should be further analysed and interpreted for this specific month. A reduction of the 2020 NO<sub>2</sub> pollution level is confirmed for March (30%, +/- 4%) and April 73%, +/- 6%), compared to 2019. However, this high value of the monthly reduction estimate for April shall be taken with caution. Indeed, the reduction rates computed for plume mass fortnightly averages, for the second half of March and the second half of April are presented in Figure 20. The value for mid- to end-March (38%, +/- 4%) is basically consistent with the March monthly reduction rate, but the reduction rate for mid- to end-April (28%, +/- 8%) is noticeably lower than for the April monthly value. This means that the high value for the monthly reduction rate is strongly impacted by small plume mass values in the first half of April 2020, not balanced by corresponding 2019 values (missing data for first half of April 2019). This could introduce a bias if the average plume mass in the first half of April 2019 had been significantly different from the average plume mass in second half of April 2019. Such verification is requiring a complementary analysis with independent data.

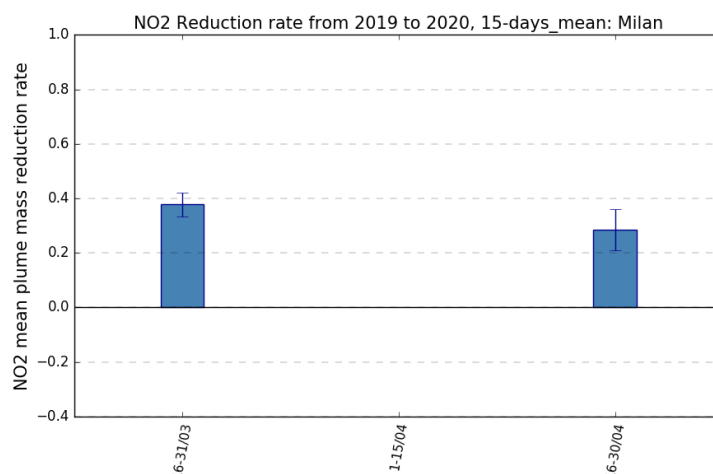
First, a similar analysis from Bauwens et al. (2020) reported a reduction rate of the NO<sub>2</sub> pollution level over Milan for the 22 February to early March 2020, as compared to the same period in 2019. The reductions are 46 % from TROPOMI data and 11% from OMI data. The corresponding estimates derived in the present analysis from Figure 16 (bottom, 26/02-04/03 values) range between 47% and 15%.

Second, we mention here the results of CAMS European air quality analyses over the North of Italy and Milan, as discussed in ECMWF/CAMS 2020-b. This electronic newsletter published

17<sup>th</sup> March 2020 on the ECMWF Copernicus Atmosphere Monitoring Service web site (<https://atmosphere.copernicus.eu/air-quality-information-confirms-reduced-activity-levels-due-lockdown-italy>) presents maps of NO<sub>2</sub> surface concentrations over northern Italy for 2 specific days, 31 January and 15 March 2020. No direct comparison with daily maps derived from S5P (as shown for example in Figure 15) is possible, since our filtering process removed the data of S5P for the corresponding days, but CAMS maps exhibit very similar patterns when compared to those derived from S5P NO<sub>2</sub> VCD values. CAMS is also providing time series of the NO<sub>2</sub> surface concentrations, based on hourly values for the grid point covering the city of Milan (from the CAMS European air quality analyses product), and presented as a 1-week running average. Again in that case, a direct comparison with weekly averages of NO<sub>2</sub> plume mass presented in Figure 16 is not possible, mainly because the spatial scale of the two quantities significantly differs (the processing used in our analysis defines a specific horizontal integration over the Po Valley removing a background threshold). The CAMS analyses of this temporal series suggest a gradual reduction trend of about 10% per week over the mid-February to mid-March period, associated to the grid point of the CAMS European analysis covering Milan (typically 10 km x 10 km). On the other hand, our results suggest a reduction trend of about 15% per week over the same period, associated to the typical 500 km diameter area of the Po Valley. Nevertheless, other comparisons would be necessary in order to conclude on the consistency of our results with CAMS results.



**Figure 19 : Reduction rates of the NO<sub>2</sub> plume mass monthly mean between 2019 and 2020 over Milan region. A positive value indicates a reduction from 2019 to 2020. The corresponding uncertainty bars are also provided.**



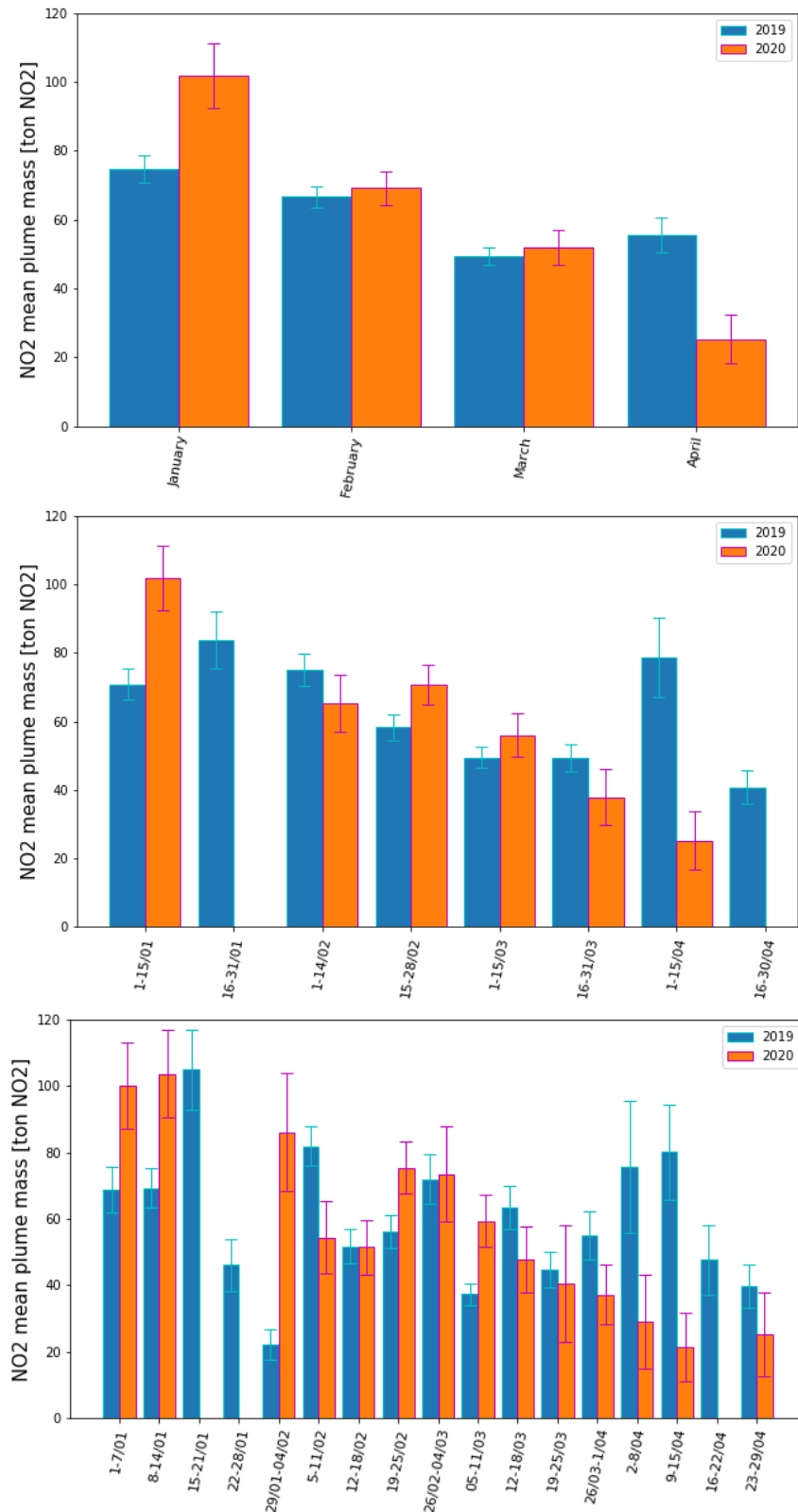
**Figure 20 : Reduction rates of the NO<sub>2</sub> plume mass fortnightly mean between 2019 and 2020 over Milan region, from mid-March to end of April, and corresponding uncertainty. The missing value for first half of April is due to insufficient number of valid data for first half of April 2019.**

## 4.5 Results for Madrid

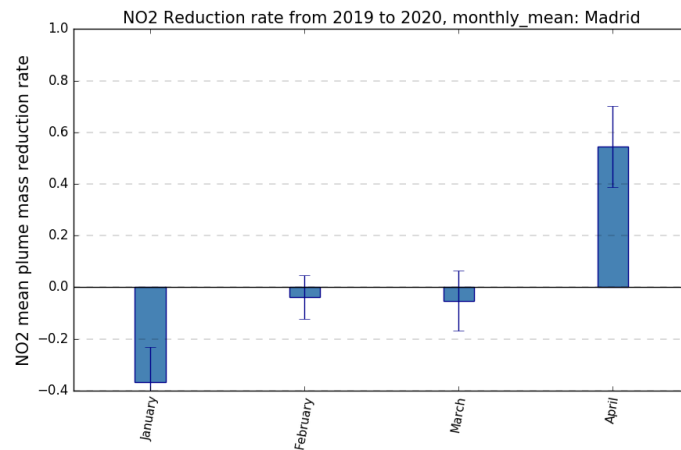
For Madrid, the number of selected S5P data is close to 76% of valid days over 4 months in 2019, but only 40% in 2020. The valid data in 2020 is evenly populated within the different periods, except for the second half of January and second half of April periods that suffer from a lack of data. The computation of reliable averages and the analysis of the NO<sub>2</sub> pollution levels trends is thus depending on the considered periods.

Figure 21 compares the monthly, fortnightly (15-days) and weekly (7-days) averages of total NO<sub>2</sub> plume mass time series from January to April 2019 and 2020. This data suggests a reduction of the total NO<sub>2</sub> plume mass starting from mid-March. However, this reduction is not reflected in the March monthly values, because the decrease in the second half of March is marginally significant (see Figure 21 middle panel and the large error bar for the 2020 value for this period) and may be balanced by the values in the first half of March. The large error bar on the 16-31/03 average estimate for 2020 is due to a small number of valid data. The reduction from 2019 to 2020 is significant for April monthly values (Figure 21 top panel), with a large reduction observed for the first half of April (Figure 21 middle panel). The detailed weekly values (Figure 21 bottom panel) even suggest a significant reduction at the weekly scale for the two first weeks of April (2-8 April and 9-15 April) and the last week of April (23-29 April).

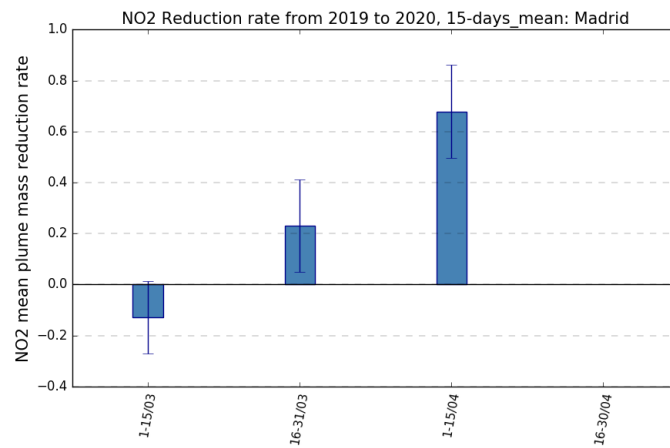
Reduction rates and corresponding uncertainties are computed from plume mass monthly averages (Figure 22), fortnightly averages (Figure 23) and weekly averages (Figure 24) for periods populated by a significant amount of daily data in 2019 and 2020.



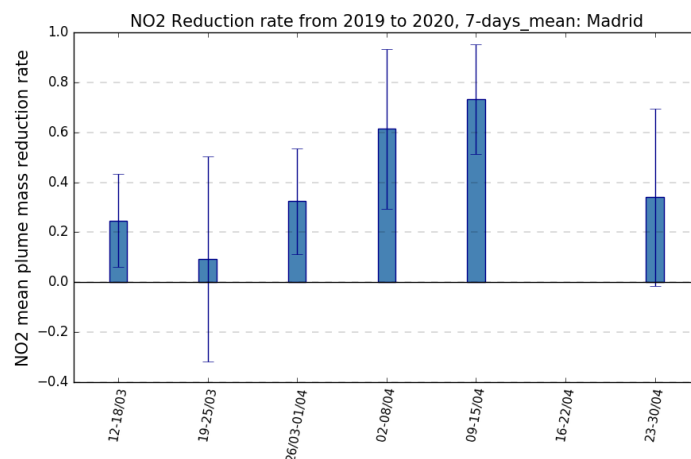
**Figure 21 : Temporal variations of total NO<sub>2</sub> plume mass over Madrid for the January-April period of 2019 (blue) and 2020 (orange). Total NO<sub>2</sub> plume mass is averaged over 1 month (top), 15-day (middle) and 7-day (bottom) temporal windows. Missing values indicate that less than 2 daily values (or less than 1 daily value for 7-day periods) are available for the considered temporal window. Uncertainty bars on the averaged values are derived from daily uncertainty estimates.**



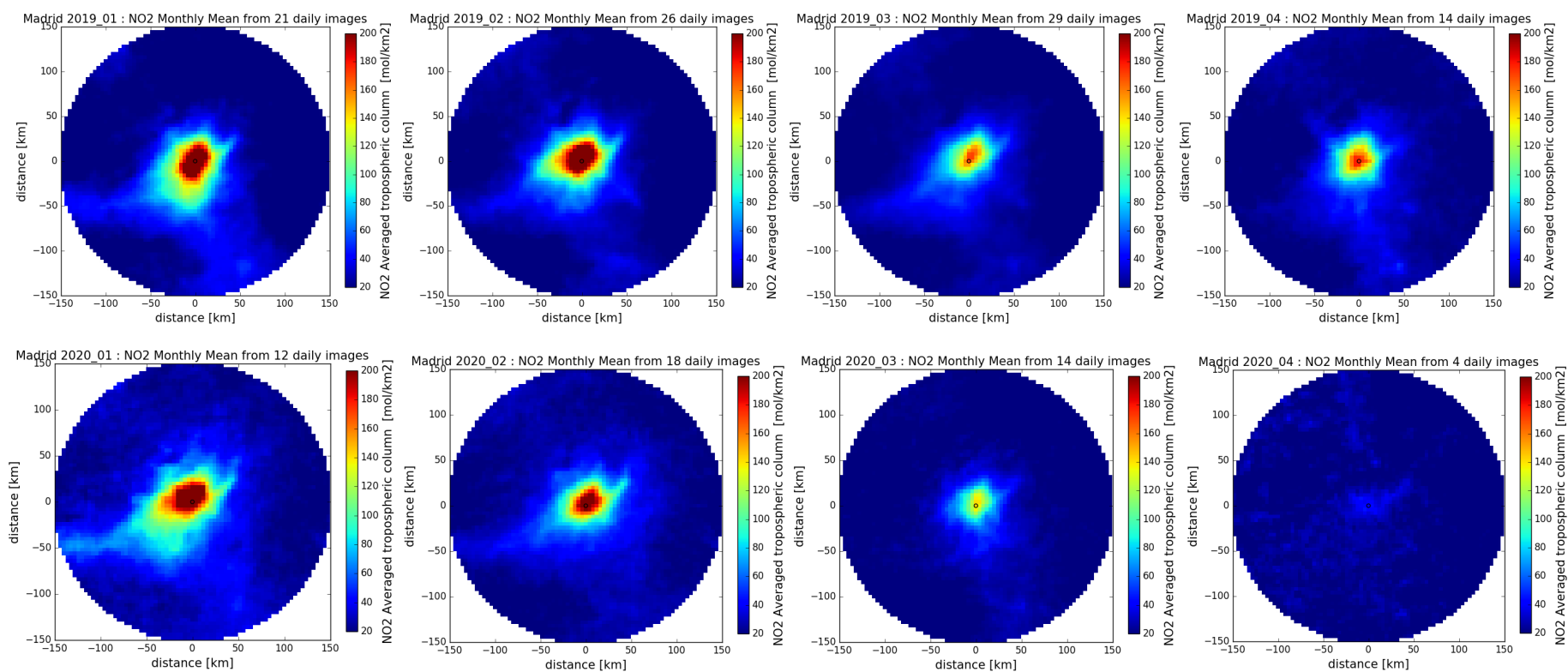
**Figure 22 : Reduction rates of the NO<sub>2</sub> plume mass monthly mean between 2019 and 2020 over Madrid. A positive value indicates a reduction from 2019 to 2020. The corresponding uncertainty bars are also provided**



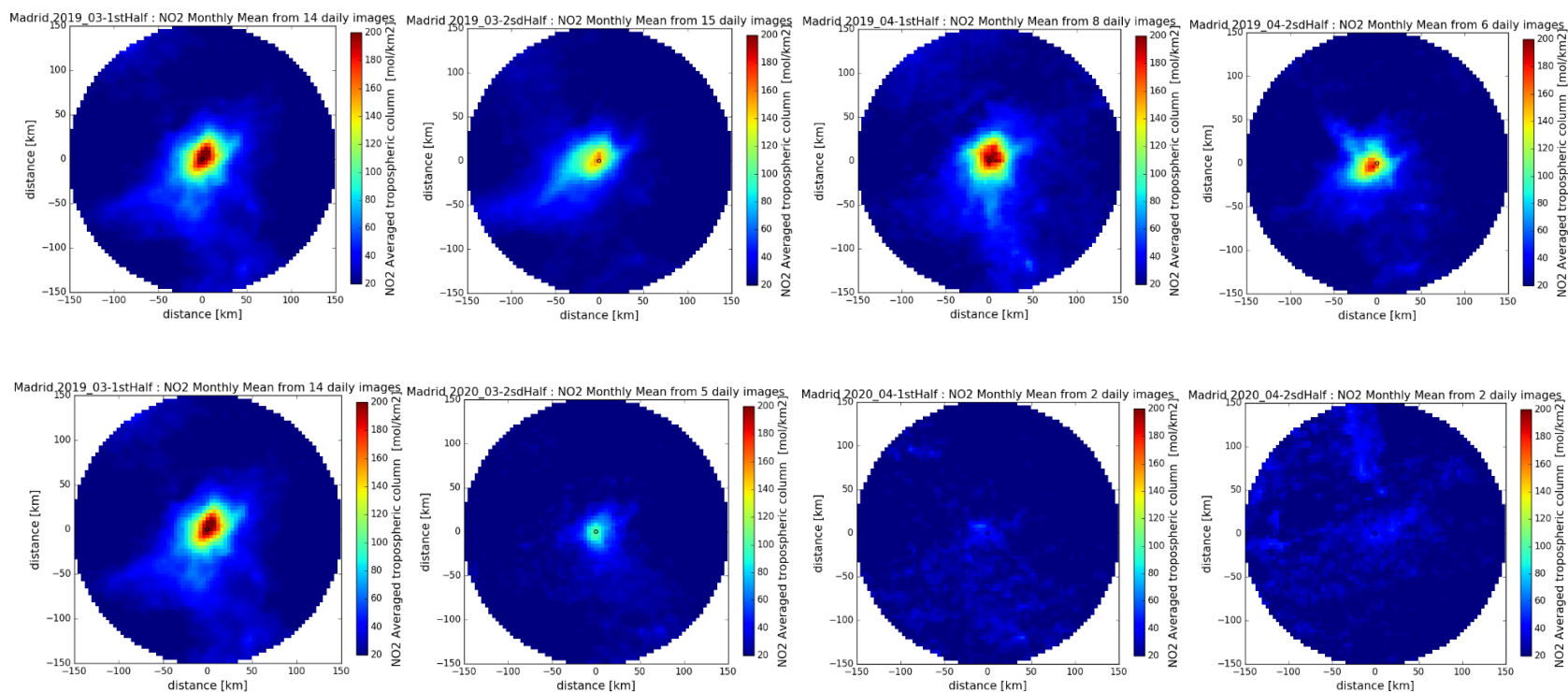
**Figure 23 : Reduction rates of the NO<sub>2</sub> plume mass fortnightly mean between 2019 and 2020 over Madrid, from mid-March to end of April, and corresponding uncertainty. The missing value for second half of April is due to insufficient number of valid data for second half of April 2020.**



**Figure 24 : Reduction rates of the NO<sub>2</sub> plume mass weekly mean between 2019 and 2020 over Madrid, from mid-March to end of April, and corresponding uncertainty. The missing value for week 16-22 April is due to insufficient number of valid data in 2020 for this week.**



**Figure 25: Monthly mean maps of NO<sub>2</sub> tropospheric column over Madrid, for January (left) February (middle) and March (right): Maps for the year 2019 are shown on the top. Maps for the year 2020 are shown at the bottom. The number of valid days used for the monthly mean is indicated in the title of each map.**



**Figure 26: March fortnightly (from 1 to 15 and from 16 to 31) maps of NO<sub>2</sub> tropospheric column (log colour scale) over Madrid: Maps for March 2019 are shown on the top. Maps for March 2020 are shown at the bottom. The number of valid days used for the fortnightly mean is indicated in the title of each map. Left column shows the 2d half March maps in linear colour scale.**

Monthly reduction rates (Figure 22) for February and March are non-significant, while for January an increase from 2019 to 2020 is observed, but is not further discussed here. A reduction of the 2020 NO<sub>2</sub> pollution level (as compared to 2019) in monthly mean is given only for April (54%, +/- 16%).

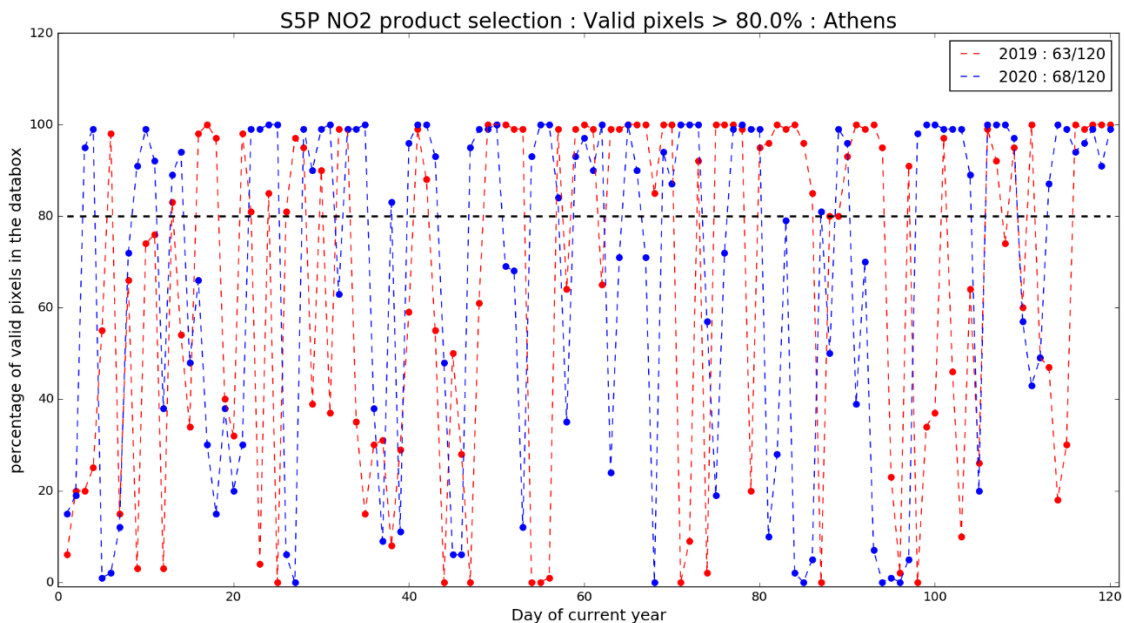
Fortnightly reduction rates (Figure 23) for the 1-15 and 16-31 March periods are basically non-significant and compensate each other. On the other hand, the reduction rate for 1-15 April is indicating a significant decrease (68%, +/- 18%) for this period and remains consistent with the April monthly rate. The rate for second half of April is not computed due to an insufficient number of valid data for second half of April 2020.

Finally, we show weekly reduction rates from mid-March to end of April (Figure 24) confirming a decrease of the NO<sub>2</sub> level in 2020, compared to 2019, starting from mid-March, with in particular, noticeable larger reduction rates in the three first weeks of April, increasing from 32% +/- 21% (week from 26 March to 01 April) to 73% +/- 22% (week from 09 to 15 April). The value for the last two weeks of April are missing or highly uncertain.

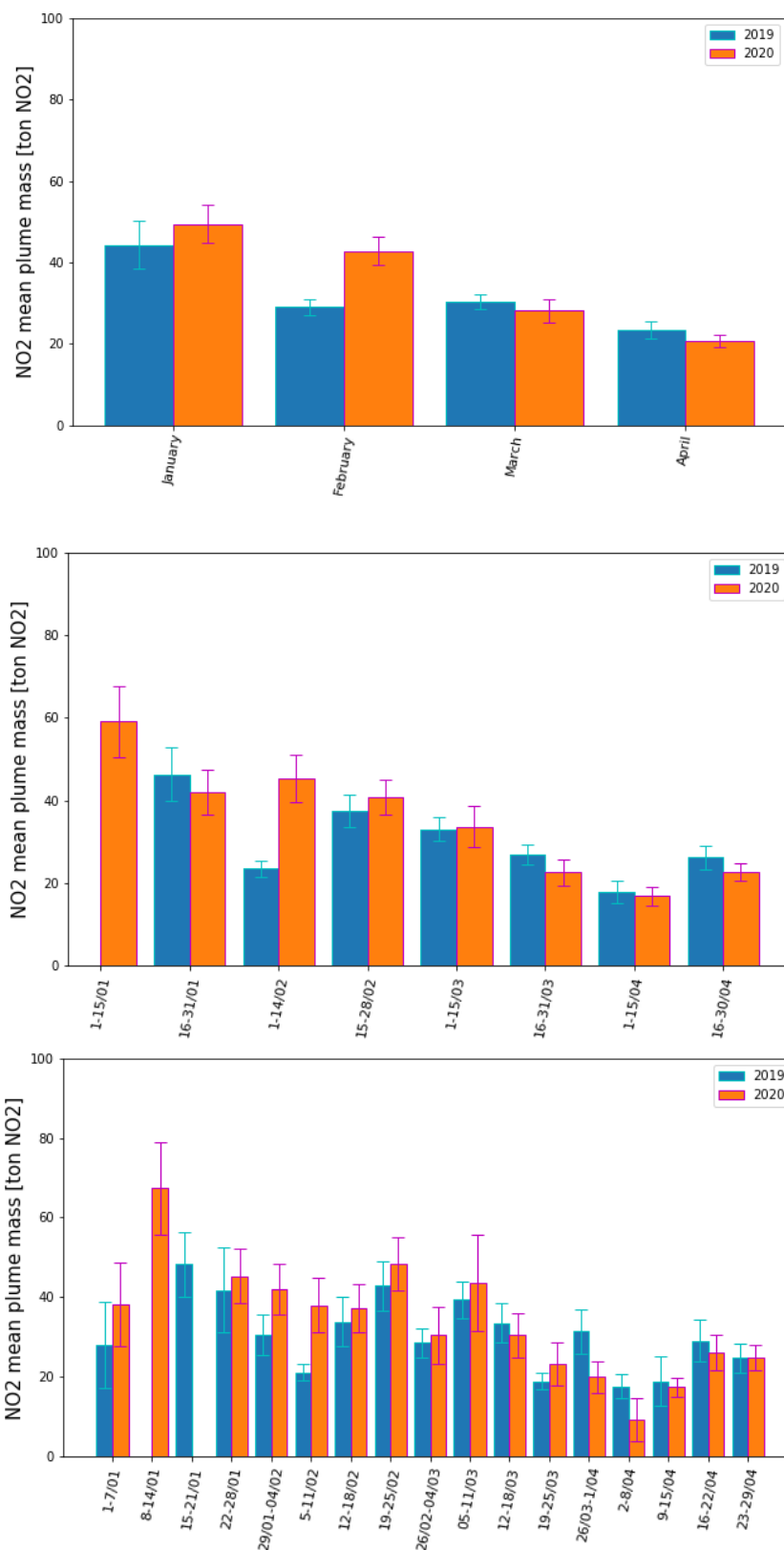
Figure 25 compares monthly mean maps for the two periods 2019 and 2020, whereas fortnightly average maps for March 2019 and 2020 are shown in Figure 26. The monthly maps already suggest a reduction in March and clearly show the April decrease of pollution. The fortnightly maps confirm the reduction of the NO<sub>2</sub> level starting in 2<sup>nd</sup> part of March.

## 4.6 Results for Athens

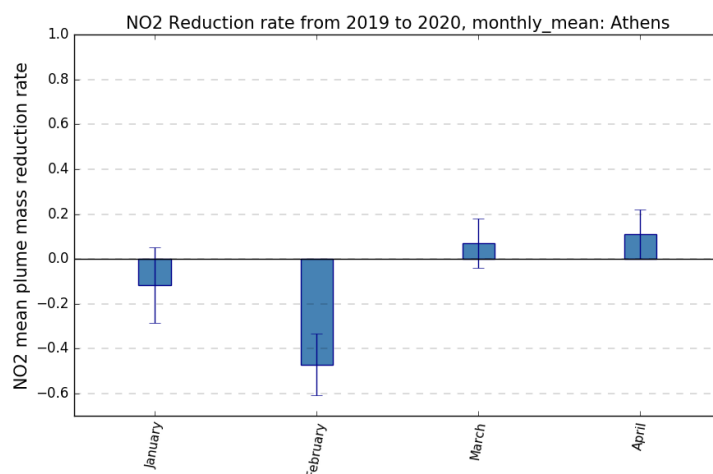
The statistics on filtered days are presented in Figure 27. For Athens, the number of selected S5P data is between 52% (2019) and 57% (2020) of valid days over 4 months, with most of the monthly and fortnightly periods populated. The computation of reliable averages and the analysis of the NO<sub>2</sub> pollution level trends is considered rather robust in this case.



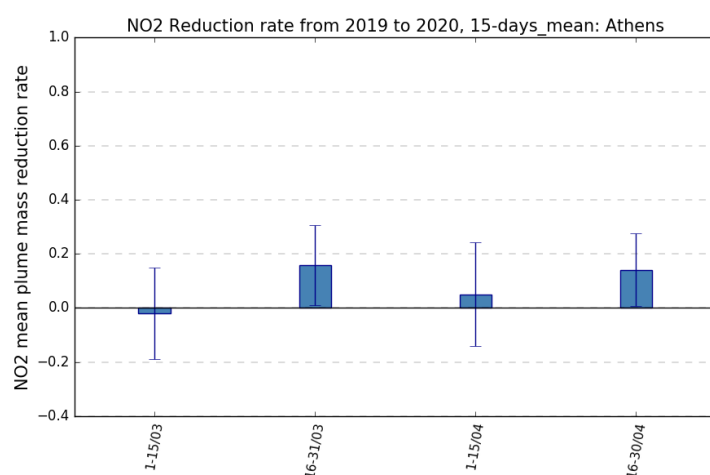
**Figure 27 : Daily selection criteria for Athens for 2019 (red) and 2020 (blue) periods (January-March) : percentage of valid pixels in the disk ( the day is selected if the percentage of valid pixels is larger than 80%), number of selected days over total number of days are indicated in the legend.**



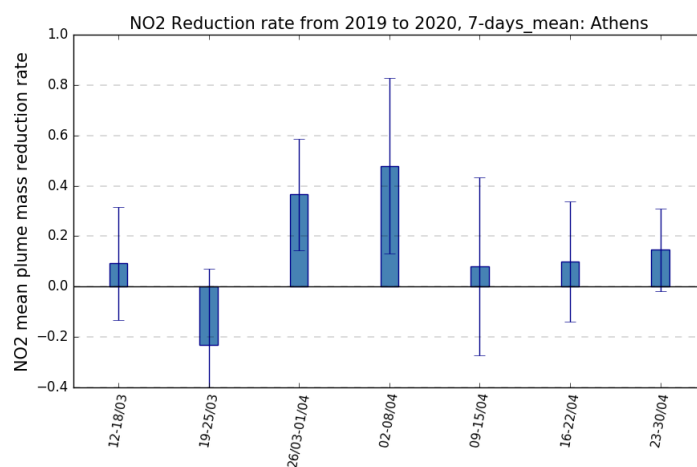
**Figure 28: Temporal variations of total NO<sub>2</sub> plume mass over Athens for the January-March period of 2019 (blue) and 2020 (orange). Total NO<sub>2</sub> plume mass is averaged over 15-day temporal windows (top) and 7-day temporal windows (bottom). Missing values indicate that less than 2 daily values are available for the considered temporal window. Uncertainty bars on the averaged values are derived from daily uncertainty estimates.**



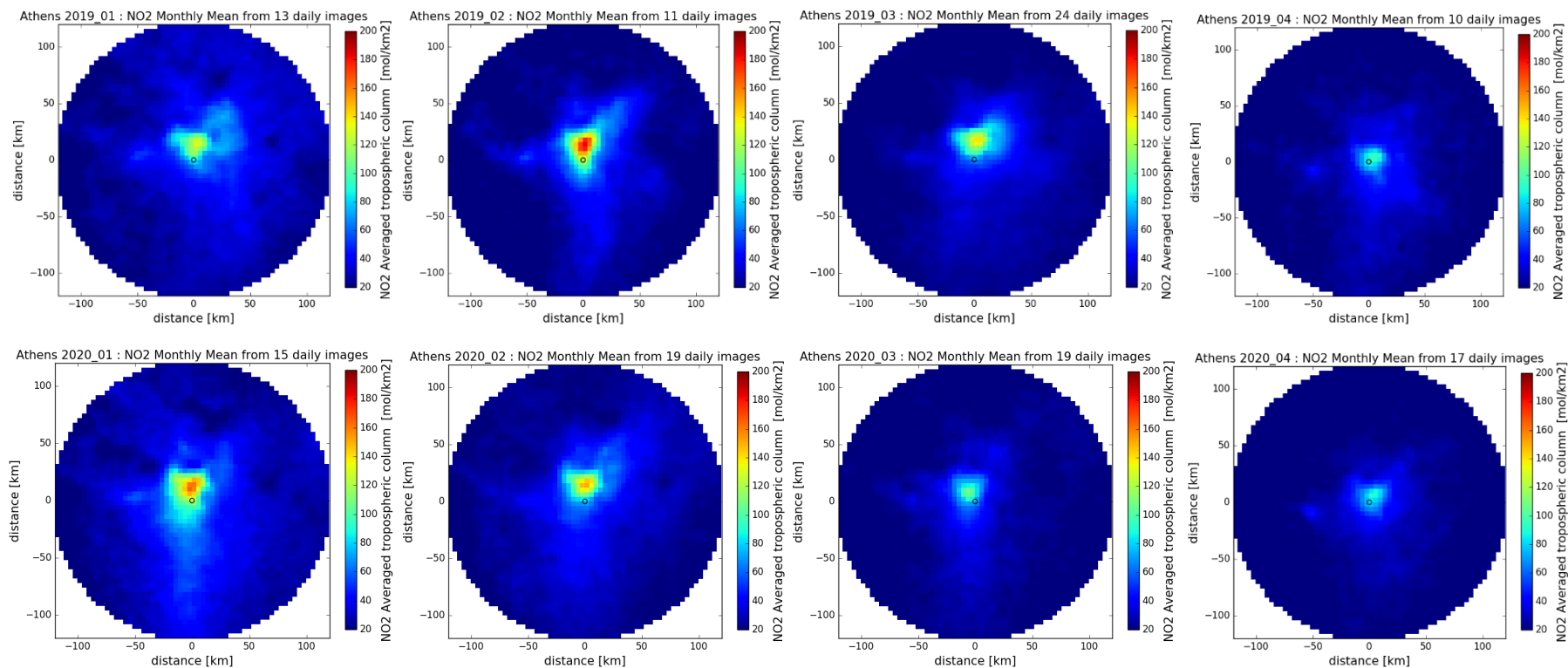
**Figure 29: Reduction rates of the NO<sub>2</sub> plume mass monthly mean between 2019 and 2020 over Athens. A positive value indicates a reduction from 2019 to 2020. The corresponding uncertainty bars are also provided**



**Figure 30: Reduction rates of the NO<sub>2</sub> plume mass fortnightly mean between 2019 and 2020 over Athens, from mid-March to end of April, and corresponding uncertainty.**



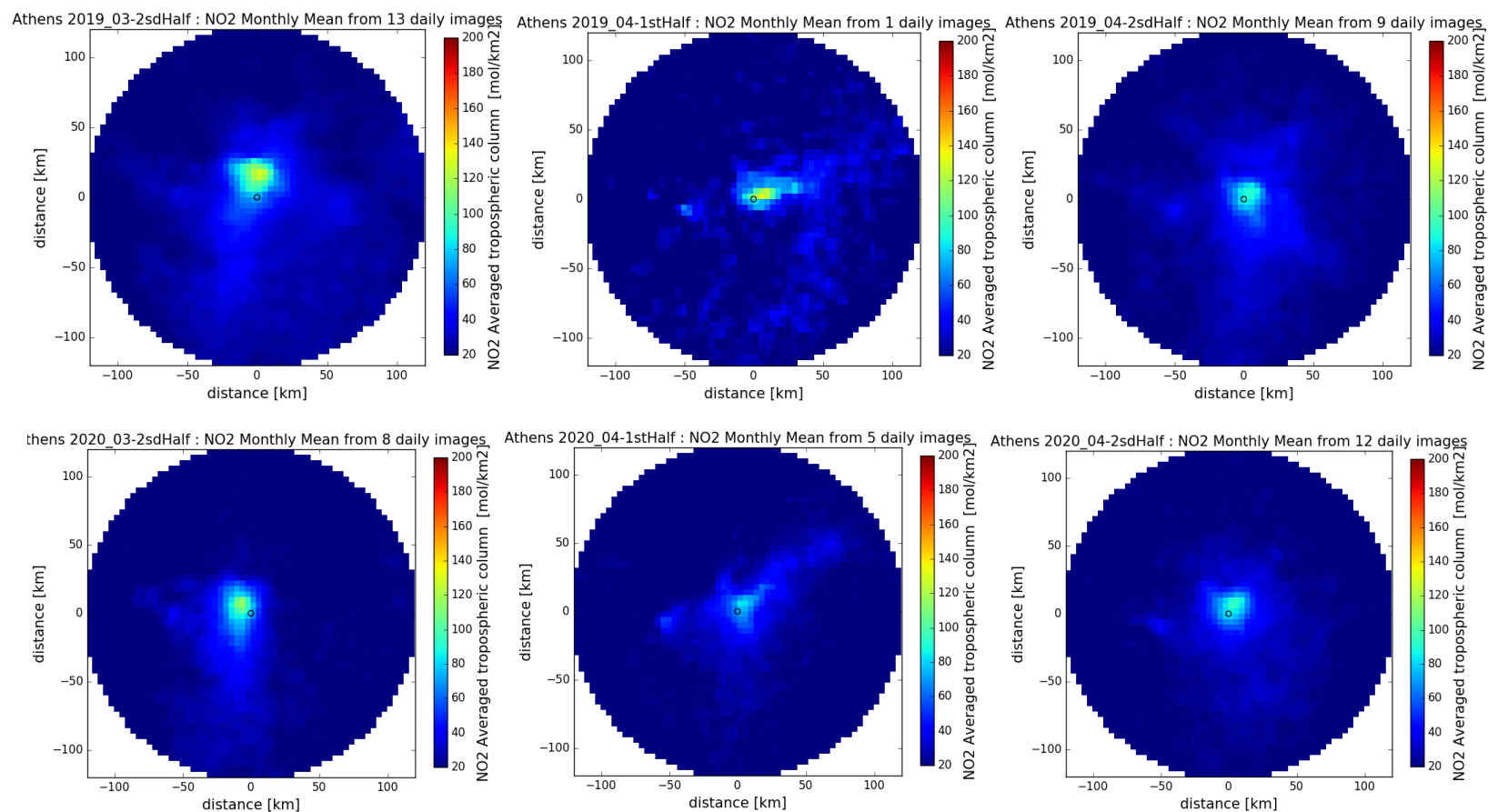
**Figure 31: Reduction rates of the NO<sub>2</sub> plume mass weekly mean between 2019 and 2020 over Athens, from mid-March to end of April, and corresponding uncertainty.**



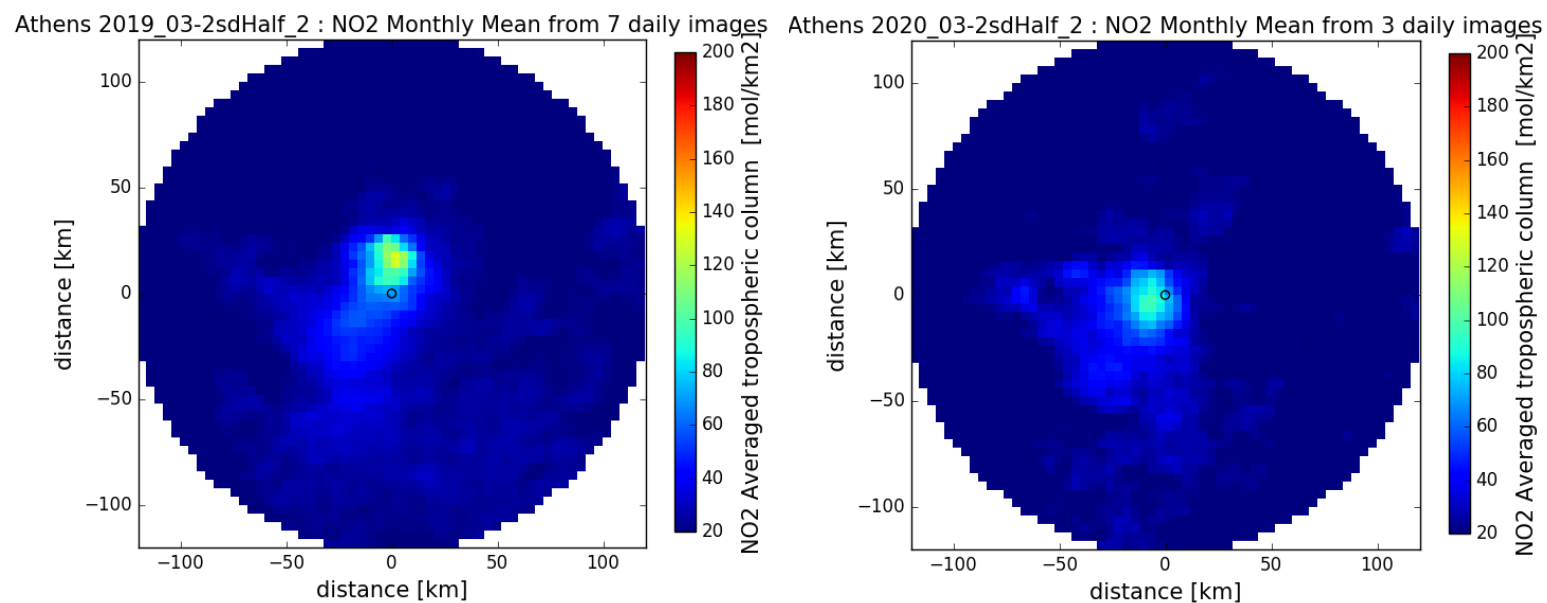
**Figure 32: Monthly mean maps of NO<sub>2</sub> tropospheric column over Athens, for January (left) February (middle) and March (right): Maps for the year 2019 are shown on the top. Maps for the year 2020 are shown at the bottom. The number of valid days used for the monthly mean is indicated in the title of each map.**



The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.



**Figure 33: March fortnightly (from 16 to 31) and April fortnightly (from 1 to 15 and from 16 to 30) maps of NO<sub>2</sub> tropospheric column over Athens: Maps for March 2019 are shown on the top. Maps for March 2020 are shown at the bottom. The number of valid days used for the fortnightly mean is indicated in the title of each map.**



**Figure 34 : Weekly maps of NO<sub>2</sub> tropospheric column for the last week of March (from 23 to 31) over Athens: Maps for March 2019 are shown on the left. Maps for March 2020 are shown on the right. The number of valid days used for the weekly mean is indicated in the title of each map.**

Figure 28 compares the monthly, fortnightly (15-days) and weekly (7-days) averages of total NO<sub>2</sub> plume mass time series from January to April 2019 and 2020. The 2019/2020 comparison of monthly estimates for March and April (Figure 28, top) does not indicate a significant reduction of the total NO<sub>2</sub> plume mass. The same conclusion applies based on fortnightly values in March and April, despite good accuracies due to the large number of valid days in the second half of March 2019 and in the two fortnightly periods of April. Only weekly values suggest slightly significant reductions for last week of March and first week of April.

This analysis is confirmed by the computation of reduction rates at monthly (Figure 29), fortnightly (Figure 30) and weekly (Figure 31) scales. Monthly and fortnightly values associated with their respective uncertainties do not show any significant decrease of NO<sub>2</sub> in the March or April periods at these scales. Only reduction rates for weeks from 26 March to 1 April and from 2 to 8 April have significant reduction values (with 37% +/- 22% and 48% +/- 35%, respectively).

The monthly mean maps for the two periods 2019 and 2020 (Figure 32) and the fortnightly average maps for March 2019 and 2020 (Figure 33) do not provide evidence for a noticeable reduction.

Figure 34 presents the weekly mean maps for the last week of March in 2019 and 2020, suggesting a NO<sub>2</sub> pollution level reduction for this period in 2020 as compared to 2019. The number of valid days available for the last weeks of March 2019 and 2020 (7 valid days in 2019, and 3 valid days in 2020) allows to be reasonably confident in this estimate. This result is consistent with that shown in Figure 28 bottom (weekly mean time series) and Figure 31 (reduction rate) for the corresponding week, in coincidence with the start date of confinement measures in Athens (23<sup>rd</sup> of March 2020), decided later than for the 3 other cities discussed before. However, correlation between this slightly visible reduction in the NO<sub>2</sub> level and some activity reduction is not possible without additional data, and the analysis of S5P data for April does not confirm any impact of confinement measures over Athens.

## 4.7 Conclusion

The TROPOMI instrument on the ESA S5P satellite provides daily measurements from space of the tropospheric column of the NO<sub>2</sub> pollutant at high spatial resolution. In this work, we have applied a dedicated processing for filtering, interpolating and averaging these L2 products at city-scale, in order to assess the impact of the human activity reduction on air pollution by comparing the first four months' periods of 2019 and 2020 on a daily, weekly and monthly basis (whenever possible), for 4 major cities in Europe i.e. Paris, Madrid, Athens and Milan (within the Po valley area).

Reductions in the pollution level (using NO<sub>2</sub> as a proxy) at the end of March 2020 have been significantly observed for the Paris, Milan and Madrid regions, as compared to the same period in 2019. The reduction is suggested but not significantly observed from space for Athens. For Paris, despite a relatively limited number of valid daily data, a quantitative estimate of this reduction of 45% +/- 13% for March and of 52% +/- 9% for April has been obtained, which appears consistent with independent estimates. For Milan, quantitative estimates can be robustly computed: a reduction of the NO<sub>2</sub> pollution level of 30% +/- 4% for March and of 28% +/- 8% for the second half of April is clearly observed. For Madrid, the

monthly mean estimated NO<sub>2</sub> mass reduction for April is of 54% +/-16%, and weekly reduction rates from mid-March to end of April are shown, suggesting a start of decreasing NO<sub>2</sub> from mid-March, with increasing reduction rates in the three first weeks of April, from 32% +/- 21% to 73% +/-22%. Finally, the analysis of the NO<sub>2</sub> pollution level over Athens does not show evidence of reduction (in average), despite a robust estimation due to the large number of available data for this period.

Except for Athens, the reported decreases of NO<sub>2</sub> pollution levels can be attributed to the exceptional city emission reductions, due for a large part to the containment measures against the spread of the Covid-19, which caused reductions in traffic and industrial activities. Some of the observed column reductions could partly result from local meteorological variability and/or impact of long-range atmospheric transport of NO<sub>2</sub>.

As discussed for example by ECMWF/CAMS (2020-1), the interpretation of tropospheric NO<sub>2</sub> concentrations observed from space in terms of actual emissions of pollutants is quite complex, as continual changes in the weather and types of emissions can cause large variations in the atmospheric concentrations of pollutants. The dedicated CAMS web site (ECMWF/CAMS, 2020-3) provided AQ information on European cities in support of the COVID-19 crisis, derived from CAMS numerical analysis on a daily and weekly basis. This web site includes time series of surface NO<sub>2</sub> for 50 cities, including Paris, Milan, Madrid, Athens. The site is providing comparisons of the CAMS 2020 daily average and weekly average values with the range of daily average values from CAMS regional reanalyses of 2017, 2018 and 2019. However, these weekly CAMS time series are not directly comparable with the results presented in the present paper (e.g., weekly time series presented in Figure 7, Figure 16, Figure 21 and Figure 28 for Paris, Milan, Madrid and Athens, respectively). This is mainly due to the specific processing used in our analysis for addressing an “effective city scale” information (processing steps 1, 3 and 4 of Section 3 defining a dedicated area selection, gridding & interpolation, then a specific integration over a background threshold) encompassing the whole city plume in the analysis and comparison of pollution levels. Together with temporal averaging over 7-day, 14-day and monthly windows, this processing allows both to mitigate the impact of weather variability and to find a compromise between small scale source variation and influence of NO<sub>2</sub> from outside the city. Consolidated comparisons would require applying similar scaling onto CAMS analysis data and this could be the subject of further validation studies. A consolidation of our results would include the improvement of the current data processing for a more quantitative estimate of NO<sub>2</sub> emission fluxes at city scales by exploiting wind information that can be obtained directly from NRTI S5P NO<sub>2</sub> products or from ECMWF field analyses.

Additional studies will also be needed to better separate and evaluate the different contributions to the NO<sub>2</sub> burden, by implementing more quantitative methods exploiting wind information to derive NO<sub>2</sub> emissions, such as those already developed by, e.g., Lorente et al. (2019) or Goldberg et al. (2019). However, this first city-scale analysis from space of the exceptional reductions of the NO<sub>2</sub> pollution levels already provides an encouraging illustration of the potential of space-based measurements for monitoring the atmospheric pollution response to current and future city emission regulations. The proposed processing method is demonstrating the usefulness of NO<sub>2</sub> measurements for monitoring AQ within and outside the perimeter of major cities in the world. This case study on 4 European cities is proposed as a testbed for applying the technique to other cities or major industrial areas. In

the long run, such an approach could help national and city authorities to base their policies for improving air quality and use the satellite data (in addition to much more local, but highly fluctuating *in situ* measurements) in their real-time AQ monitoring and forecasting systems. Since pollution can be exported by long range transport outside the emission area or country, international agreements and implemented regulations would also need a strong and independent observing strategy to monitor atmospheric pollution from space. The proposed approach is a step forward in this direction resulting from the European space infrastructure developed by ESA with S5P and further planned on an operational basis with Sentinel 5 and Sentinel 4, in coordination with the Metop-Second-Generation (Metop-SG) and MeteoSat Third Generation (MTG) infrared atmospheric sounding components.

## 5 PCA-based screening of L1 data: Processing development status

---

### ***5.1 Implementation of the processing***

The implementation of the processing can be separated into 4 modules: the module to read the input data and select the different needed products, the module to eventually filter and correct the measurements (L1b spectra), the module to generate the set of eigenvalues (offline) and the module to perform the principal component decomposition/reconstruction. All modules are written in python language.

#### **Input data reader**

This module reads L1B Sentinel-5P data from band07 and band 08. It also selects products and flags needed for the processing.

#### **Filtering and correcting measurements module**

This module detects erroneous measurements thanks to the flags and the analysis of the spectrum. Because principal components method needs a complete spectrum without missing value, it is important to correct bad values and/or to filter not valid spectrum.

#### **Generation of a set of eigenvectors/eigenvalues**

A set of more than 75000 spectra are used, generated by taking one day in every five from one complete year (2018). About 70 spectra per orbit are selected using an oriented-random method that allows to have representative values of longitude, latitude, solar angles. From that reference database, the variance/covariance matrix of the centred and normalised spectra is computed. The principal component decomposition is applied to compute the corresponding eigenvectors, ranked from the highest to the lowest eigenvalues.

#### **Principal components analysis**

Each measured spectra are projected into the set of eigenvectors to obtain principal components coefficients. The spectrum is then reconstructed using principal components and its coefficients, by using only the first eigenvectors up to a given rank. The rank, and thus the number of eigenvectors used for the reconstruction, can be selected by the user. The difference between the reconstructed spectrum and the measured one allows to detect extreme atmospheric events looking at the whole spectral range or some spectral range associated to some gas or molecules.

## ***5.2 Interfacing and pre-processing of S5P L1b spectra***

On going work for dealing with Sentinel 5 Precursor level 1b data, that strongly differ from IASI L1c data. Implementation of the dedicated pre-processing of S5P L1b data, and analysis and evaluation of this first step, are ongoing.