



UNIVERSITÀ DEGLI STUDI DI BERGAMO  
CORSO DI LAUREA MAGISTRALE IN INGEGNERIA  
INFORMATICA  
SCUOLA DI INGEGNERIA

---

## Progetto di Sistemi infomativi per il territorio

---

Open Data Cube e Analisi degli inquinanti nella Lombardia pre  
e post lockdown

*Autori*

Lorenzo CONTI

Samuele FERRI

Fabio SANGREGORIO

Simone SUDATI

30 marzo 2021

# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Struttura dell'elaborato . . . . .	1
1.2	Dati utilizzati e fonti . . . . .	2
1.2.1	Sentinel-5P . . . . .	2
1.2.2	ARPA . . . . .	2
<b>2</b>	<b>Open Data Cube</b>	<b>5</b>
2.1	Cosa è un Open Data Cube? . . . . .	5
2.2	Componenti di un Open Data Cube . . . . .	7
2.3	Installazione di un Open Data Cube . . . . .	8
2.3.1	Requisiti . . . . .	8
2.3.2	Database . . . . .	9
2.3.3	Datacube . . . . .	10
2.3.4	Dataset Sentinel5P . . . . .	11
2.3.5	Generatore automatico YAML . . . . .	15
2.4	Utilizzo . . . . .	15
<b>3</b>	<b>Analisi degli inquinanti</b>	<b>17</b>
3.1	NO <sub>2</sub> : biossido di azoto . . . . .	17
3.2	NO <sub>x</sub> : ossidi di azoto . . . . .	19
3.3	PM <sub>10</sub> . . . . .	20
3.4	PM <sub>2.5</sub> . . . . .	23
3.5	Correlazioni tra inquinanti . . . . .	24
<b>4</b>	<b>Analisi dell'impatto del lockdown</b>	<b>27</b>
4.1	Dati di base . . . . .	27
4.2	Dati degli inquinanti . . . . .	30
4.3	NO <sub>2</sub> . . . . .	31
4.3.1	Concentrazione media nel 2019 . . . . .	31
4.3.2	Concentrazione media nel 2020 . . . . .	33

4.3.3	Impatto del periodo di lockdown . . . . .	35
4.4	$\text{NO}_x$ . . . . .	37
4.4.1	Concentrazione media nel 2019 . . . . .	37
4.4.2	Concentrazione media nel 2020 . . . . .	37
4.4.3	Impatto del periodo di lockdown . . . . .	37
4.5	$\text{PM}_{10}$ . . . . .	41
4.5.1	Concentrazione media nel 2019 . . . . .	41
4.5.2	Concentrazione media nel 2020 . . . . .	41
4.5.3	Impatto del periodo di lockdown . . . . .	44
4.6	$\text{PM}_{2.5}$ . . . . .	46
4.6.1	Concentrazione media nel 2019 . . . . .	46
4.6.2	Concentrazione media nel 2020 . . . . .	46
4.6.3	Impatto del periodo di lockdown . . . . .	46
4.7	Composizione dell'aria . . . . .	50
4.7.1	Composizione degli inquinanti nel 2019 . . . . .	50
4.7.2	Composizione degli inquinanti nel 2020 . . . . .	50
<b>5</b>	<b>Analisi dei regressori</b>	<b>53</b>
5.1	Analisi del traffico veicolare . . . . .	54
5.1.1	Reperimento ed elaborazione dati . . . . .	54
5.1.2	Risultati analisi . . . . .	54
5.2	Analisi della rete stradale . . . . .	57
5.2.1	Reperimento ed elaborazione dati . . . . .	57
5.2.2	Risultati analisi . . . . .	57
5.3	Analisi della popolazione . . . . .	59
5.3.1	Reperimento ed elaborazione dati . . . . .	59
5.3.2	Risultati analisi . . . . .	59
5.4	Analisi della temperatura . . . . .	61
5.4.1	Reperimento ed elaborazione dati . . . . .	61
5.4.2	Risultati analisi . . . . .	62
<b>6</b>	<b>Interpretazione dei risultati</b>	<b>65</b>
6.1	Confronto tra i regressori . . . . .	65
<b>7</b>	<b>Conclusioni</b>	<b>67</b>
7.1	Sviluppi futuri . . . . .	68
<b>8</b>	<b>Sviluppo</b>	<b>69</b>
8.1	Python . . . . .	69
8.1.1	xarray . . . . .	69

8.1.2	pandas . . . . .	69
8.1.3	geopandas . . . . .	69
8.1.4	matplotlib . . . . .	70
8.1.5	numpy . . . . .	70
<b>A</b>	<b>Preparazione dataset degli inquinanti</b>	<b>79</b>
<b>B</b>	<b>Kriging degli inquinanti</b>	<b>81</b>

# Capitolo 1

## Introduzione

In questo elaborato viene presentato uno studio riguardante l'inquinamento dell'aria nella regione Lombardia, relativamente a quattro fattori inquinanti principali: PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> e, più in generale, la famiglia degli NO<sub>x</sub>.

Viene posta una particolare attenzione al cambiamento che il lockdown ha comportato in termini di concentrazione di inquinanti dell'atmosfera. Si vedrà come diverse zone e diversi inquinanti abbiano subito cambiamenti differenti: dove possibile si cercherà di spiegare le cause che comportano la presenza di un determinato inquinante e si correleranno queste informazioni in modo qualitativo attraverso mappe e grafici.

Per ciascun inquinante sono stati reperiti i dati nel periodo da Marzo a Luglio sia del 2019 che del 2020. Si considera che questo intervallo temporale permetta di analizzare l'effetto che il lockdown ha avuto sulle condizioni atmosferiche<sup>1</sup>.

### 1.1 Struttura dell'elaborato

In questo capitolo introduttivo verranno presentate le fonti dei dati ed elencati i trattamenti preliminari svolti.

Nel [Capitolo 2](#) si presenta Open Data Cube, una tecnologia innovativa per l'accesso e l'indicizzazione dei dati. Si spiegano i concetti alla base di tale tecnologia e si spiega la procedura di implementazione per importare un file in formato NetCDF all'interno del datacube.

Nel [Capitolo 3](#) si presenta una descrizione degli inquinanti e dei dati ricavati, con particolare attenzione alle caratteristiche quantitative.

Nel [Capitolo 4](#) si presenta la vera e propria analisi su cui verte questo elaborato. Si mostra dal punto di vista qualitativo come gli inquinanti si distribuiscono sul territorio, quali

---

<sup>1</sup>Considerando che il lockdown sia iniziato verso la fine di Febbraio 2020 e che gli effetti si siano manifestati nel breve termine.

sono i fattori che li influenzano, come il lockdown ha impattato sulle loro concentrazioni e le loro conseguenze.

Nel [Capitolo 5](#) vengono elencati i principali regressori identificati per gli inquinanti e il loro effetto sulla concentrazione di inquinamento nel territorio.

Nel [Capitolo 6](#) viene mostrata una conclusione qualitativa dei risultati ottenuti dalla sperimentazione.

Nel [Capitolo 7](#) si elencano le conclusioni principali dell'elaborato, dell'esperienza del gruppo e delle conoscenze acquisite durante lo svolgimento degli esperimenti.

Nel [Capitolo 8](#) si elencano gli strumenti utilizzati nella conduzione degli esperimenti e nella stesura dell'elaborato.

## 1.2 Dati utilizzati e fonti

### 1.2.1 Sentinel-5P

Sentinel-5P [1], per esteso *Sentinel-5 Precursor*, è un satellite lanciato il 13 ottobre 2017 per il telerilevamento. Esso è stato sviluppato dall'ESA<sup>2</sup> (Agenzia spaziale europea) come parte del programma Copernicus per restringere l'intervallo di discontinuità d'osservazione tra l'Envisat ed il Sentinel-5. Copernicus<sup>3</sup> è il programma di osservazione terrestre dell'Unione europea che sfrutta una serie di satelliti dedicati (le famiglie Sentinel). La missione che si è preposto il Sentinel-5P è di riuscire a fornire servizi e informazioni circa la qualità dell'aria e il clima dal 2017 al 2023. Per accedere ai dati del Sentinel-5P è necessario inviare una richiesta POST alla API di processo. I dati richiesti verranno restituiti come risposta alla richiesta. Ogni richiesta POST può essere personalizzata per ottenere esattamente i dati di cui si ha bisogno.

### 1.2.2 ARPA

ARPA [2], acronimo per agenzia regionale per la protezione dell'ambiente, è un ente della pubblica amministrazione italiana, costituito e operante in ogni regione d'Italia. In particolare, l'ARPA Lombardia<sup>4</sup> mette a disposizione indicatori e dati ambientali raccolti da sensori dislocati sul territorio lombardo. In questo elaborato ci limitiamo allo studio di quattro inquinanti presenti nell'aria.

I dati forniti da ARPA sono scaricati in formato tabellare (.csv), con una cadenza giornaliera e per un periodo temporale che copre i due intervalli di interesse (Marzo-Luglio del 2019 e del 2020).

Viene quindi fatta una prima analisi preliminare dei dati, viene pulito il dataset da even-

---

<sup>2</sup>ESA (<http://www.esa.int/>)

<sup>3</sup>Copernicus (<https://www.copernicus.eu/it>)

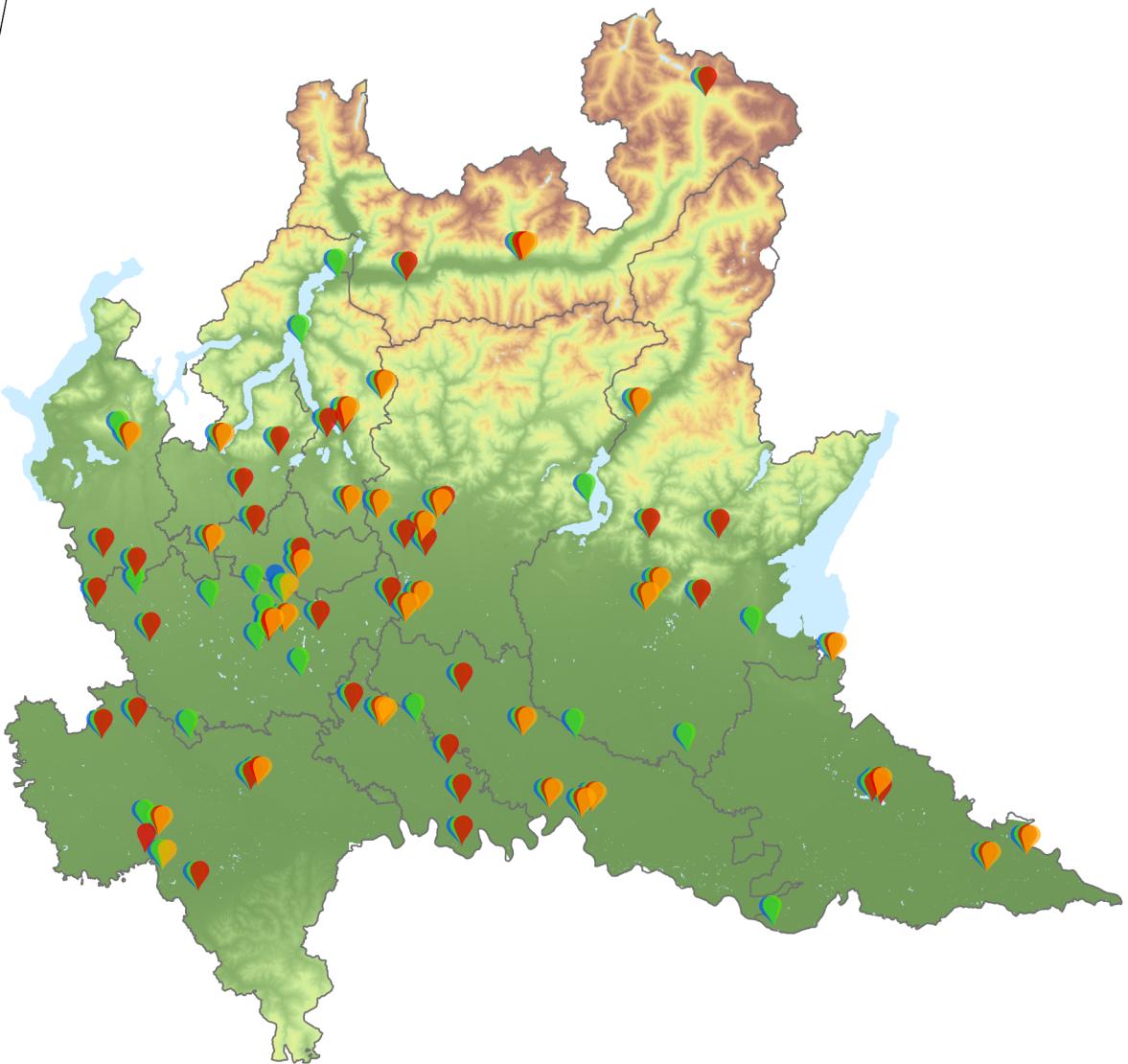
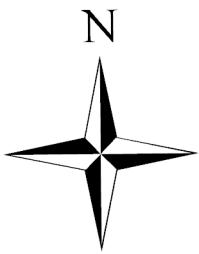
<sup>4</sup>ARPA (<https://www.arpalombardia.it/>)

tuali valori mancanti, viene filtrato sul periodo di interesse ed i dati vengono aggregati tramite media ad un singolo valore per ciascuno dei due intervalli di riferimento precedentemente menzionati.

Questo permette di avere una stima della concentrazione media di inquinante nello stesso periodo di due anni successivi e permette di osservare variazioni dovute all'impatto del lockdown distribuite su un ampio arco temporale.

A ciascun sensore quindi corrispondono due valori di inquinante (2019 e 2020). È importante osservare che ciascuna centralina ARPA è dotata di uno o più sensori ma non necessariamente di un sensore per inquinante.

Nella mappa seguente si può osservare la distribuzione dei sensori sul territorio. In primo luogo si nota come i sensori relativi a NO<sub>2</sub> e NOx siano molti di più rispetto a quelli di PM10 e PM2.5, in secondo luogo si osserva come siano distribuiti in modo disomogeneo sul territorio, soprattutto concentrati sui capoluoghi di provincia (Milano soprattutto), molto meno in montagna e nella pianura bresciana e mantovana. Sono presenti delle stazioni che possiedono sensori per tutti e quattro gli inquinanti ed è presente almeno una stazione per provincia.



📍 Sensori NO<sub>2</sub> (84)

📍 Sensori NOX (83)

📍 Sensori PM10 (64)

📍 Sensori PM2.5 (32)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# Geolocalizzazione dei sensori

# Capitolo 2

## Open Data Cube

Nella realizzazione del progetto è stato sperimentato l'utilizzo della tecnologia Open Data Cube (anche noto come ODC) per l'accesso e l'indicizzazione dei dati geografici, eseguendo la creazione di un Open Data Cube costruito sul territorio lombardo riguardante i dati di sostanze inquinanti.

In questo capitolo si presenta una breve introduzione sulla tecnologia e sul suo utilizzo, si mostrano i componenti necessari alla realizzazione di un Open Data Cube, la sua installazione ed esecuzione, e infine le procedure adottate per l'importazione e indicizzazione dei dati.

### 2.1 Cosa è un Open Data Cube?

Open Data Cube [3] è un progetto open source per la gestione di dati satellitari derivanti da sorgenti multiple: la sua potenzialità principale è la possibilità di accedere, organizzare e analizzare grandi quantità di osservazioni territoriali eterogenee (Geographic Information Sysyem (GIS) data, ossia Earth Observations, o dati EO). Esso è un framework comune a questi dati, il quale mette a disposizione una serie di strumenti per facilitare l'analisi di grandi quantità di dati e tracciarne la provenienza. Il suo scopo finale può essere riassunto dalla [Figura 2.3](#), ossia la raccolta di dati satellitari eterogenei tramite la cui gestione ed analisi è possibile prendere decisioni sull'ambiente analizzato.

Al suo interno, ODC si compone di una serie di librerie Python ed un database PostgreSQL. Il core di ODC consiste in un layer posto tra i provider di dati satellitari e le applicazioni.

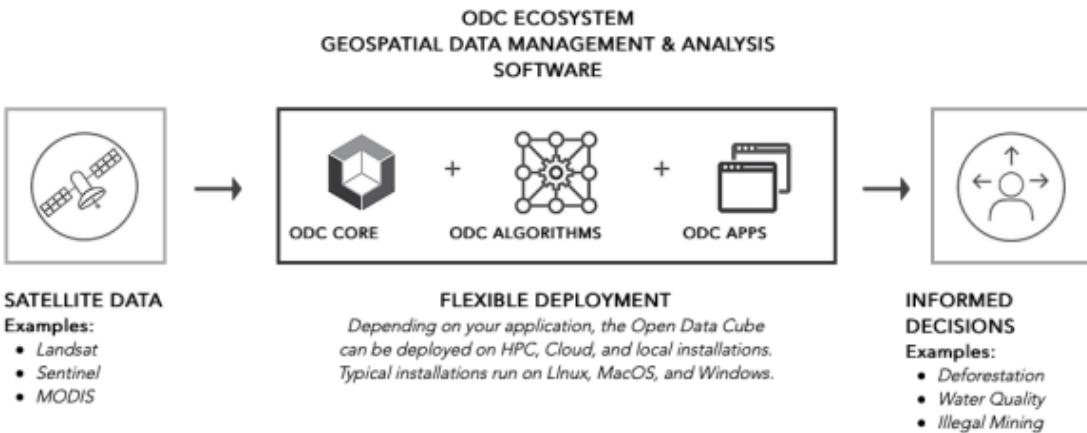


Figura 2.1: ODC come layer tra provider ed applicazioni

Il progetto è in continuo sviluppo e ogni anno nascono nuovi Data Cube, sia governativi che di associazioni private, i quali collezionano i dati di specifici stati o regioni mondiali. L'iniziativa ha attualmente tre importanti Data Cube operativi:

- Digital Earth Australia: è l'implementazione del governo australiano, la cui piattaforma comprende documentazione, guide, tutorial e supporto per utenti internazionali.
- Africa Regional Data Cube: comprende Kenya, Senegal, Sierra Leone, Ghana e Tanzania, e si concentra sull'affrontare le esigenze locali e nazionali, oltre agli obiettivi sostenuti da *Group on Earth Observations (GEO)* e *United Nations Sustainable Development Goals (UN-SDG)*.
- Swiss Data Cube: uno dei primi ad adottare la tecnologia su scala nazionale. Contiene attualmente 30 anni di osservazioni Landsat ed osservazioni di Sentinel 1 e Sentinel 2 per l'intero paese. L'interesse di questo Data Cube è l'urbanizzazione, meteo e manti nevosi.

In Figura 2.2 si può vedere la distribuzione geografica mondiale delle installazione attuali dell'iniziativa.



Figura 2.2: La posizione delle installazioni attuali di Open Data Cube nel mondo

## 2.2 Componenti di un Open Data Cube

Da un punto di vista tecnico, un'implementazione di ODC è composta da tre componenti: i dati, un indice ed il software.

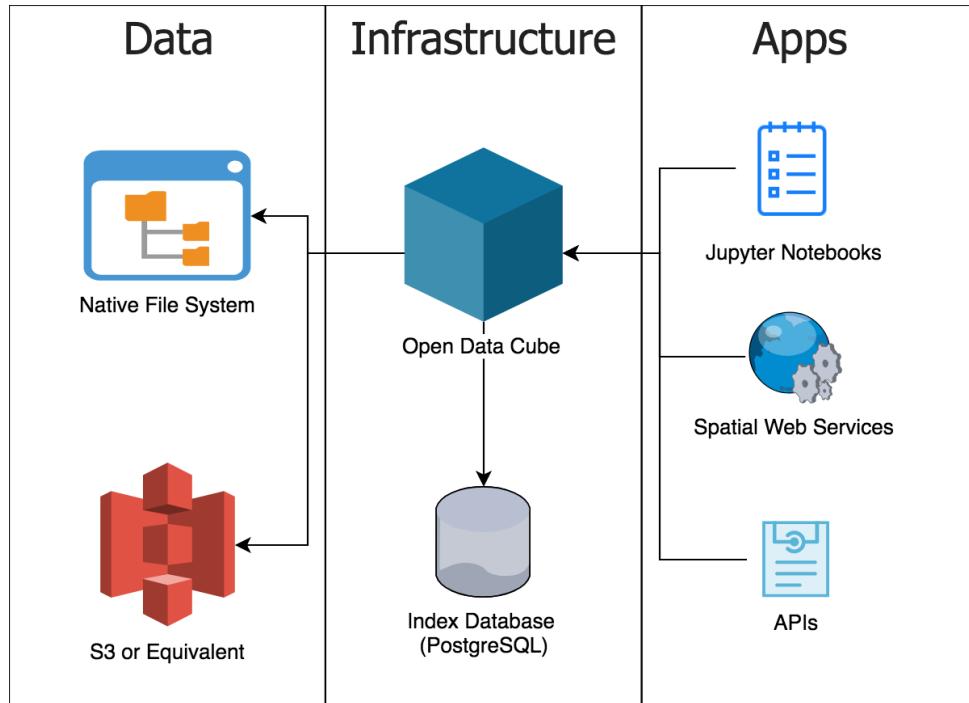


Figura 2.3: Schema dell'ecosistema Open Data Cube

I **dati** sono generalmente ottenuti da file in formato GeoTIFF o NetCDF. In generale però

essi possono essere in qualsiasi formato che sia leggibile e gestibile da GDAL, una libreria di traduttori per formati di dati geospaziali raster e vettoriali utilizzata da ODC.

Il **software** al centro di ODC è una libreria Python che permette all’utente di indicizzare i dati e, grazie a questo, gestirli in maniera ottimizzata in termini di performance. Permette inoltre la richiesta di dati (ossia permette di effettuare *query*) ritornando i risultati in formati di dati standard ed una vasta gamma di funzioni per la gestione dei dati.

Per l’**indicizzazione** Open Data Cube utilizza un database PostgreSQL per memorizzare una lista di *Product* (un tipo di dato specifico relativo alla sorgente dei dati come, ad esempio, *Landsat 8 Analysis Ready Data*) e di *Dataset* (una singola istanza di un *Product* come, ad esempio, una singola scena di Landsat 8). L’indice permette all’utente di richiedere dati ad una certa posizione spaziale e temporale, senza la necessità di sapere specificatamente dove i file richiesti sono stati memorizzati e come accedervi.

## 2.3 Installazione di un Open Data Cube

Per procedere con il setup di un Open Data Cube, ci siamo basati (in parte) sulla documentazione ufficiale<sup>1</sup>. Nelle prossime sottosezioni si spiegano tutti i passi effettuati ed i comandi utilizzati.

Nella Figura 2.4 sono rappresentati schematicamente tutti i passi necessari per la realizzazione di un ODC.

### 2.3.1 Requisiti

Si è usato Python v3.8.6, altre versioni potrebbero causare a problemi di incompatibilità. Innanzitutto si crea un nuovo ambiente virtuale in Python, tramite i seguenti comandi:

```
1 python -m venv env  
2 .\env\Scripts\activate
```

Listing 2.1: Creazione ambiente virtuale in Python

Successivamente si installano tutti i pacchetti necessari presenti, con le relative versioni compatibili, presenti nel file `requirements.txt`. Da notare che alcuni pacchetti non sono gestiti dal package manager di Python ma sono installati tramite i file binari con formato `*.whl files`.

---

<sup>1</sup>Open Data Cube Documentation (Installation) (<https://datacube-core.readthedocs.io/en/latest/ops/install.html>)

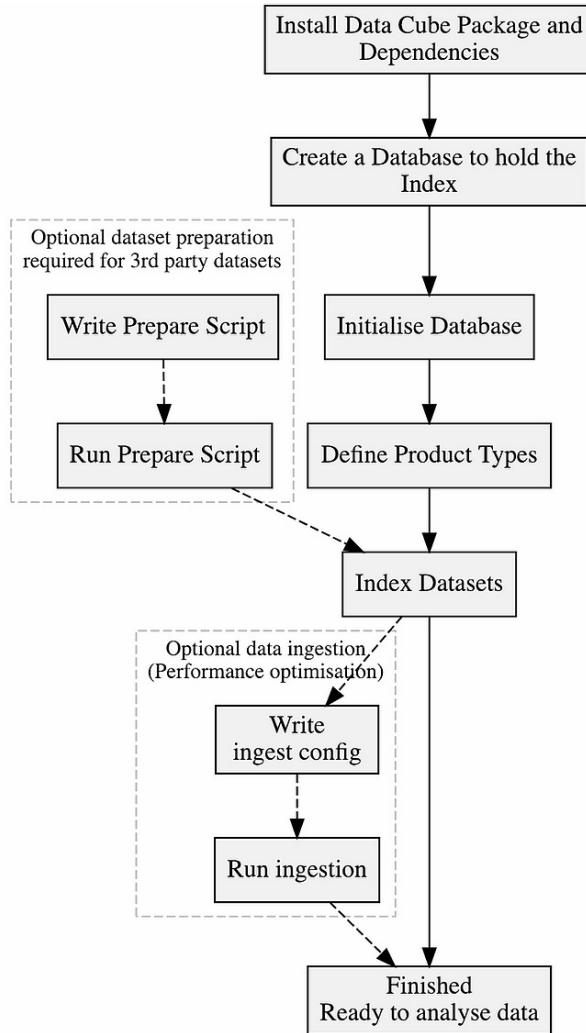


Figura 2.4: Schema realizzazione ODC

```
1 pip install -r requirements.txt
```

Listing 2.2: Installazione dei requisiti

Successivamente si installa Jupyter per gestire tutti i notebook prodotti contenenti le analisi effettuate.

```
1 pip install jupyter
2 jupyter notebook
```

Listing 2.3: Installazione e uso dei Jupyter Notebooks

### 2.3.2 Database

Si procede con il setup del database necessario per ODC. In particolare abbiamo utilizzato PostgreSQL<sup>2</sup>.

---

<sup>2</sup>PostgreSQL (<https://www.postgresql.org/download/>)

Innanzitutto si aggiunge PostgreSQL alle variabili d'ambiente di sistema indicandone il percorso in questo modo:

```
1 C:\Program Files\PostgreSQL\13\bin  
2 C:\Program Files\PostgreSQL\13\lib
```

Listing 2.4: Variabili d'ambiente di sistema di PostgreSQL

Successivamente da terminale è possibile creare il database da chiamato (da noi per convenzione) **datacube** con i seguenti comandi:

```
1 psql -U postgres  
2 > CREATE DATABASE datacube;
```

Listing 2.5: Creazione database

Successivamente si crea un file di configurazione per il database, nominato **.datacube.conf** (importante notare il punto all'inizio) e posizionato nella cartella utente. All'interno del file si inseriscono le seguenti configurazioni:

```
1 [datacube]  
2 db_database: datacube  
3  
4 # A blank host will use a local socket. Specify a hostname (such as  
# localhost) to use TCP.  
5 db_hostname: localhost  
6  
7 # Credentials are optional: you might have other Postgres authentication  
# configured.  
8 # The default username otherwise is the current user id.  
9 db_username: datacube  
10 db_password: 0
```

Listing 2.6: Configurazione database (File **.datacube.conf**)

### 2.3.3 Datacube

Si procede con l'installazione del datacube vera e propria. È necessario scaricare i file contenuti nella repository GitHub **datacube-core**<sup>3</sup> dal branch **develop**.

Per procedere con l'installazione, si esegue il file **setup.py** presente dentro la cartella clonata.

```
1 cd datacube-core-develop
```

---

<sup>3</sup>Repository **datacube-core** (<https://github.com/opendatacube/datacube-core>)

```
2 python setup.py install
```

Listing 2.7: Installazione datacube

Si può verificare la versione del datacube con il seguente comando:

```
1 datacube --version
```

Listing 2.8: Versione del datacube

Nel caso di errori con il pacchetto `numpy` si può risolvere installando la versione 1.19.3 con il seguente comando:

```
1 pip install numpy==1.19.3
```

Listing 2.9: Versione del datacube

Successivamente si inizializza lo schema del database

```
1 datacube -v system init
```

Listing 2.10: Inizializzazione database

### 2.3.4 Dataset Sentinel5P

In questa sezione facciamo riferimento al dataset Sentinel5P fornito dal Politecnico di Milano etichettato come `dataset.nc` in formato NetCDF con le convenzioni HARP.

L'indicizzazione dei dati viene fatta tramite due file in formato YAML (*Dataset* e *Product*), i quali indicano le informazioni dei metadati necessarie alla corretta importazione dei dati. Nel caso non fossero forniti con il file NetCDF, si possono creare manualmente ad hoc o utilizzato lo script presentato nella [sottosezione 2.3.5](#).

#### Product YAML

La definizione del *Product* fornisce un **nome**, una **descrizione**, alcuni **metadati di origine** di base e un elenco di **misurazioni** che descrivono il tipo di dati che sono contenuti nei dataset. Le misurazioni sono un elenco ordinato di dati che specificano un **nome** ed alcuni **alias**, un tipo di dati (o **dtype**) e alcune opzioni extra, tra cui il tipo di unità in cui si trova la misurazione ed un valore **nodata** da inserire come valore di default nel caso non siano presenti i dati della misurazione.

Dopo averlo creato si aggiunge il prodotto al datacube con il seguente comando

```
1 datacube product add product.yaml
```

Listing 2.11: Aggiunta del prodotto al datacube

Nella nostra realizzazione del Data Cube abbiamo utilizzato i dati sulle misurazioni `cloud_pressure` incluse nel file NetCDF di Sentinel5P, fornite dal Politecnico di Milano. Il relativo file `product.yaml` costruito per l'inclusione di questi dati si può trovare in [Listing 2.12](#).

```
1 name: sentinel5p
2 description: Sentinel5P NetCDF with various layers
3 metadata_type: eo3
4
5 metadata:
6     product:
7         name: sentinel5p
8
9 measurements:
10    - name: 'cloud_pressure'
11        dtype: float64
12        nodata: -1
13        units: 'Pa'
14
15 storage:
16     driver: NetCDF CF
17     crs: EPSG:4326
18     dimension_order: [time, latitude, longitude]
```

Listing 2.12: Esempio del file `product.yaml`

## Dataset YAML

Per quanto riguarda il dataset YAML abbiamo usato un notebook di Jupyter chiamato `dataset.ipynb`.

Ogni dataset richiede un documento di metadati che descriva ciò che i dati rappresentano e da dove provengono, nonché il formato in cui sono archiviati. Come minimo, sono necessarie le dimensioni o i campi in base ai quali si desidera eseguire la ricerca, come latitudine, longitudine e tempo, ma si possono includere tutte le informazioni ritenute utili. Viene memorizzato nell'indice per la ricerca, l'interrogazione e l'accesso ai dati.

Bisogna specificare tutti i parametri richiesti dalla convenzione EO3<sup>4</sup>, estrapolandoli se necessario dal NetCDF.

I *dataset metadata documents* definiscono dei metadati critici riguardo al dataset tra cui:

- Dati dei sensori disponibili
- Nome della piattaforma e dei sensori

---

<sup>4</sup>EO3 Format ([https://datacube-core.readthedocs.io/en/latest/ops/dataset\\_documents.html](https://datacube-core.readthedocs.io/en/latest/ops/dataset_documents.html))

- Estensioni e proiezione geospaziali
- Data di acquisizione
- Informazioni sulla provenienza

Nel notebook si è analizzato il `dataset.nc` dopo aver creato i due file di supporto.

Per validare il file `dataset.yaml` prodotto si può usare il seguente comando (l'opzione `--thorough` tenta di leggere i dati e le misure andando a verificare che le loro proprietà coincidano con quelle richieste dal product):

```
1 eo3-validate "dataset.yaml"
2 eo3-validate --thorough "dataset.yaml"
```

Listing 2.13: Validazione del `dataset.yaml`

Per aggiungere il file `dataset.yaml` validato al datacube bisogna usare il seguente comando

```
1 datacube dataset add --auto-match dataset.yaml
```

Listing 2.14: Aggiunta del `dataset.yaml` al datacube

In caso di errore con il pacchetto `shapely` questo comando può risolvere andando a installare direttamente dai binari (usando file `.whl`):

```
1 pip uninstall shapely
2 pip install C:\GitHub\unibg-gis\whls\Shapely-1.7.1-cp38-cp38-win_amd64.whl
3
4 # Alternative, inside Jupyter Notebook
5 from shapely import speedups
6 speedups.disable()
```

Listing 2.15: Risolvere problemi con `shapely`

In caso di errore con il pacchetto `sqlalchemy` questo comando può risolvere andando a installare direttamente dai binari (usando file `.whl`):

```
1 pip install sqlalchemy==1.3.20
```

Listing 2.16: Risolvere problemi con `sqlalchemy`

Nel Listing 2.17 è riportato un esempio riguardante il file NetCDF riguardante Sentinel5P.

```
1 $schema: https://schemas.opendatacube.org/dataset
2
```

```

3 # UUID of the dataset
4 id: fb3afcb0-4301-57c5-8455-35a64e3b0c53
5
6 # Product name
7 product:
8   name: sentinel5p
9   href: https://www.sentinel-hub.com/sentinel-5/
10
11 # Native CRS, assumed to be the same across all bands
12 crs: epsg:4326
13
14 # Optional GeoJSON object in the units of native CRS
15 # Defines a polygon such that, all valid pixels across all bands are inside
16 # this polygon
16 geometry:
17   type: Polygon
18   coordinates: [[[15, 47], [15, 42], [5, 42], [5, 47], [15, 47]]]
19
20 # Mapping name:str -> { shape: Tuple[ny: int, nx: int]
21 #                           transform: Tuple[float x 9]}
22 # Captures image size, and geo-registration
23 grids:
24   default:
25     shape: [500,1000]
26     transform: [1,0,0,0,1,0,0,0,1]
27
28 # Lineage only references UUIDs of direct source datasets
29 # Mapping name:str -> [UUID]
30 lineage: {} # Set to empty object if no lineage is defined
31
32 # Per band storage information and references into `grids`
33 # Bands using "default" grid should not need to reference it
34 measurements:
35   cloud_pressure:
36     layer: cloud_pressure
37     path: dataset.nc
38
39 # Dataset properties, prefer STAC standard names here
40 # Timestamp is the only compulsory field here
41 properties:
42   # ODC specific "extensions"
43   odc:file_format: NetCDF
44   odc:processing_datetime: 2020-01-01T07:02:54.188Z
45   # If it's a single time instance use datetime
46   datetime: 2020-01-01T07:02:54.188Z # Use UTC

```

Listing 2.17: Esempio del file dataset.yaml

### 2.3.5 Generatore automatico YAML

Se i dati sono già stati preparati dalla sorgente per l'inclusione nel progetto Open Data Cube, essi potranno essere importati nel database senza ulteriori passi, in quanto saranno già pacchettizzati con documenti *dataset* e *product* compatibili con ODC e saranno quindi già pronti all'indicizzazione immediata.

Nel caso in cui invece i dati siano da sorgenti esterne o non compatibili, sarà necessario generare questi documenti *dataset* e *product* a mano. Questo è ciò che abbiamo fatto per la prima importazione dei dati nel Data Cube. Successivamente abbiamo creato un cosiddetto "Data Preparation Script", ossia uno script python che legge il formato dei metadati del file di input e genera automaticamente i documenti *dataset* e *product* necessari all'importazione.

Lo script di generazione automatica è presente nel file  
`/odc/yaml_generator/generate_yaml.py` della repository.

## 2.4 Utilizzo

Una volta installato l'ambiente ed istanziato il datacube (specificando ed inserendo i due file Product e Dataset prima menzionati) è possibile importare i dati in un notebook Python nel seguente modo.

```
1 import datacube
2
3 dc = datacube.Datacube()
4 print(dc)
5
6 Output >> Datacube<index=Index<db=PostgresDb<engine=Engine(postgresql://
    postgres:***@:5432/opendatacube)>>>
```

Listing 2.18: Importare un datacube in Python

In questo modo si crea un'istanza del Datacube all'interno del codice. È possibile accedere alla lista dei product inseriti nel datacube tramite il comando:

```
1 dc.list_products()
```

Listing 2.19: Lista dei product nel datacube

che mostrerà la presenza del product `sentinel5p`. A questo punto è possibile accedere al prodotto caricandolo direttamente dal datacube come segue:

```
1 ds = dc.load('sentinel5p', output_crs="EPSG:4326", resolution=(0.01, 0.01))
```

Listing 2.20: Accesso al product

Il contenuto di `ds` a questo punto risulta:

```
▶ Dimensions:      (latitude: 500, longitude: 1000, time: 1)
▼ Coordinates:
  time          (time)           datetime64[ns] 2020-01-01T07:02:54...
  latitude       (latitude)        float64 42.01 42.02 42.03 ...
  longitude      (longitude)       float64 5.005 5.015 5.025 ...
  spatial_ref    ()              int32 4326
▼ Data variables:
  cloud_pressure (time, latitude, longitude)   float64 -1.0 -1.0 -1.0 ...
▼ Attributes:
  crs :           EPSG:4326
  grid_mapping :  spatial_ref
```

Figura 2.5: Esempio del contenuto di un product

Infine, la libreria `xarray` su cui Datacube si basa permette di mostrare graficamente i dati contenuti nel datacube. Nell'esempio in [Figura 2.6](#) si mostra una rappresentazione grafica di uno dei layer contenuti all'interno del dataset sentinel5p inserito.

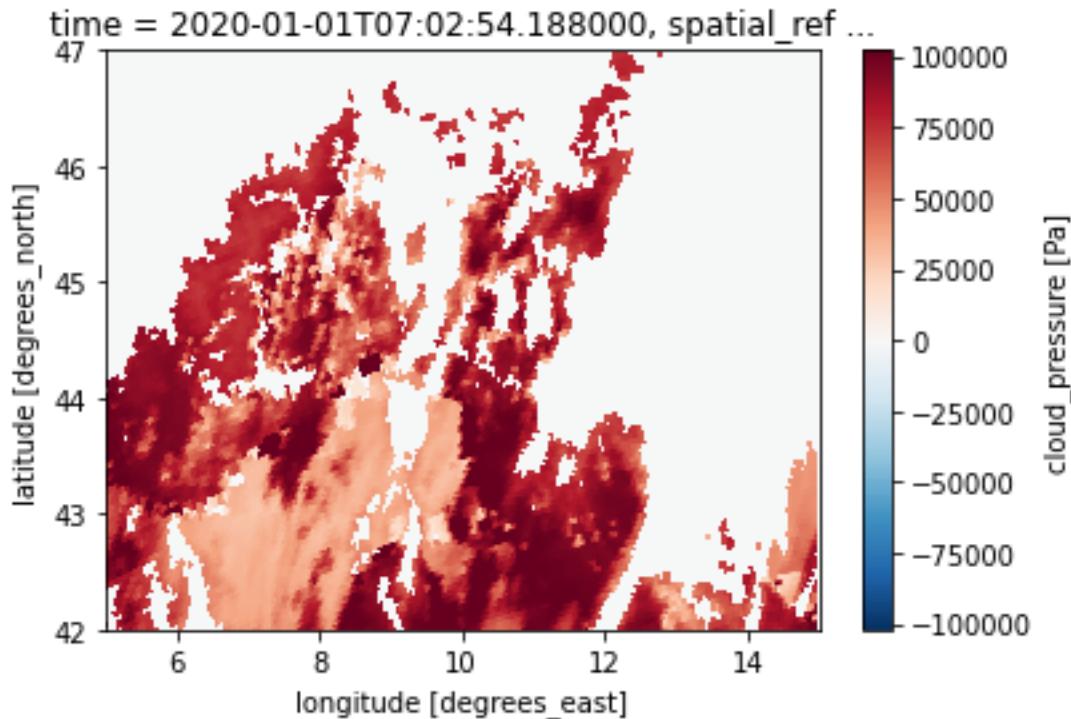


Figura 2.6: Visualizzazione grafica del contenuto del product

# Capitolo 3

## Analisi degli inquinanti

In questo capitolo si elencano gli inquinanti oggetto dell'analisi e si forniscono alcune informazioni legate alla loro natura.

Il sistema di riferimento utilizzato per ogni mappa è il WGS 1984 UTM Zone 32N.

### 3.1 $\text{NO}_2$ : biossido di azoto

Il biossido di azoto  $\text{NO}_2$  [4] è un gas di colore rosso bruno, dall'odore forte e pungente, altamente tossico e irritante. Essendo più denso dell'aria tende a rimanere a livello del suolo. E' un forte agente ossidante e reagisce violentemente con materiali combustibili e riducenti mentre in presenza di acqua è in grado di ossidare diversi metalli. Il biossido di azoto è un inquinante secondario, che si forma in atmosfera principalmente per ossidazione del monossido di azoto (NO). Dal punto di vista chimico è una molecola fortemente reattiva composta da un atomo di azoto e da due atomi di ossigeno. In quanto fortemente reattivo partecipa a numerose reazioni chimiche che portano alla formazione di altri inquinanti, tra i quali l'ozono. Il biossido di azoto si forma in ambienti ad alta densità energetica (alte temperature) come, ad esempio, nelle combustioni di origine antropica e all'interno dei fulmini.

Le emissioni da fonti antropiche derivano sia da processi di combustione (centrali termoelettriche, riscaldamento, traffico), che da processi produttivi senza combustione (produzione di acido nitrico, fertilizzanti azotati, ecc.). Il traffico veicolare risulta essere la principale fonte di emissione di  $\text{NO}_2$ , come si può notare nel [Capitolo 5](#) in cui si mostra il legame del traffico urbano e la presenza di sostanze inquinanti.

## Parametri normativi correnti

Inquinante	Tipo di limite	Limite
NO <sub>2</sub>	Limite orario	200 $\mu\text{g}/\text{m}^3$ (non oltre >18 volte l'anno)
NO <sub>2</sub>	Limite annuale	40 $\mu\text{g}/\text{m}^3$ media annua
NO <sub>2</sub>	Soglia di allarme	400 $\mu\text{g}/\text{m}^3$ misurata su tre ore consecutive

Tabella 3.1: Parametri normativi NO<sub>2</sub>

In Figura 3.1 viene mostrata la distribuzione dei valori giornalieri negli ultimi due anni (2019 e 2020) in tutte le centraline ARPA e, nello stesso intervallo di tempo, viene mostrato l'andamento della concentrazione nell'aria in una centralina campione nella città di Milano.

Dal grafico di sinistra si può notare che mediamente le concentrazioni rimane al di sotto del limite annuale, senza mai superare il limite orario e senza mai raggiungere la soglia d'allarme. Più interessante è sicuramente capire come l'inquinante si distribuisce sul territorio e quali siano le zone maggiormente inquinate. Le risposte a queste domande vengono meglio mostrate nei capitoli seguenti.

Dal grafico di destra, invece, si può notare come la concentrazione tenda ad alzarsi nei mesi freddi e ad essere minima nei mesi più caldi. Questo è facilmente motivabile da un maggiore utilizzo di processi di combustione per il riscaldamento e ad un maggiore utilizzo dei mezzi di trasporto.

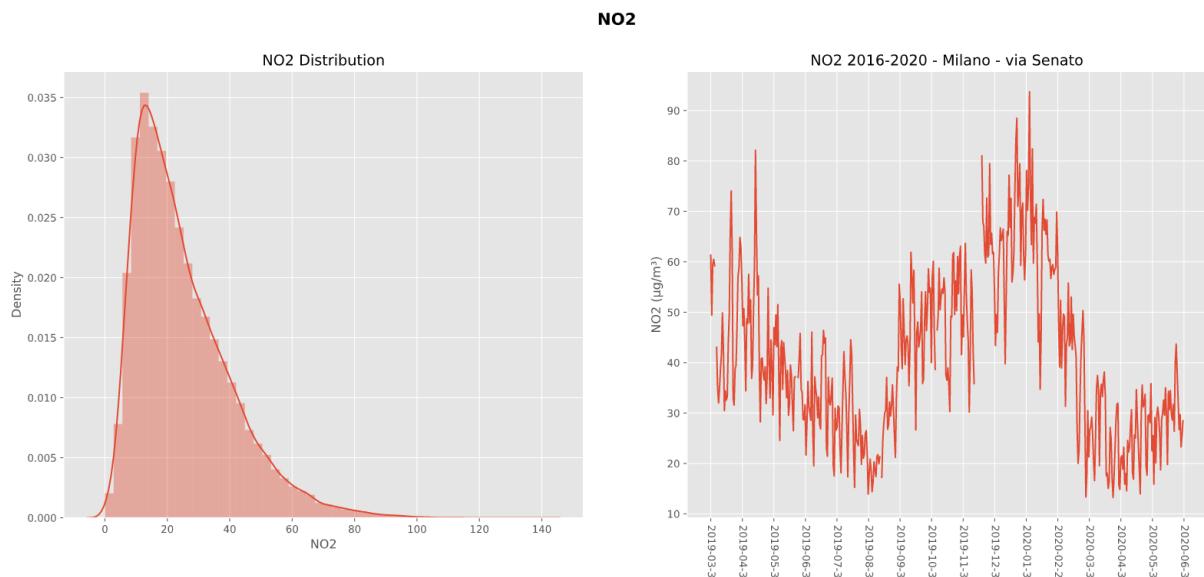


Figura 3.1: NO<sub>2</sub>, distribuzione e serie storica

## 3.2 NO<sub>x</sub> : ossidi di azoto

NO<sub>x</sub> è una sigla generica che identifica collettivamente tutti gli ossidi di azoto e le loro miscele. L'azoto, combinandosi con l'ossigeno, dà luogo a diversi composti denominati NO<sub>x</sub>. In particolare, è in grado di formare i seguenti ossidi:

- monossido di azoto (NO)
- diossido di azoto o ipoazotide (NO<sub>2</sub>, anche come dimero N<sub>2</sub>O<sub>4</sub>)
- ossido di diazoto o protossido di azoto (N<sub>2</sub>O)
- triossido di diazoto o anidride nitrosa (N<sub>2</sub>O<sub>3</sub>)
- pentossido di diazoto o anidride nitrica (N<sub>2</sub>O<sub>5</sub>)

Tra questi, i più rilevanti dal punto di vista dell'inquinamento atmosferico sono l'ossido di azoto NO ed il biossido di azoto NO<sub>2</sub>. L'ossido di azoto è un gas incolore ed inodore ed ha origine in qualsiasi processo di combustione in cui l'aria è impiegata come comburente, per reazione tra ossigeno e azoto ad alte temperature. Circa il 10% dell'NO, una volta immesso in atmosfera, viene trasformato in biossido di azoto per azione della radiazione solare. In condizioni di forte irraggiamento, inoltre, gli ossidi di azoto partecipano a reazioni fotochimiche che originano inquinanti secondari (ozono, smog fotochimico). Gli ossidi di azoto contribuiscono anche alla formazione delle piogge acide e favoriscono l'accumulo di nitrati al suolo che, a loro volta, possono alterare significativamente gli equilibri ecologici ambientali.

I composti della famiglia NO<sub>x</sub> sono prodotti in natura da fulmini e dalle eruzioni vulcaniche. Una delle principali fonti antropiche sono i motori a combustione interna, le elevate temperature raggiunte durante la combustione del carburante contribuiscono alla formazione di ossidi di azoto. Infine, altra fonte di produzione sono i fornelli a gas delle cucine e le caldaie di riscaldamento per acqua e/o ambiente.

### Parametri normativi correnti

Inquinante	Tipo di limite	Limite
NO <sub>x</sub>	Media annuale delle medie orarie	30 $\mu\text{g}/\text{m}^3$

Tabella 3.2: Parametri normativi NO<sub>x</sub>

In Figura 3.2 viene mostrata la distribuzione dei valori giornalieri negli ultimi due anni (2019 e 2020) in tutte le centraline ARPA e, nello stesso intervallo di tempo, viene mostrato l'andamento della concentrazione in atmosfera in una specifica centralina campione

nella città di Milano.

Rispetto al caso precedente, è meno evidente che il valor medio sia inferiore alla media annuale. Si può inoltre notare che, nel corso dei due anni, sono stati rilevati valori particolarmente alti di inquinante.

Analogamente al caso precedente, si nota un incremento della concentrazione nei periodi freddi dell'anno, imputabile al maggiore utilizzo di processi di combustione.

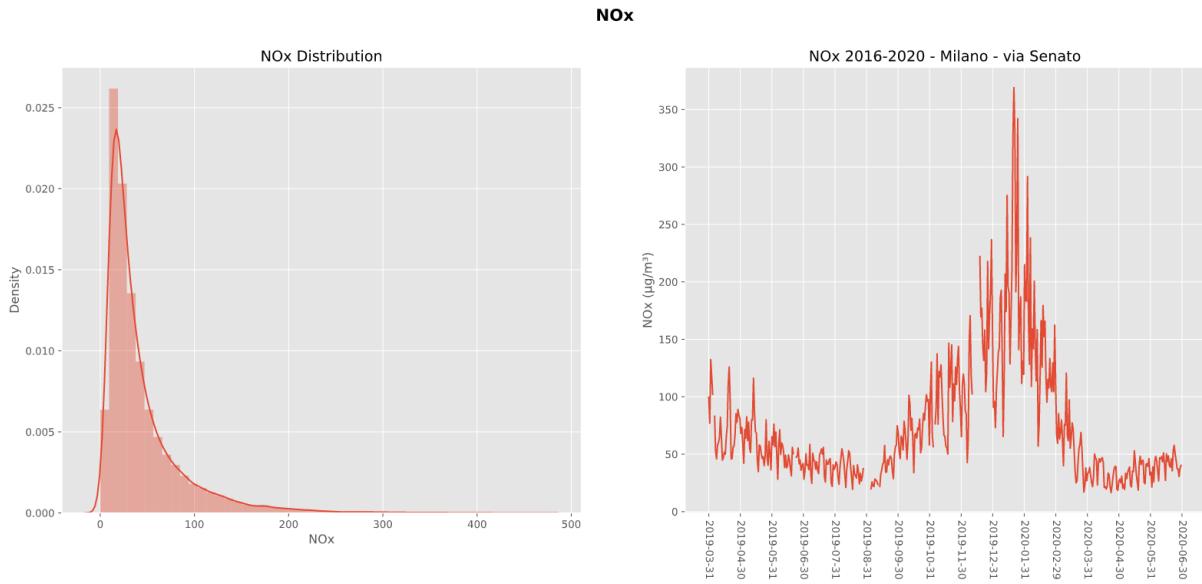


Figura 3.2: NO<sub>x</sub>, distribuzione e serie storica

### 3.3 PM<sub>10</sub>

L'acronimo PM [5] deriva dal termine inglese “*particulate matter*” (materiale particolato) e viene utilizzato per indicare le polveri sottili (o pulviscolo), nonché insieme di particelle microscopiche, solide e liquide, di diversa natura e composizione chimica, che si trovano in sospensione nell'aria che respiriamo. A seconda della loro dimensione convenzionalmente si distinguono in PM<sub>10</sub>, PM<sub>2.5</sub> e PM<sub>1</sub> dove il numero dopo la sigla PM indica la grandezza del diametro della particella. Nello specifico il termine PM<sub>10</sub> individua una delle frazioni in cui viene classificato il particolato atmosferico, quelle il cui diametro aerodinamico (corrispondente al diametro di una sfera di densità uguale a 1 g/cm<sup>3</sup> che cade nell'aria con la stessa velocità della particella considerata) è compreso tra i 10  $\mu$  e i 2,5  $\mu$  dove 1  $\mu$ , o micron, corrisponde ad un millesimo di millimetro.

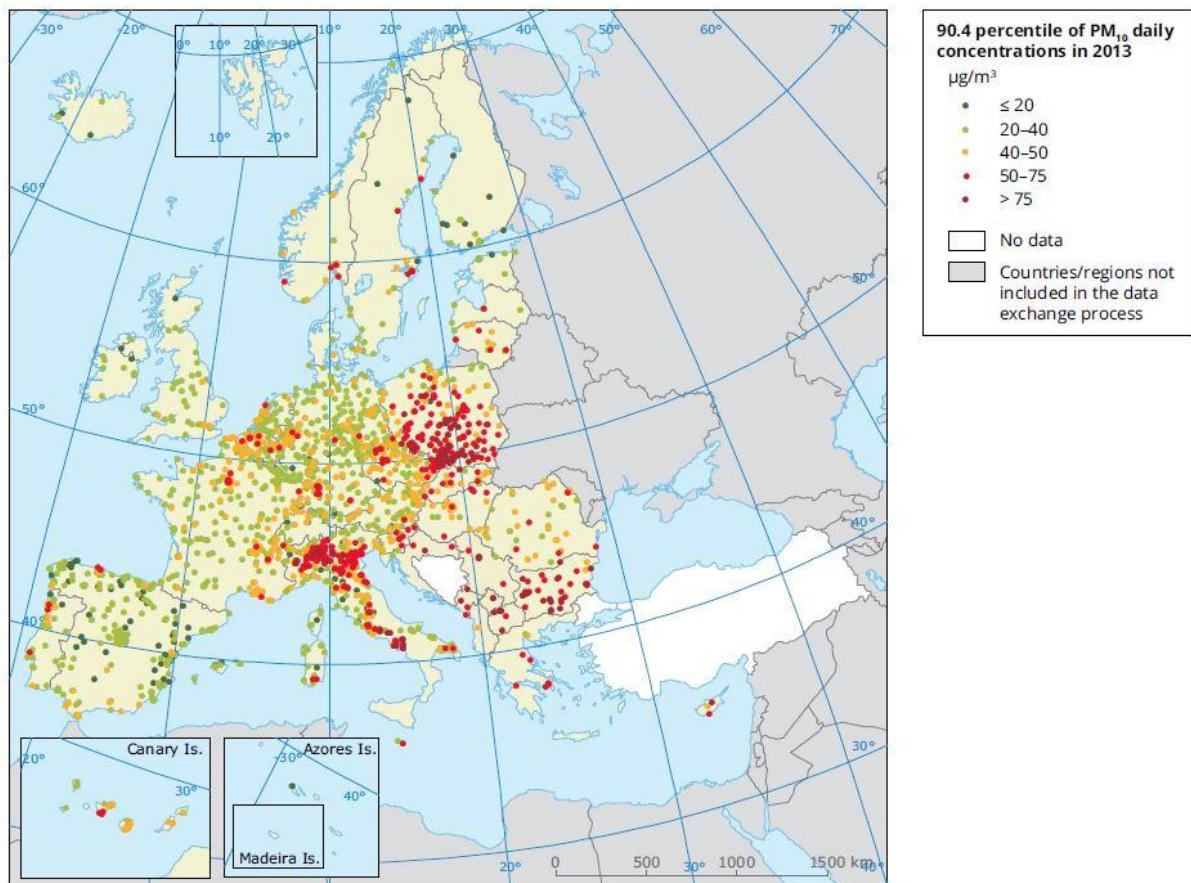


Figura 3.3: Concentrazioni PM<sub>10</sub> in Europa, 2013

### Parametri normativi correnti

Inquinante	Tipo di limite	Limite
PM <sub>10</sub>	Limite giornaliero	50 µg/m <sup>3</sup> (non oltre >35 giorni l'anno)
PM <sub>10</sub>	Limite annuale	40 µg/m <sup>3</sup> media annua

Tabella 3.3: Parametri normativi PM<sub>10</sub>

Secondo un articolo<sup>1</sup> apparso sul *Il Sole 24 Ore* il 21 dicembre 2018 a cura di Riccardo Saporiti, non sono le auto o in generale i mezzi di trasporto su gomma le principali responsabili della presenza di PM<sub>10</sub> nell'aria delle città italiane, bensì gli impianti di riscaldamento, tra i quali annoveriamo caldaie, stufe e caminetti.

<sup>1</sup>Il Sole 24 Ore (https://www.infodata.ilsole24ore.com/2018/12/21/pm10-le-caldaie-le-principali-responsabili-dell'inquinamento/)

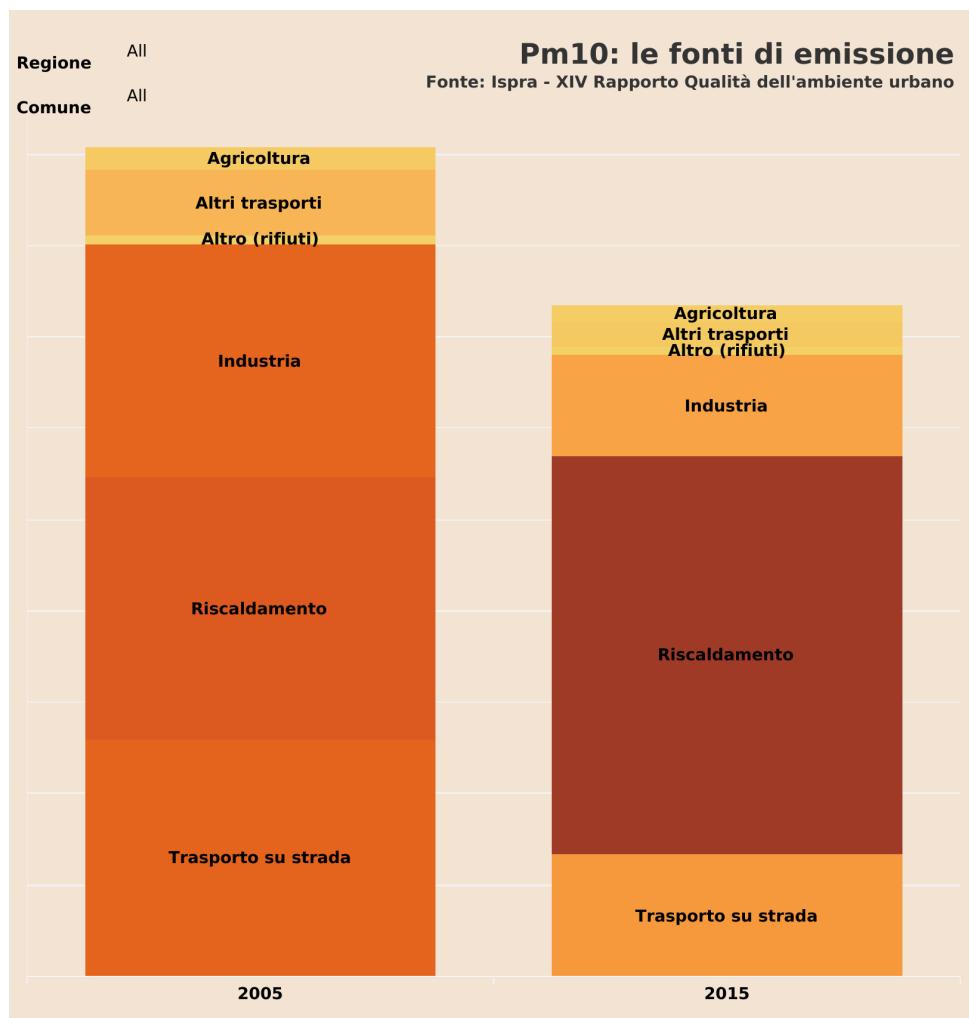


Figura 3.4: Le fonti di emissioni di PM<sub>10</sub>, 2005-2015

In Figura 3.5 viene mostrata la distribuzione dei valori giornalieri negli ultimi due anni (2019 e 2020) in tutte le centraline ARPA e, nello stesso intervallo di tempo, viene mostrato l'andamento della concentrazione nell'aria in una centralina campione nella città di Milano.

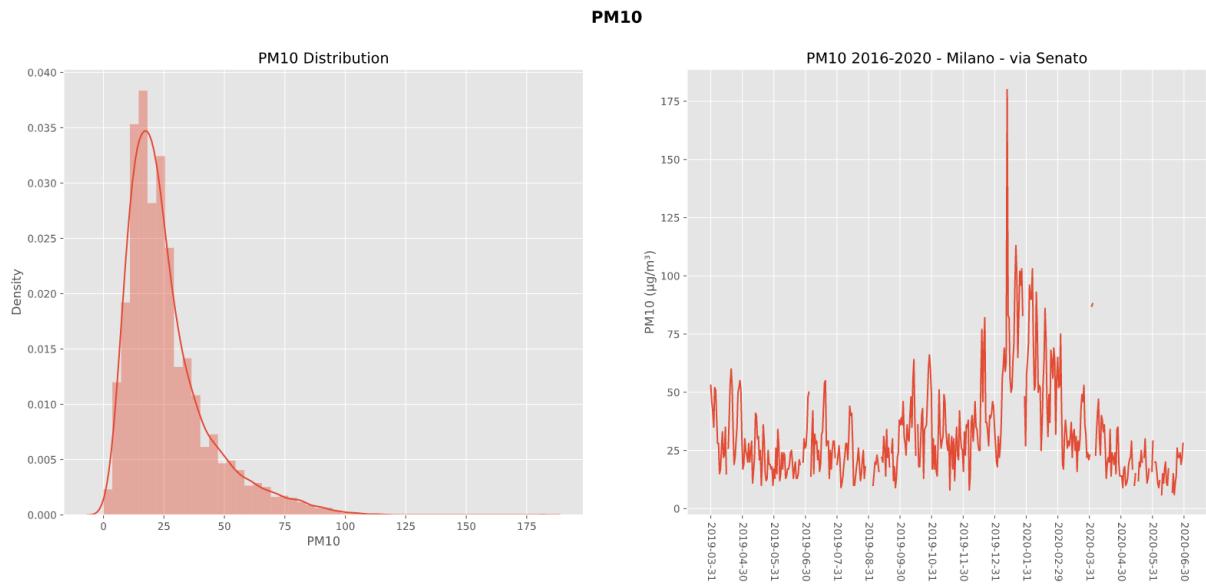


Figura 3.5:  $\text{PM}_{10}$ , distribuzione e serie storica

### 3.4 $\text{PM}_{2.5}$

Con  $\text{PM}_{2.5}$  si identificano quelle particelle il cui diametro (corrispondente al diametro di una sferetta di densità uguale a  $1 \text{ g}/\text{m}^3$  che cade nell'aria con la stessa velocità della particella considerata) è inferiore o uguale ai  $2.5 \text{ micron}$  dove  $1 \text{ micron} (\mu)$  corrisponde ad un millesimo di millimetro. Il particolato  $\text{PM}_{2.5}$  è chiamato anche genericamente “particolato fine” in contrapposizione a “particolato grossolano” che invece indica quelle particelle sospese il cui diametro aerodinamico è minore di  $2.5 \mu$  o, all'interno della frazione del  $\text{PM}_{10}$ .

Le fonti di emissione di  $\text{PM}_{2.5}$  posso essere ricondotte alle medesime del  $\text{PM}_{10}$  precedentemente illustrate. In particolare, le polveri di dimensione inferiore a  $2.5\mu \text{ m}$  hanno origine sia naturale che antropica; quest'ultima fonte copre più del 90% del totale nel 2015<sup>2</sup>.

I principali settori emissivi da cui si originano le emissioni di particolato in atmosfera sono:

- il settore della combustione non industriale (riscaldamento nel settore commerciale, istituzionale e residenziale).
- il settore del trasporto su strada.
- i settori della combustione nell'industria manifatturiera e dei processi industriali.

---

<sup>2</sup>Ispra Ambiente (<https://annuario.isprambiente.it/pon/basic/>)

## Parametri normativi correnti

Inquinante	Tipo di limite	Limite
PM <sub>2.5</sub>	Limite annuale	25 $\mu\text{g}/\text{m}^3$ media annua (dal 2015)

Tabella 3.4: Parametri normativi PM<sub>2.5</sub>

In Figura 3.6 viene mostrata la distribuzione dei valori giornalieri negli ultimi due anni (2019 e 2020) in tutte le centraline ARPA e, nello stesso intervallo di tempo, viene mostrato l'andamento della concentrazione nell'aria in una centralina campione nella città di Milano.

Nell'immagine di sinistra si può notare che i valori di concentrazione tendono ad essere maggiori del valore limite annuale. Analogamente agli altri inquinanti, la concentrazione aumenta nei periodi invernali.

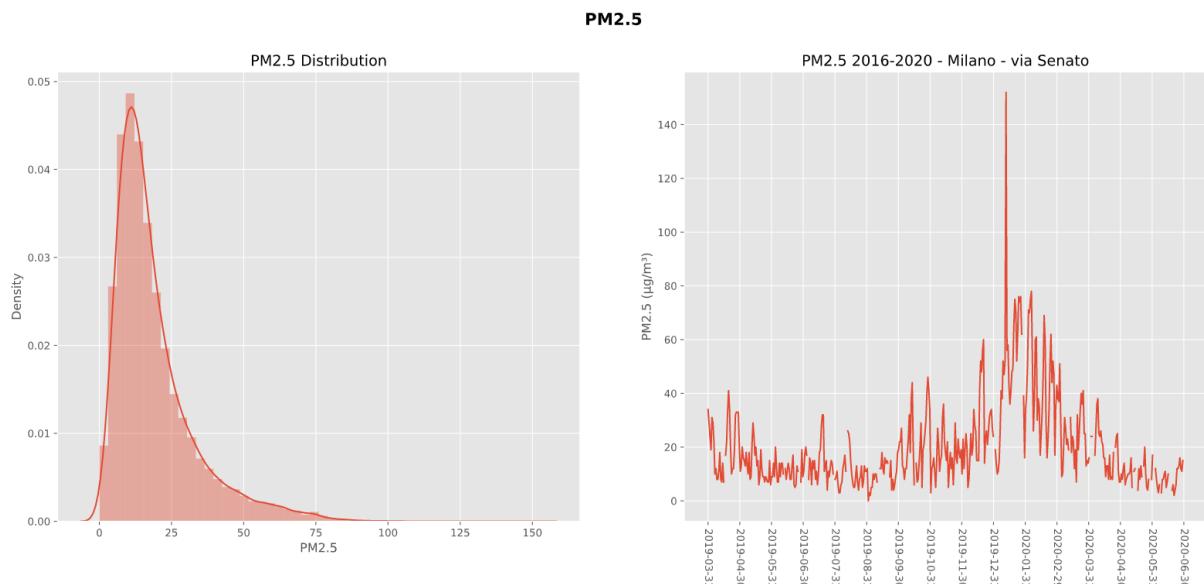


Figura 3.6: PM<sub>2.5</sub>, distribuzione e serie storica

## 3.5 Correlazioni tra inquinanti

Le proprietà sopra elencate si traducono in una forte correlazione tra i vari inquinanti. I quattro inquinanti condividono cause comuni e il grafico sottostante illustra chiaramente questo tipo di relazione.

I grafici mostrano la correlazione di inquinante. Tanto più la distribuzione dei valori si avvicina alla bisettrice del quadrante tanto più i due inquinanti sono correlati tra

loro. Inoltre se la retta è crescente la correlazione è positiva, mentre se è decrescente la correlazione è negativa.

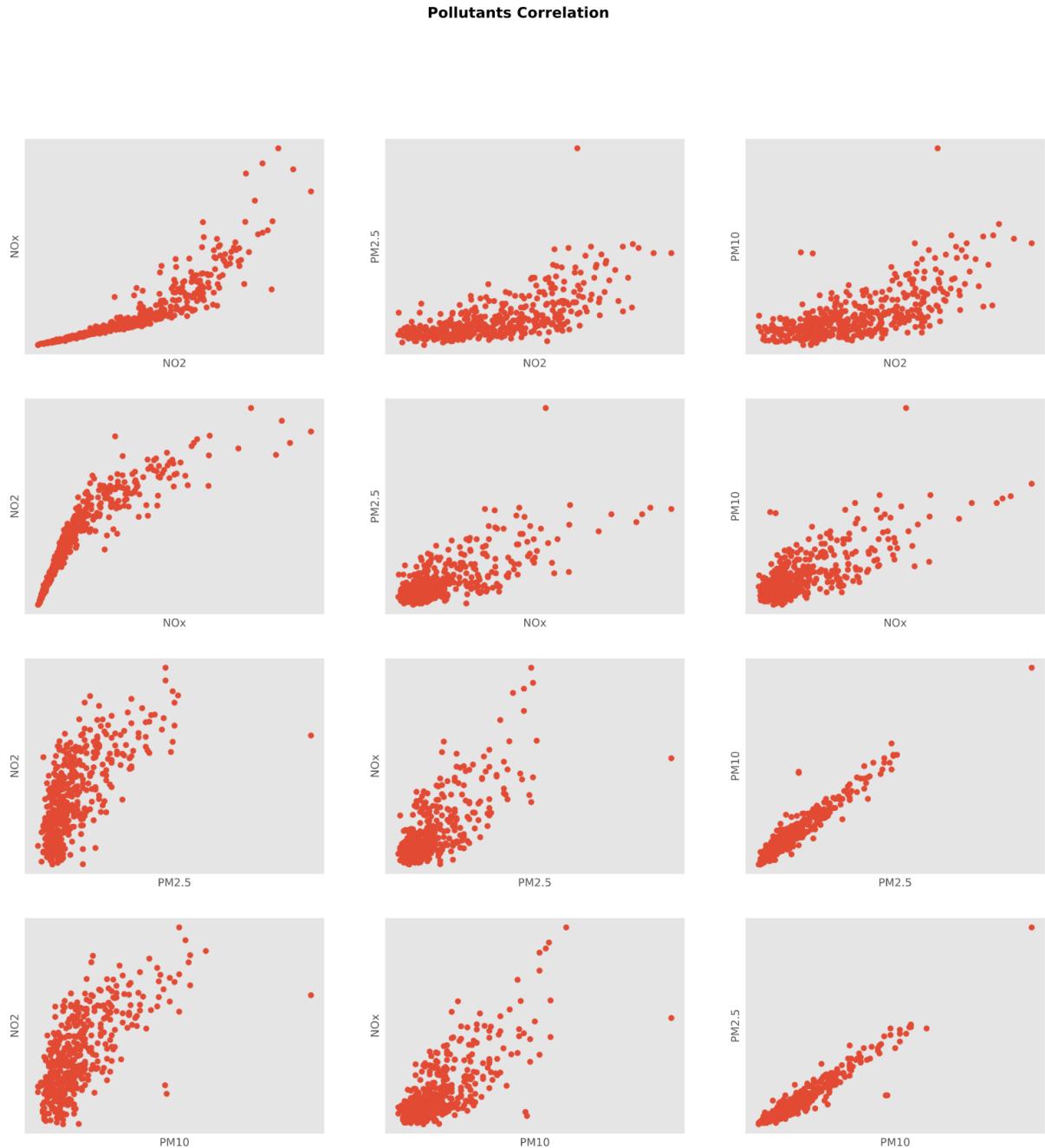


Figura 3.7: Correlazioni tra inquinanti

Si può notare che la correlazione tra  $\text{PM}_{2.5}$  e  $\text{PM}_{10}$  è molto forte dato che i punti si dispongono sulla bisettrice del quadrante, non è un caso visto che sono due inquinanti della stessa famiglia e ciò che cambia è solo la dimensione delle particelle.

Discorso simile si può fare tra  $\text{NO}_x$  e  $\text{NO}_2$ , anche se i punti non vanno a coincidere perfettamente sulla bisettrice.

In tutti i quadranti si può notare che c'è una relazione positiva, prova del fatto che esiste una correlazione positiva tra tutte le coppie di inquinanti, all'aumentare di uno, aumenta anche l'altro.

# Capitolo 4

## Analisi dell'impatto del lockdown

In questo capitolo, dopo una breve introduzione ai dati di base usati nel progetto, si analizzeranno uno ad uno gli inquinanti focalizzando l'attenzione sulle concentrazioni nel 2019, nel 2020 e la differenza tra i due anni per evidenziare se esiste o meno un impatto dovuto al lockdown.

Ogni file usato è espresso nel sistema di coordinate proiettato UTM32N, riferito al sistema di riferimento geodetico WGS84 (codice EPSG: 4326).

### 4.1 Dati di base

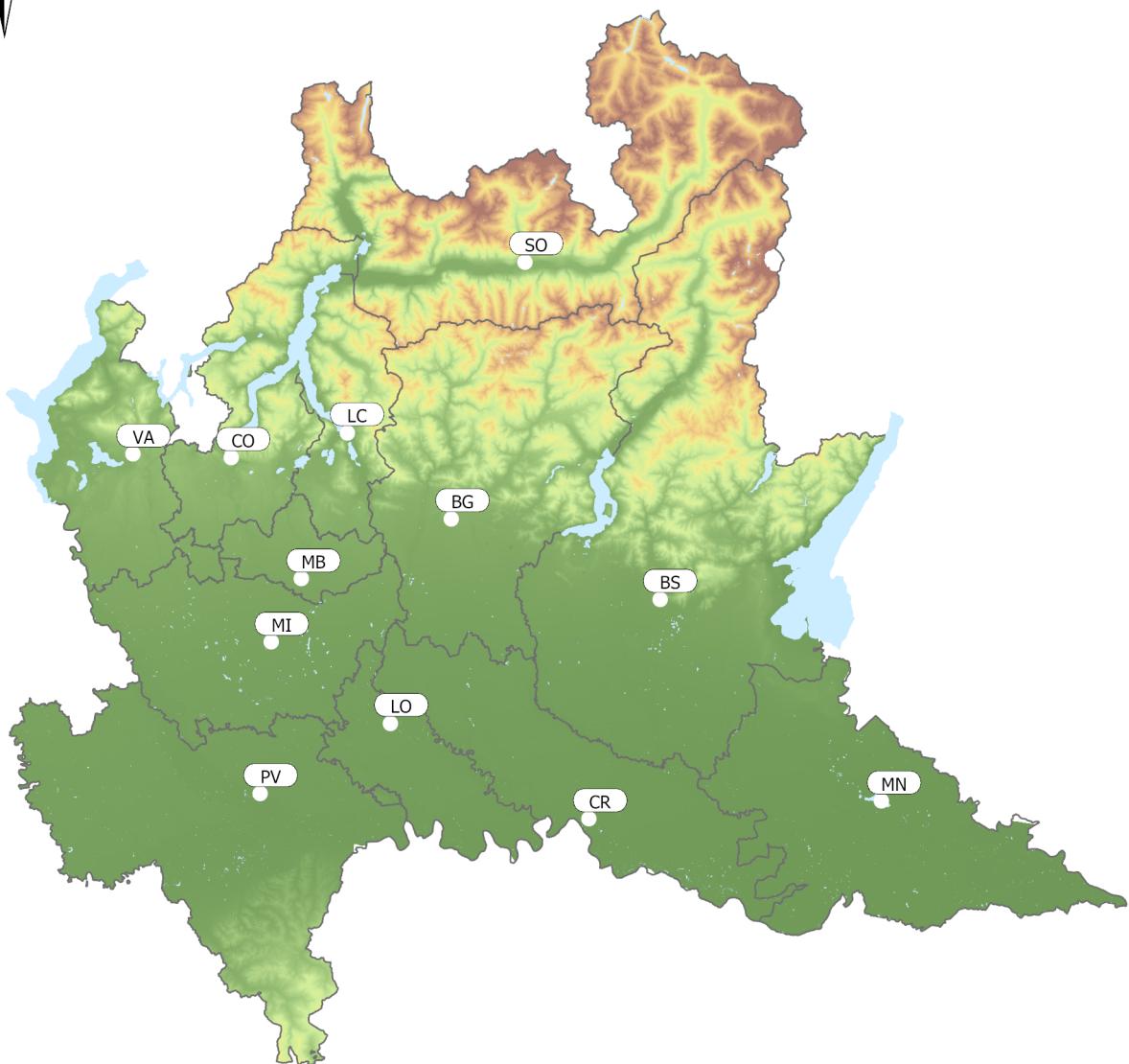
In ArcGIS Pro abbiamo usato i seguenti dati di base con le relative fonti che verranno usati man mano nell'analisi:

- DTM 20 - ESRI GRID [6]: modello digitale del terreno rappresentato da un DTM con struttura a griglia con passo di campionamento pari a 20 metri. La copertura comprende le zone collinari e montane della Regione Lombardia. Il GRID è un file raster sviluppato da Esri e rappresenta lo spazio geografico come una matrice di punti a griglia quadrata di uguale dimensione composta da righe e colonne. Ogni punto della griglia memorizza un valore numerico che rappresenta un attributo geografico (ad esempio elevazione o pendenza della superficie). Ogni cella della griglia è georeferenziata.
- Lago 10000 CT10 [7]: shapefile contenuto nel gruppo "*Carta Tecnica Regionale (CTR) 10.000 Vettoriale alla scala 1:10.000 (CT10)*" che è stata costruita a partire dalla digitalizzazione della CTR al tratto in scala 1:10.000.
- Limiti amministrativi [8]: shapefile contenuto nel gruppo della base dati di sintesi in scala 1:250.000. La base dati di sintesi è costituita da una serie di cartografie numeriche in formato vettoriale, realizzate per costituire una base informativa ter-

ritoriale utile per gli interventi di analisi e di programmazione generale interessanti l'intero territorio regionale.

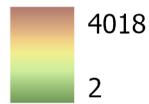
- Capoluoghi: si è costruito un file .csv con i nomi e le coordinate dei capoluoghi di provincia della Lombardia.

Usando questi di dati di base combinati si è ottenuta una mappa di base (vedi pagina seguente) su cui partire per fare tutte le analisi degli inquinanti e regressori.



## Lombardia DTM

Altitudine (m)



Laghi

Laghi

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# Dati di base

## 4.2 Dati degli inquinanti

Come già descritto nelle fonti di dati (vedi [sottosezione 1.2.2](#)) abbiamo a disposizione 84 sensori per l' $\text{NO}_2$ , 83 per l' $\text{NO}_x$ , 64 per il  $\text{PM}_{10}$  e 32 per il  $\text{PM}_{2.5}$  come mostrato in [Figura 4.1](#). Possiamo notare che le centraline per la rilevazione del  $\text{PM}_{2.5}$  sono solo 32, per esempio nelle provincie di Como, Sondrio e Lodi vi è solo una centralina per l'intera provincia. In ogni caso, tutte le province dispongono di almeno una centralina per la rilevazione di ognuna delle 4 tipologie di sensori. I mesi campionati sono gli stessi sia per il 2019 che per il 2020 e vanno da Marzo a Luglio (vedi [Appendice A](#)).

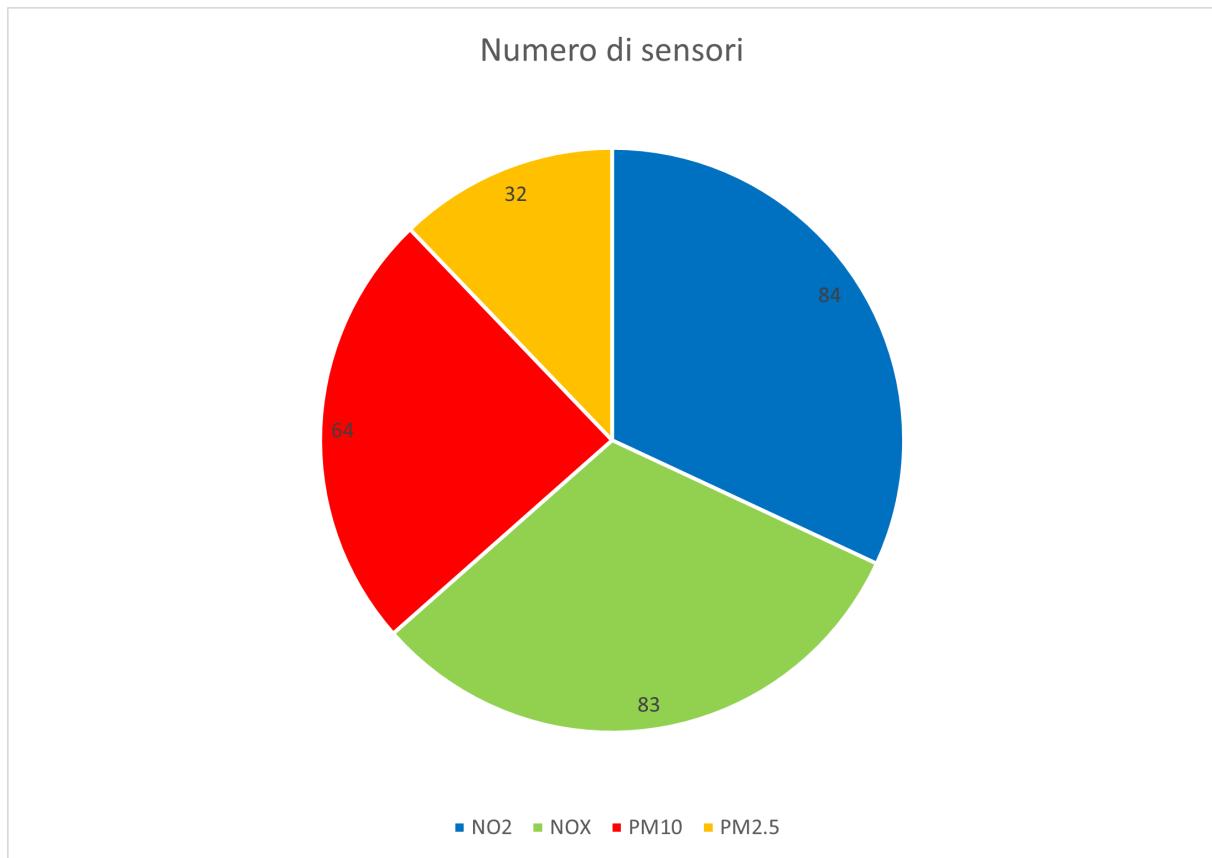


Figura 4.1: Numero di sensori nel territorio lombardo

## 4.3 NO<sub>2</sub>

### 4.3.1 Concentrazione media nel 2019

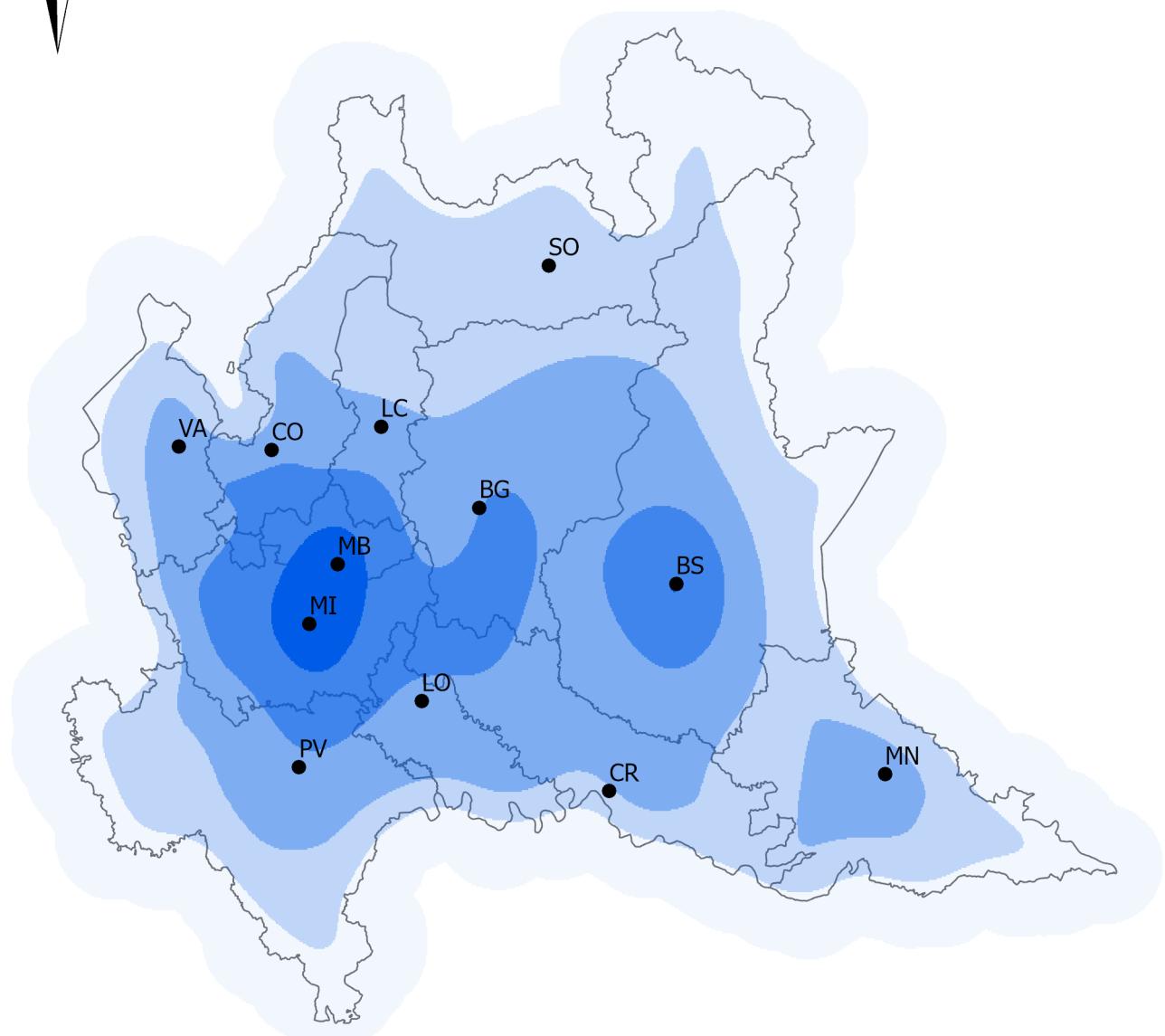
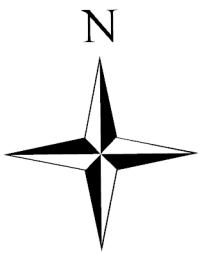
Nei mesi campionati che vanno dal Marzo 2019 a Luglio 2019, vi è un'elevata concentrazione di NO<sub>2</sub> soprattutto nell'area urbana di Milano e dintorni, comprendendo anche in parte la Monza e Brianza. Vi è inoltre una elevata concentrazione nelle zone della Bassa Bergamasca e nei dintorni di Brescia. Nella provincia di Sondrio si rileva bassa concentrazione.

La mappa seguente mostra la distribuzione della concentrazione media di NO<sub>2</sub> sul territorio lombardo nel periodo di interesse dell'anno 2019. Il risultato è ottenibile tramite la funzione **Heatmap** di ArcGIS Pro. Tuttavia, sono stati presi piccoli accorgimenti ed accortezze: in primo luogo il numero di valori (circa 80, uno per ogni stazione) era insufficiente per coprire completamente il territorio e l'analisi sarebbe risultata di poco impatto. Per risolvere questo problema, è stato realizzato uno script Python (vedi [Appendice B](#)) che adottasse particolari tecniche di stima in grado di creare una vera e propria mappa di colore riempiendo le porzioni di territorio non coperte da centraline con valori approssimati. Sono stati quindi stimati i dati in una superficie rettangolare che contenesse completamente l'intera regione, applicando in seguito la funzione di **Clip** per limitarne la rappresentazione al solo territorio lombardo.

In secondo luogo, sono stati normalizzati i valori, in modo da far corrispondere al punto di maggior concentrazione il colore più scuro, con conseguente sbiancamento al diminuire della concentrazione stessa.

Riassumendo, è stato applicato un profilo di colore discreto in funzione del valore di concentrazione stimato in ogni punto.

La legenda sul grafico riporta i valori di minimo e massimo dei dati effettivamente rilevati dalle stazioni ARPA, con buona approssimazione posti in relazione ai valori stimati.



NO<sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )



5,26

62,50

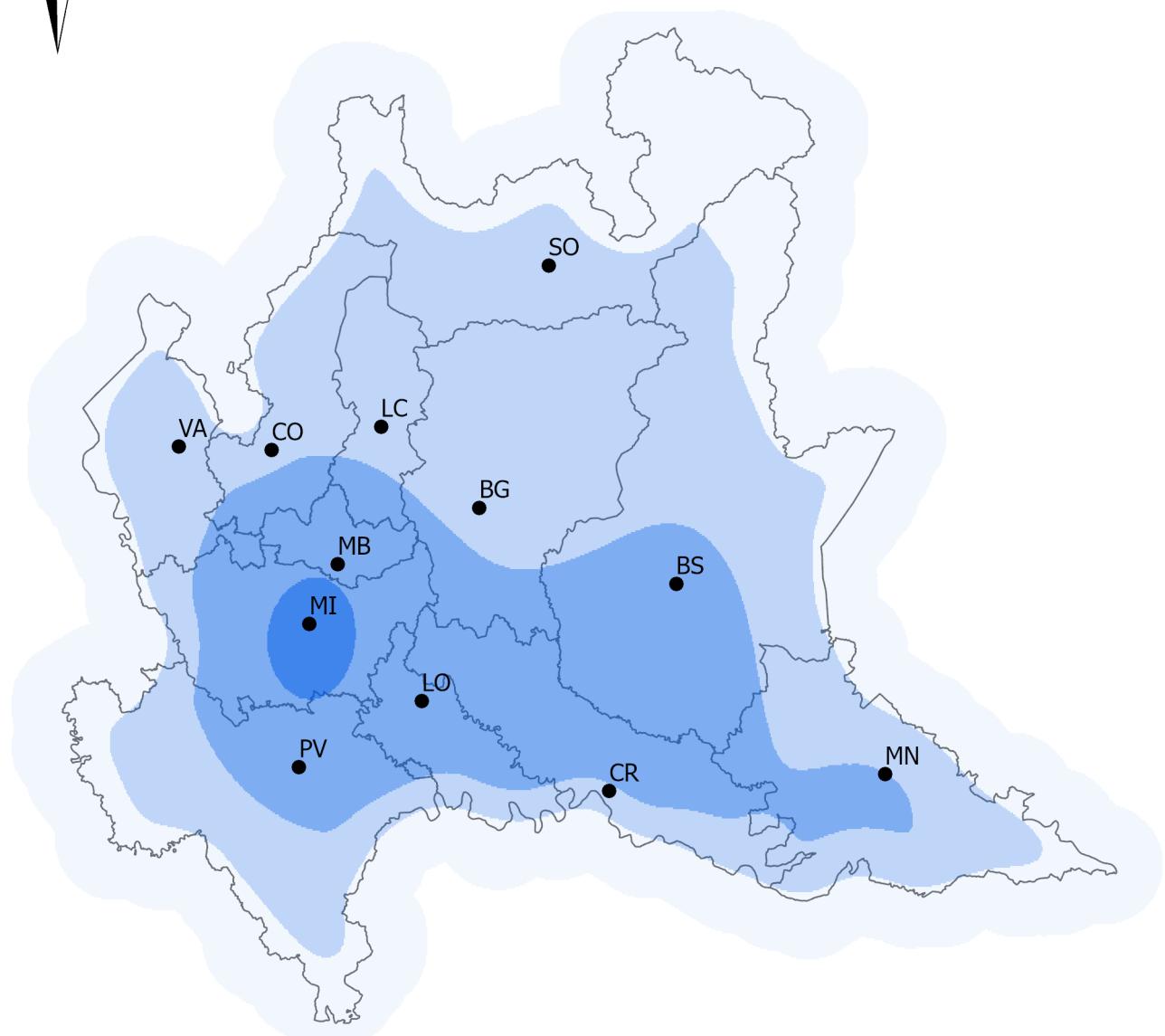
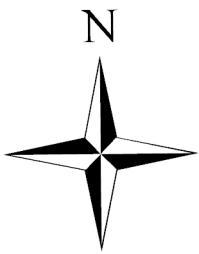
30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>2</sub>: Concentrazioni 2019

### **4.3.2 Concentrazione media nel 2020**

La mappa seguente è ottenuta con lo stesso procedimento prima citato e riporta la concentrazione di NO<sub>2</sub>. Anche in questo caso quindi, a partire dai dati ARPA rilevati per nell'anno 2020 sono stati stimati i dati in ogni altro punto mancante. I dati vengono normalizzati rispetto al valore di concentrazione massimo tra 2019 e 2020. In questo modo si può percepire una distinzione qualitativa direttamente dall'immagine. Si può notare che la scala in legenda sia uguale all'immagine precedente e che nessuna zona della mappa risulti essere di colore blu scuro, ovvero del quinto livello di colore. Questo si traduce in valori di concentrazione inferiori rispetto all'anno precedente. È molto evidente che vi sia stata una riduzione di livelli di concentrazione in modo omogeneo in tutta la Lombardia. È opportuno ricordare che l'analisi si limita ad un approccio qualitativo, basandosi su dati approssimati da stima.



NO<sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )



5,26

62,50

30 15 0 30 60 90 120 Kilometers

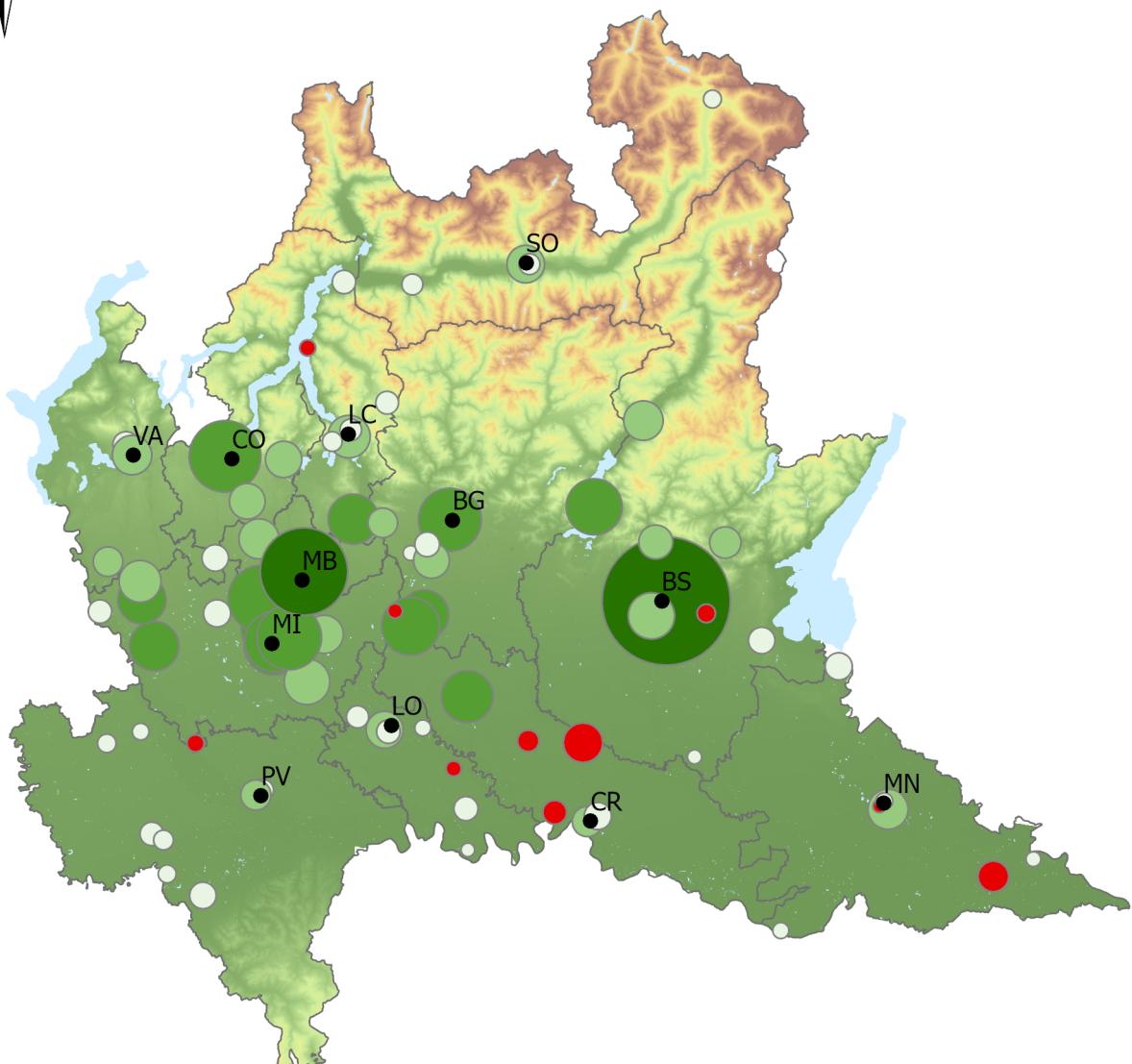
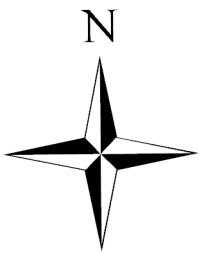
Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>2</sub>: Concentrazioni 2020

### 4.3.3 Impatto del periodo di lockdown

L'effetto del Lockdown in termini di riduzione di NO<sub>2</sub> è stato significativo. La mappa seguente vuol mostrare quali siano state le zone soggette e quale la tipologia di cambiamento. L'analisi, pur rimanendo qualitativa, assume un maggior carattere quantitativo legato esclusivamente ai dati veri, rilevati.

Attraverso una rielaborazione in Python sono state calcolate le medie nel periodo di interesse. Come si può osservare, la *symbology* della mappa è stata relazionata alla differenza di concentrazione. La dimensione dei cerchi è in funzione del valore assoluto di differenza di concentrazione, mentre il colore è in funzione dell'aumento o della diminuzione di concentrazione e della relativa intensità. La differenza è calcolata come media del 2020 sottratto alla media del 2019, così che valori positivi traducono diminuzione di concentrazione e quindi un cambiamento positivo, delineato dal colore verde. La zona che ne ha maggiormente beneficiato è l'area bresciana, così come Monza e Brianza e Como. Il motivo è da ricercarsi dalla diminuzione di traffico veicolare che è la maggior fonte di emissione di NO<sub>2</sub>. Da segnalare che nelle province di Mantova e Cremona si rilevano dei piccoli aumenti di NO<sub>2</sub> in contrasto alle aspettative. Quasi irrilevante la variazione di NO<sub>2</sub> a Sondrio.



#### Differenza di concentrazione

- -7,49 - -0,08 (11)
- -0,08 - 4,08 (32)
- 4,08 - 9,26 (24)
- 9,26 - 16,28 (15)
- 16,28 - 31,90 (2)

30 15 0 30 60 90 120  
Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>2</sub>: Impatto Lockdown

## 4.4 $\text{NO}_x$

Le successive tre mappe sono analoghe a quanto presentato per  $\text{NO}_2$ : i dati sono soggetti alle stesse elaborazioni ed accortezze e le mappe sono ottenute seguendo la medesima procedura.

### 4.4.1 Concentrazione media nel 2019

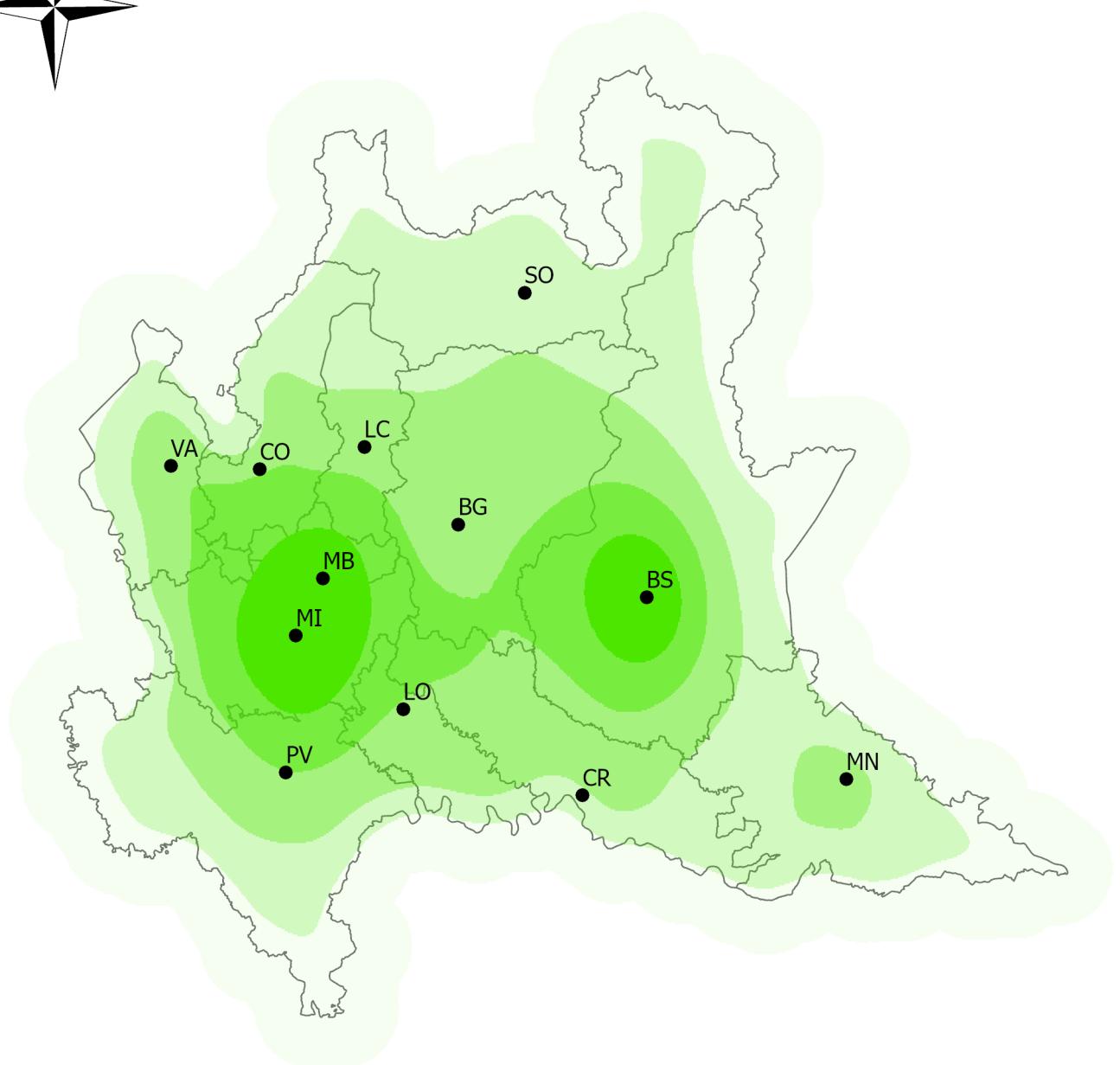
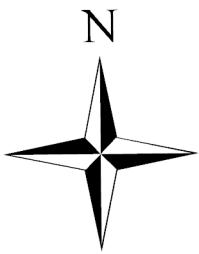
Le zone a maggior concentrazione sono 3: Milano, Monza e Brianza e Brescia. Il valore in queste zone raggiunge i  $126,22 \mu\text{g}/\text{m}^3$ . Ritroviamo come da aspettativa una situazione molto coerente con l'andamento dell'  $\text{NO}_2$  nel 2019 con un valore nettamente maggiore solo a Brescia. Da notare come i picchi di  $\text{NO}_2$  raggiungevano nel 2019 i  $62,50 \mu\text{g}/\text{m}^3$  che sono precisamente la metà dei picchi di  $\text{NO}_x$ , questo ci dimostra nuovamente come  $\text{NO}_x$  sia composta al 50% da  $\text{NO}_2$ .

### 4.4.2 Concentrazione media nel 2020

Graficamente riscontriamo una sola zona che presenta maggior valori ed è quella milanese. A seguire una serie di province quali Monza e Brianza, Lodi, Pavia e Bergamo ma con valori nettamente minori.

### 4.4.3 Impatto del periodo di lockdown

Dalla seguente mappa si evince che Brescia è la provincia che più ha beneficiato dal lockdown in termini di riduzione di  $\text{NO}_x$ . Seguita da Como, Monza e Brianza e Milano. Da segnalare nella zona sud della Lombardia una tendenza all'aumento di  $\text{NO}_x$  diversamente dal resto della regione dove la tendenza alla riduzione dell'inquinante è predominante. Riscontriamo nuovamente una bassa variazione in Sondrio.



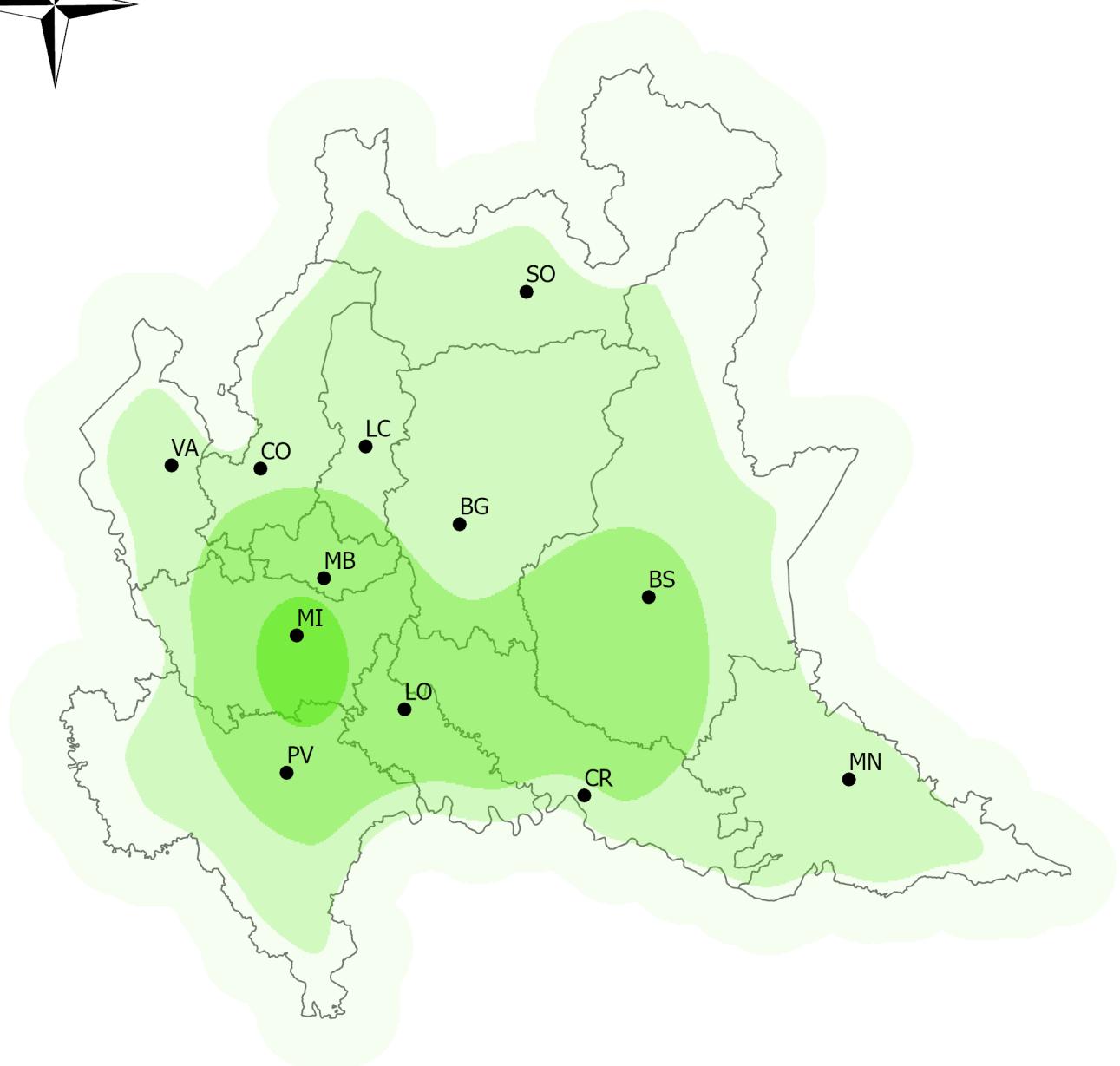
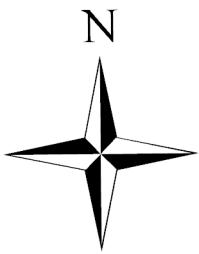
NO<sub>x</sub> ( $\mu\text{g}/\text{m}^3$ )



30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>x</sub>: Concentrazioni 2019



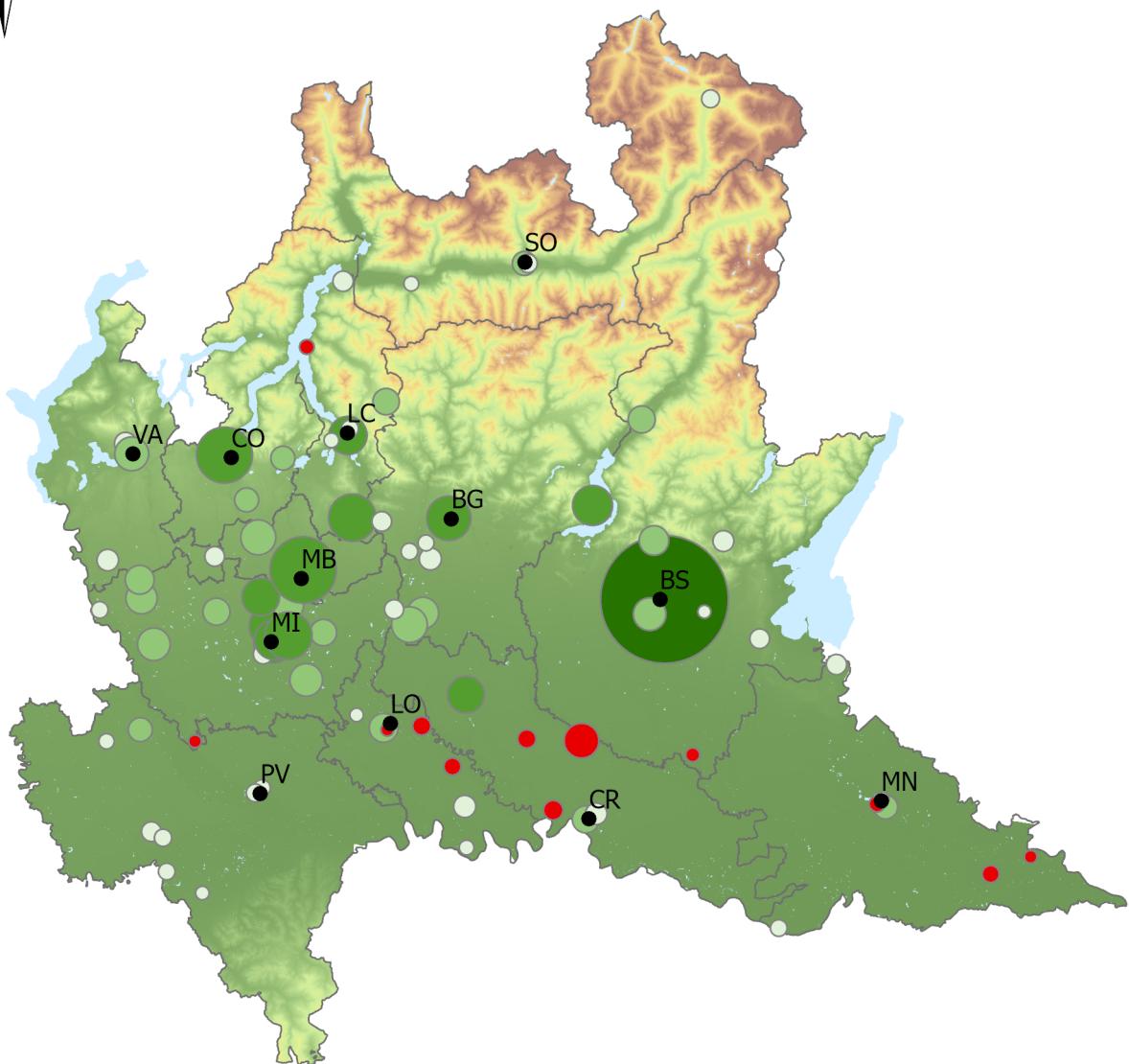
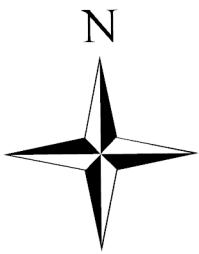
NO<sub>x</sub> ( $\mu\text{g}/\text{m}^3$ )



30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>x</sub>: Concentrazioni 2020



#### Differenza di concentrazione

- -14,18 - -0,30 (12)
- -0,30 - 6,32 (33)
- 6,32 - 14,48 (24)
- 14,48 - 33,68 (13)
- 33,68 - 71,94 (1)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NOx: Impatto Lockdown

## 4.5 PM<sub>10</sub>

### 4.5.1 Concentrazione media nel 2019

Nella mappa seguente si mostra la concentrazione media di PM<sub>10</sub> per provincia. Si ricorda che per ogni provincia è presente almeno una centralina, ma è bene tenere in considerazione che possono essere presenti province con poche centraline che generalizzano per l'intera provincia. Si è comunque preferita questa tipologia di mappa perché l'approccio con mappa di calore non è applicabile: il numero così ridotto di centraline rende qualitativamente scarsa l'informazione di interesse, con poche macchie distribuite in modo sparso su tutta la regione. Allo stesso modo, la scarsa quantità di dati non permette di adottare un approccio basato su stima come nel caso precedente. I risultati non risultavano coerenti con quanto osservato.

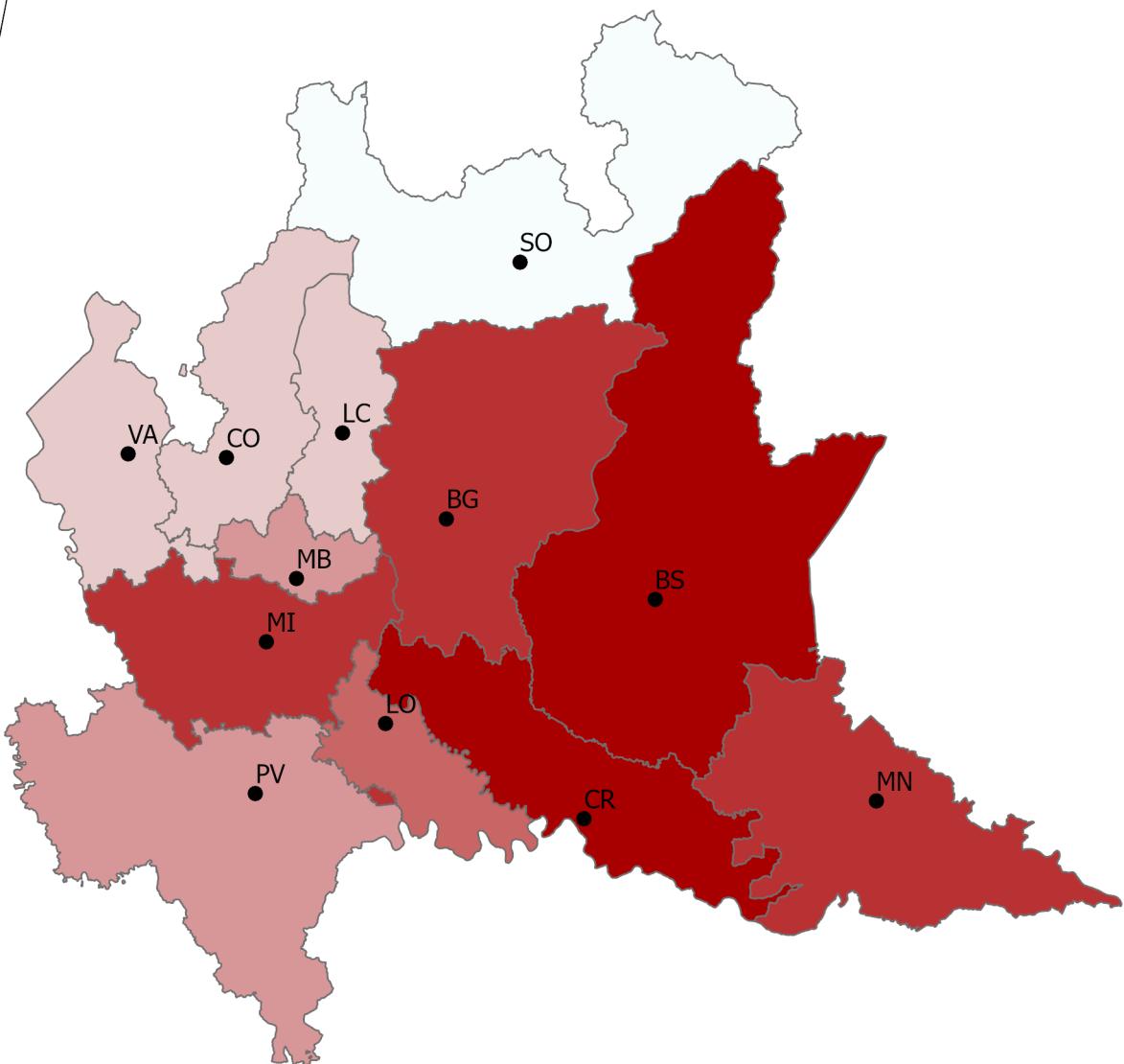
Per realizzare la mappa seguente sono stati utilizzati e intersecati due diversi file di input: una tabella contenente la concentrazione media in specifici punti georeferenziati ed uno shapefile contenente le informazioni relative alle province ed alla loro superficie. Attraverso un join spaziale punto a punto (Spatial join one-to-one, Match option: contains, Merge rule: maximum), mappando ogni centralina con la provincia nella quale è contenuta, è stata calcolata una media tra tutti i punti appartenenti alla stessa provincia.

Le province con più presenza di PM<sub>10</sub> sono Brescia e Cremona con media di 23,56  $\mu\text{g}/\text{m}^3$ . Seguite da Bergamo, Milano e Mantova. Si segnala una bassa concentrazione di PM<sub>10</sub> a Sondrio pari a 13,36  $\mu\text{g}/\text{m}^3$ , nonché la minore registrata.

### 4.5.2 Concentrazione media nel 2020

Allo stesso modo è stata realizzata la mappa relativa all'anno 2020.

In quest'anno si nota come la sola provincia di Cremona abbia concentrazione nettamente maggiore di PM<sub>10</sub> con media di 22,93  $\mu\text{g}/\text{m}^3$  e diverse zone quali Milano, Bergamo, Brescia, Lodi e Mantova con concentrazione superiore alle restanti. Sondrio resta sempre la provincia meno inquinata con medie di circa 11,50  $\mu\text{g}/\text{m}^3$ .



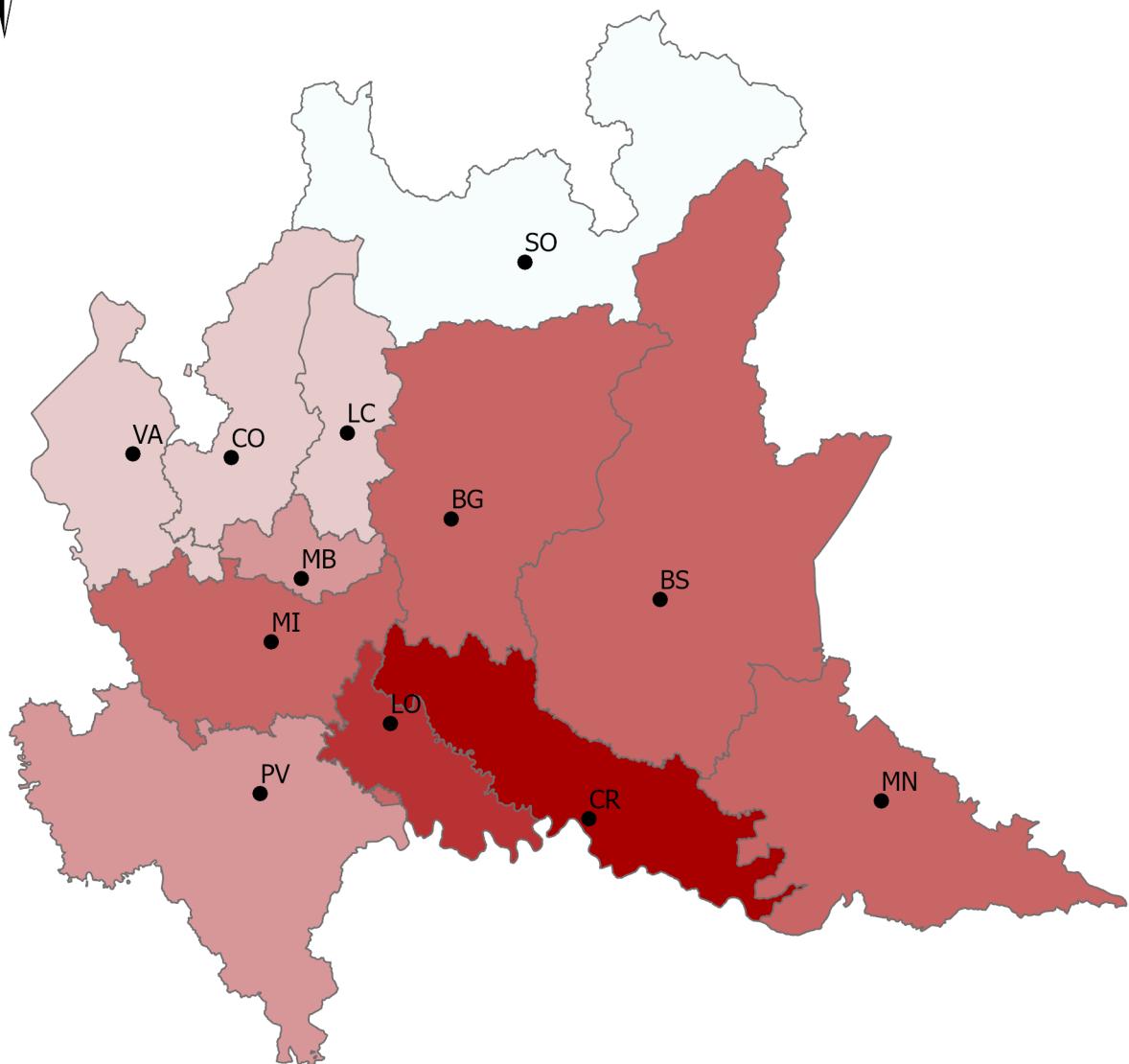
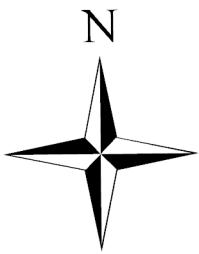
PM10 ( $\mu\text{g}/\text{m}^3$ )

- 13,36 (1)
- 13,36 - 18,05 (3)
- 18,05 - 19,31 (2)
- 19,31 - 20,37 (1)
- 20,37 - 21,42 (3)
- 21,42 - 23,56 (2)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

PM10: Concentrazione media 2019



PM10 ( $\mu\text{g}/\text{m}^3$ )

- White: 11,50 (1)
- Pale pink: 11,50 - 15,83 (3)
- Dark pink: 15,83 - 17,40 (2)
- Red: 17,40 - 18,69 (4)
- Dark red: 18,69 - 19,47 (1)
- Black: 19,47 - 22,93 (1)

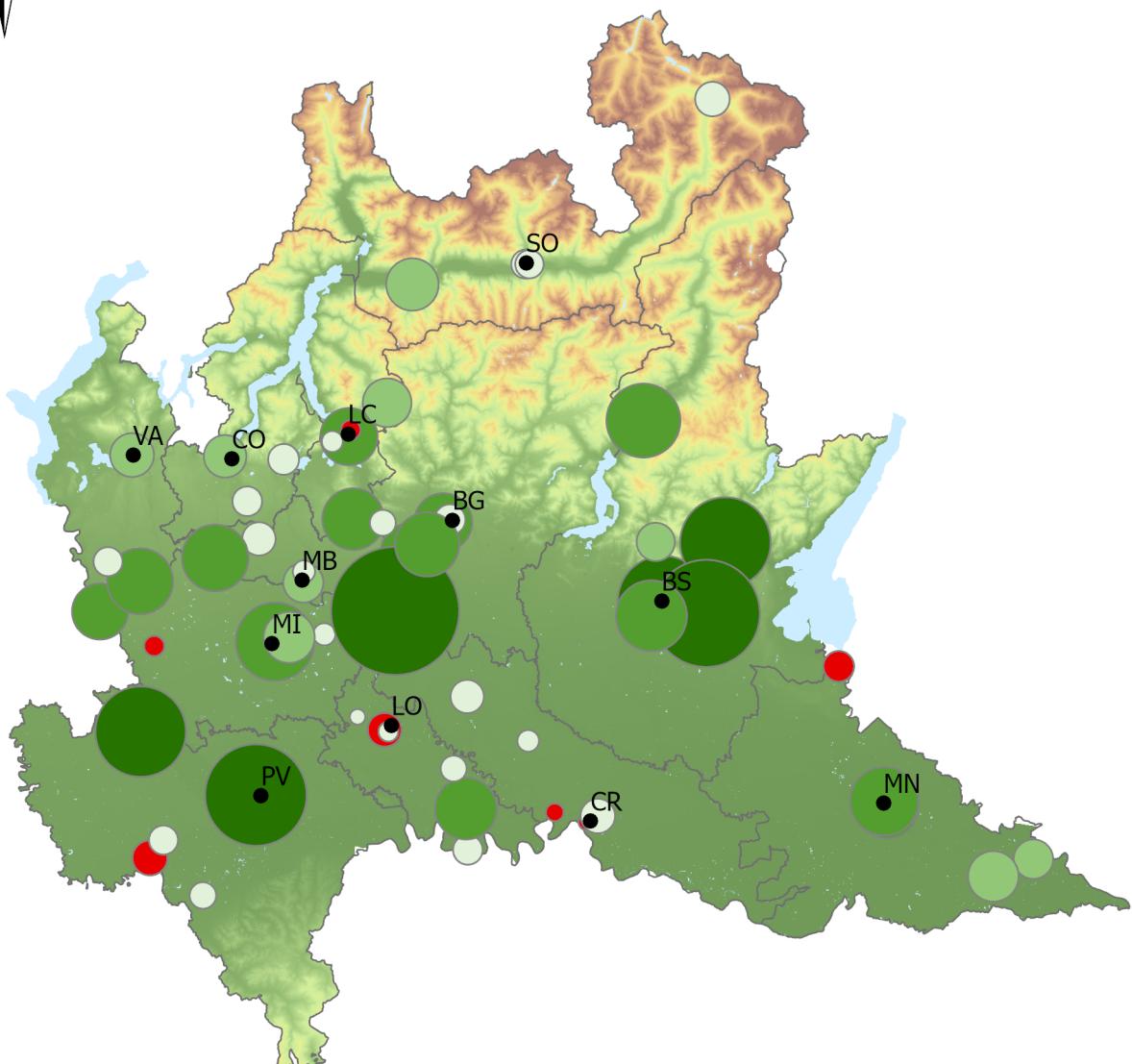
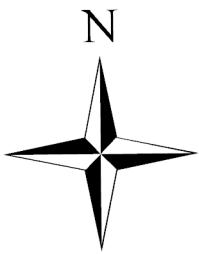
30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

PM10: Concentrazione media 2020

### 4.5.3 Impatto del periodo di lockdown

La seguente mappa permette di apprezzare meglio la variazione di concentrazione nei due successivi anni. La procedura di realizzazione risulta la stessa adottata nella [sotto-sezione 4.3.3](#). Analizzando la variazione dell'inquinante dovuta al periodo di lockdown possiamo notare che la zona di Cremona ha riscontrato la minor variazione e sono rare le zone che hanno rilevato addirittura un aumento dello stesso. Riscontriamo nuovamente una bassa variazione in Sondrio come per l'analisi degli altri inquinanti. Le zone che più hanno beneficiato del decremento di  $PM_{10}$  dovuto al Lockdown sono la Bassa Bergamasca, la zona di Pavia e dintorni e la zona Bresciana con medie fino a  $8,92 \mu g/m^3$ .



#### Differenza di concentrazione

- -1,78 - 0,07 (9)
- 0,07 - 1,75 (23)
- 1,75 - 3,27 (12)
- 3,27 - 5,10 (14)
- 5,10 - 8,92 (6)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# PM10: Impatto Lockdown

## 4.6 PM<sub>2.5</sub>

### 4.6.1 Concentrazione media nel 2019

Ugualmente a quanto fatto per il PM<sub>10</sub> si riporta l'analisi svolta sull'inquinante PM<sub>2.5</sub>. In questo caso, le mappe sono ottenute non più come aggregazione media delle concentrazioni medie per provincia ma come selezione del massimo per ogni provincia. Essendo il comportamento di PM<sub>10</sub> e PM<sub>2.5</sub> molto simile (vedi [sezione 3.5](#)), si vuole porre in questo caso attenzione ai valori massimi rilevati.

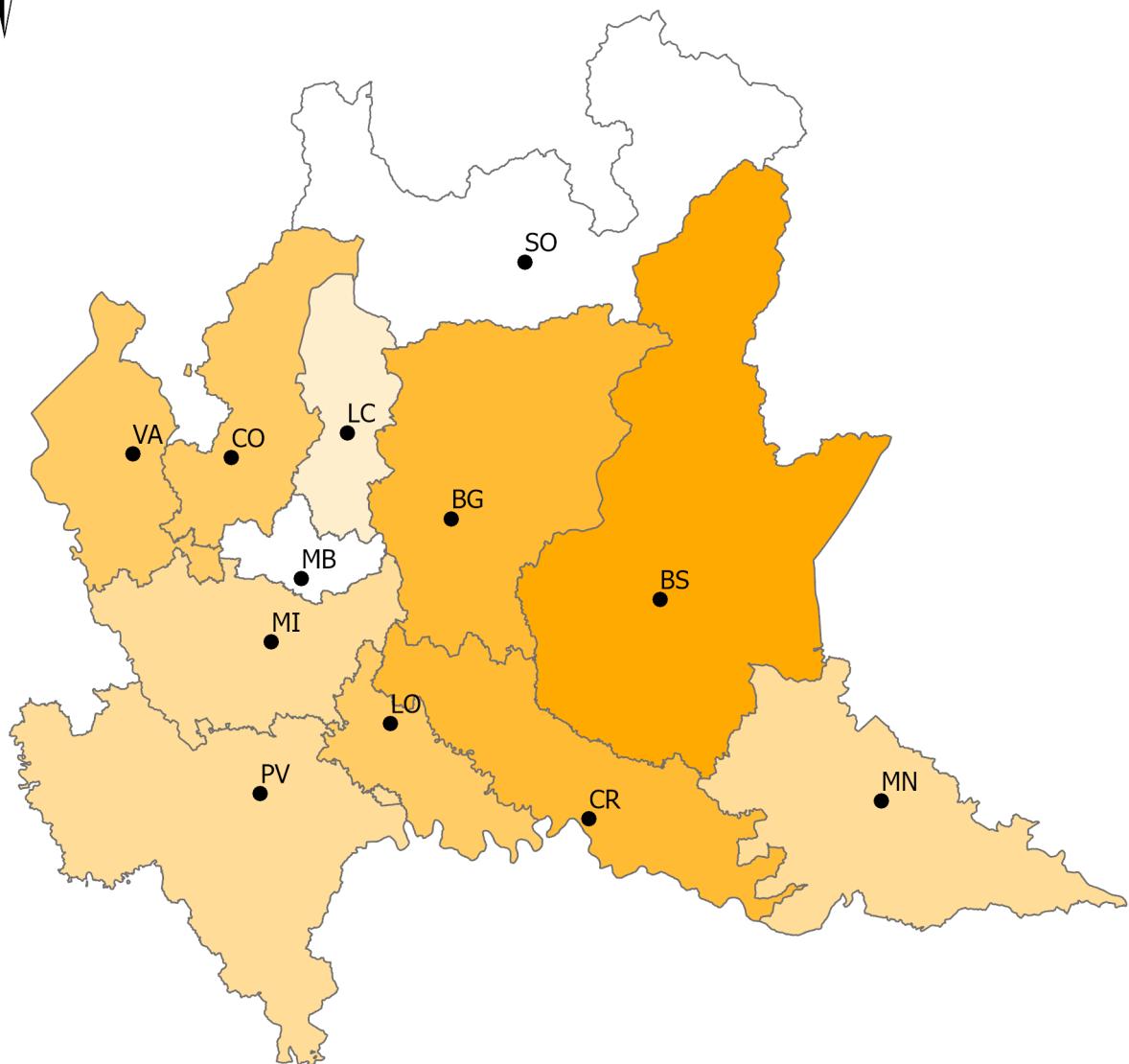
Le zone con minor concentrazione rilevata sono Monza e Brianza e Sondrio con valori medi di circa 11  $\mu\text{g}/\text{m}^3$ . Le altre province fanno registrare valori simili che variano da medie di 13  $\mu\text{g}/\text{m}^3$  fino a medie di 18  $\mu\text{g}/\text{m}^3$ . La provincia che registra la maggior quantità di PM<sub>2.5</sub> è quella bresciana.

### 4.6.2 Concentrazione media nel 2020

Per quanto riguarda l'anno 2020, la zona con minor concentrazione rilevata è Sondrio con valori medi di circa 11  $\mu\text{g}/\text{m}^3$ . La zone con maggior concentrazione invece risultano essere Bergamo e Cremona con valori medi di 16  $\mu\text{g}/\text{m}^3$ .

### 4.6.3 Impatto del periodo di lockdown

Notiamo che la zona di Brescia e di Mantova sono quelle che più hanno beneficiato del Lockdown con rilevante diminuzione di PM<sub>2.5</sub>. Invece Monza e Brianza e Cremona nonostante il Lockdown registrano un deciso incremento dell'inquinante in contrasto alla tendenza delle restanti province che registrano comunque un decremento dello stesso. Le variazioni registrate in media rientrano tra i range di -4  $\mu\text{g}/\text{m}^3$  e +4  $\mu\text{g}/\text{m}^3$ . Inoltre bisogna notare che nell'analisi del lockdown di questo inquinante si riscontra un ratio di 45% tra sensori che hanno riscontrato un miglioramento e quelli che hanno riscontrato un peggioramento dal 2019 al 2020.



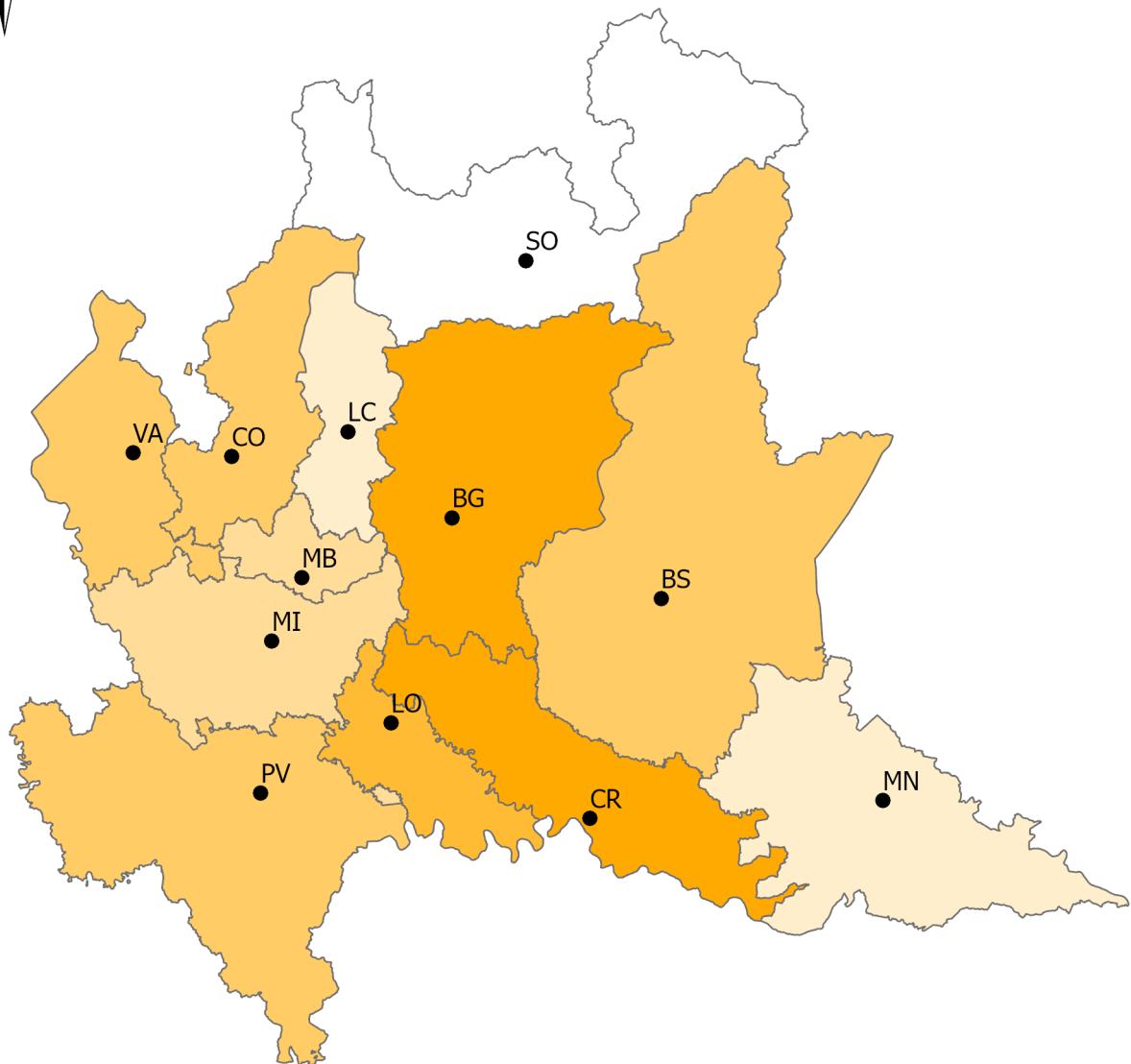
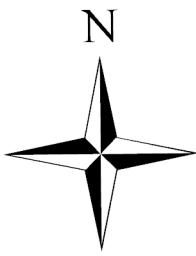
PM2.5 ( $\mu\text{g}/\text{m}^3$ )

- [Light Yellow] 10,76 - 11,41 (2)
- [Pale Orange] 11,41 - 13,24 (1)
- [Orange] 13,24 - 13,59 (3)
- [Darker Orange] 13,59 - 15,51 (3)
- [Medium Orange] 15,51 - 17,54 (2)
- [Dark Orange] 17,54 - 18,24 (1)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

PM2.5: Concentrazione massima 2019



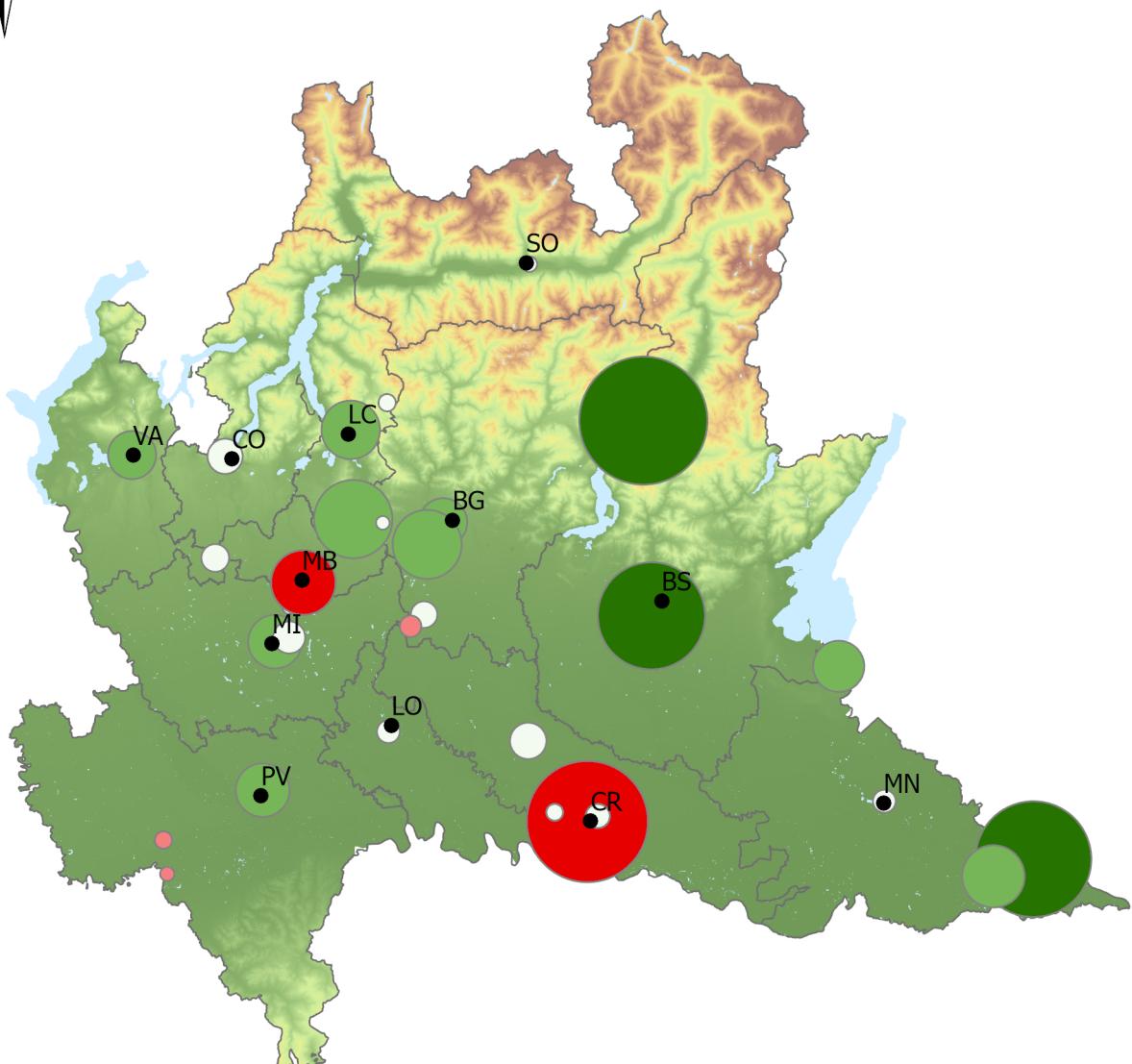
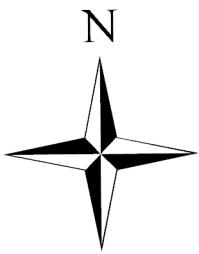
PM2.5 ( $\mu\text{g}/\text{m}^3$ )

- 10,99 (1)
- 10,99 - 11,73 (2)
- 11,73 - 12,87 (2)
- 12,87 - 13,97 (4)
- 13,97 - 14,18 (1)
- 14,18 - 16,25 (2)

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

PM2.5: Concentrazione massima 2020



#### Differenza di concentrazione

- -4,06 - -2,10 (2)
- -2,10 - -0,32 (4)
- -0,32 - 1,13 (13)
- 1,13 - 2,57 (10)
- 2,57 - 4,30 (3)

30 15 0 30 60 90 120  
Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# PM2.5: Impatto Lockdown

## 4.7 Composizione dell'aria

Nelle due figure seguenti sono rappresentati degli istogrammi rappresentanti la composizione degli inquinanti presenti nell'aria per provincia nel 2019 e nel 2020.

Aggregando in modo analogo alle mappe precedenti i dati per provincia e calcolando la media per ogni aggregato, è possibile realizzare una mappa che utilizzi dei grafici come simbologia.

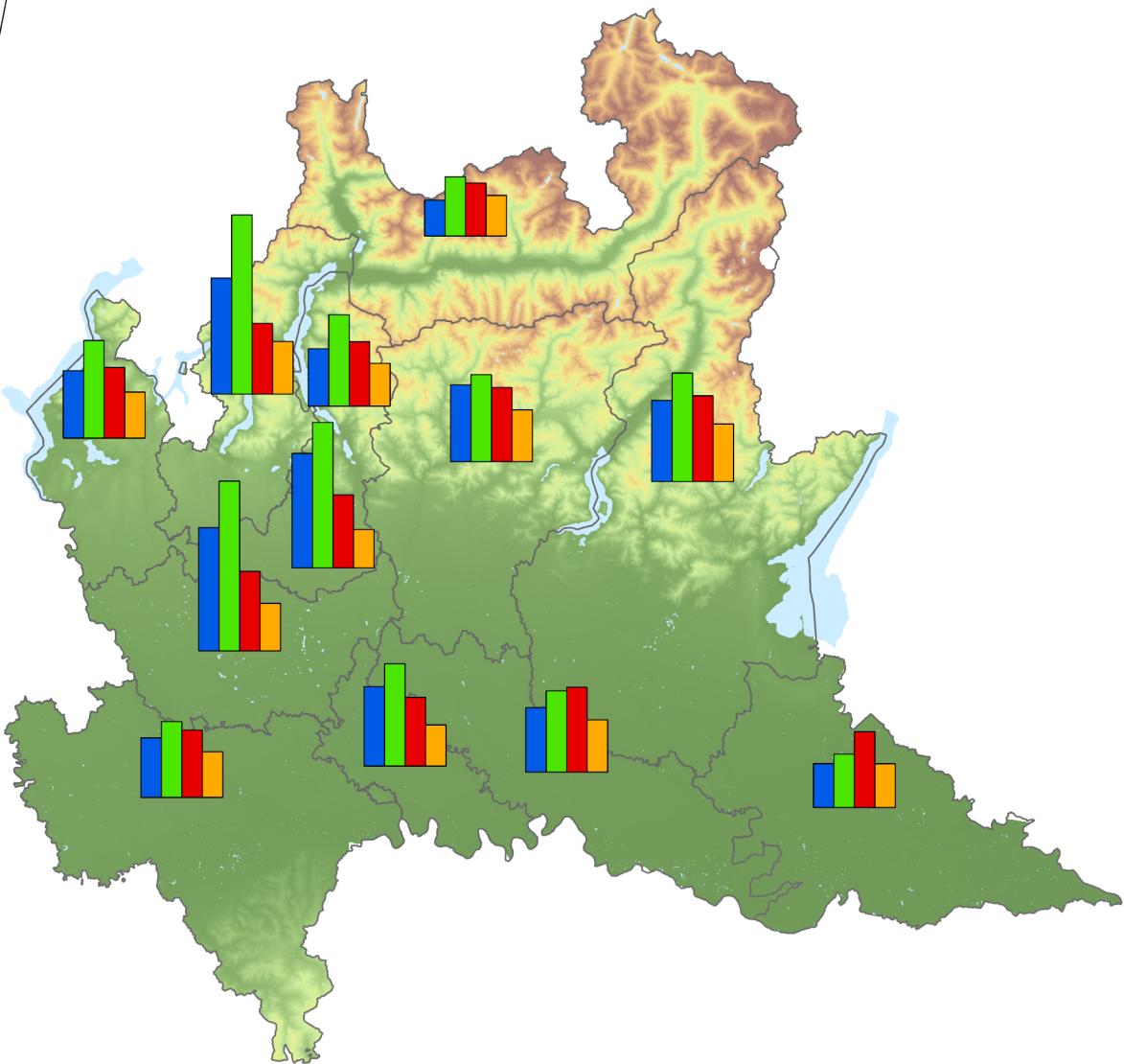
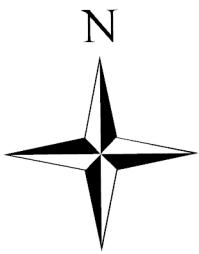
In questo caso, le due mappe condividono la stessa scala di valori. L'istogramma vuol quindi dare un'informazione di proporzione e mostrare (a pari unità di misura) quale inquinante sia maggiormente presente nell'aria. Allo stesso tempo, tuttavia, la condivisione della stessa scala permette di apprezzare una diminuzione o un aumento della concentrazione media in una data provincia.

### 4.7.1 Composizione degli inquinanti nel 2019

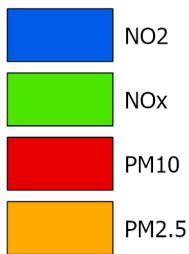
La province meno inquinate sono Sondrio e Mantova con valori di NO<sub>x</sub> e PM<sub>2.5</sub> molto più bassi delle altre province. Quelle che risultano essere più inquinate sono Como, Monza e Brianza e Milano soprattutto con valori nettamente maggiori di NO<sub>2</sub> e NO<sub>x</sub>.

### 4.7.2 Composizione degli inquinanti nel 2020

Evidente con il passaggio dello studio degli inquinanti dal 2019 al 2020 è l'impatto del lockdown che porta a una netta diminuzione degli inquinanti. Sondrio rimane la più virtuosa con bassi rilevazioni degli inquinanti seppur la differenza registrata con gli inquinanti delle altre province si è decisamente assottigliata rispetto al 2019.



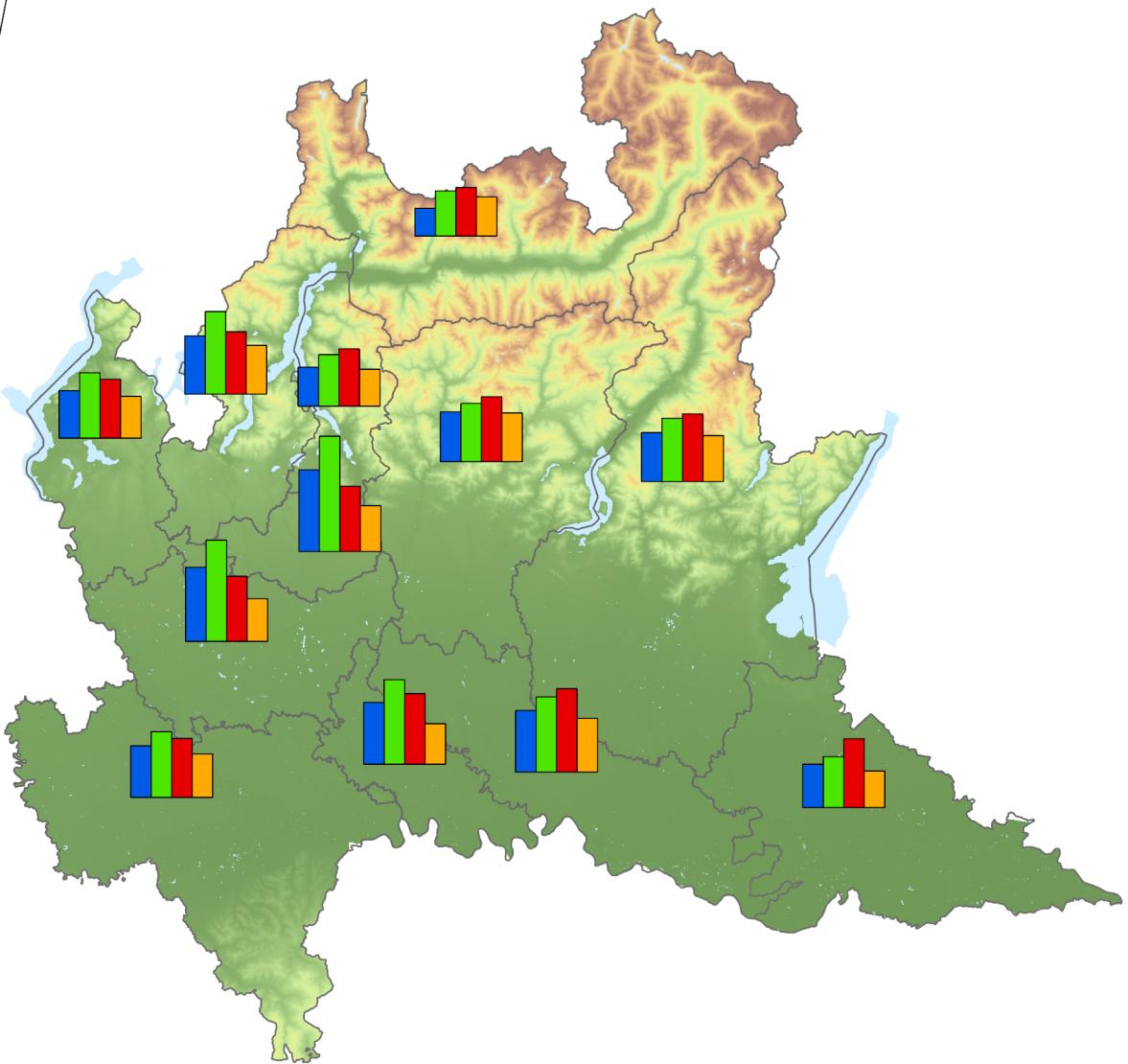
Composizione aria 2019 (Media provincia)



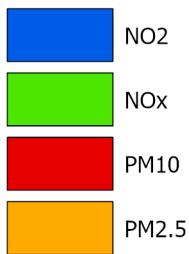
30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

Composizione aria per provincia 2019



Composizione aria 2020 (Media provincia)



30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

Composizione aria per provincia 2020

# Capitolo 5

## Analisi dei regressori

In questo capitolo si studiano più nel dettaglio le cause dell'NO<sub>2</sub>. Come abbiamo visto nel [Capitolo 3](#), tra le maggiori fonti di emissione di NO<sub>2</sub> abbiamo i processi di combustione, tra i quali annoveriamo come maggiormente influenti il traffico veicolare e il riscaldamento nelle abitazioni/industrie. Con questa prospettiva abbiamo svolto un studio sulle cause (regressori) dell'NO<sub>2</sub> nel 2020 con l'obiettivo di dimostrare che fossero effettivamente determinanti e in relazione con i valori dell'inquinante riscontrati.

In particolare si analizzano i seguenti quattro regressori:

- Traffico veicolare
- Rete stradale
- Numero di abitanti
- Temperatura

Per le concentrazioni di NO<sub>2</sub> si farà riferimento alle mappe del 2019 e 2020 visti nel [Capitolo 4](#).

Gli script Python utilizzati per la preparazione dei dati ottenuti dalle varie fonti sono disponibili nel codice allegato.

## 5.1 Analisi del traffico veicolare

Il primo regressore di cui ci siamo occupati è il traffico veicolare. La combustione dei motori impiegati nei veicoli di trasporto è infatti indicata come la maggior causa di NO<sub>2</sub>.

### 5.1.1 Reperimento ed elaborazione dati

I dati "*Matrice OD2020 - Passeggeri*" sono stati reperiti dal Geoportale della Lombardia [9]. I dati si presentano in formato .csv e strutturati in una matrice Origine-Destinazione di valori stimati relativa agli spostamenti all'interno della regione. La matrice fa riferimento ad un giorno feriale medio e considera 1525 zone (di cui 1450 zone interne alla regione e le rimanenti sono quindi state scartate). I dati sugli spostamenti sono distinti per origine, destinazione, fascia oraria, motivo e metodo di spostamento. La matrice regionale Origine-Destinazione include 8 modalità (auto conducente, auto passeggero, TPL gomma, TPL ferro, moto, bici, piedi e altro) e 5 motivi (lavoro, studio, occasionali, affari, rientri a casa).

Sulla matrice è stato effettuato un lavoro di pulizia dei dati: si sono tenute solo le colonne relative ai mezzi a motore e si sono aggregati tra loro i valori relativi ad una stessa coppia origine-destinazione in diverse fasce orarie. In questo modo è possibile ottenere il numero totale di persone su mezzi a motore che si spostano tra una coppia di comuni lombardi nell'intera giornata. Nella Tabella 5.1 si osserva il risultato della fase di preparazione. La tabella in realtà riporta altre informazioni di contorno, non mostrate per semplicità.

ZONA_ORIG	ZONA_DEST	LAT_ORIG	LAT_DEST	LON_ORIG	LON_DEST	Total
LECCO	LECCO	45.857915	45.857915	9.422787	9.422787	36210.96
VOGHERA	VOGHERA	45.008314	45.008314	9.013631	9.013631	33483.05
GALLARATE	GALLARATE	45.662789	45.662789	8.793204	8.793204	33410.71
CREMA	CREMA	45.366003	45.366003	9.683362	9.683362	31037.17
ARESE	ARESE	45.550203	45.550203	9.073570	9.073570	27960.58

Tabella 5.1: File della matrice dei trasporti (dopo il pre processing)

I dati "*Shape Matrice OD2020 - Passeggeri - Zone interne*" sono stati reperiti dal Geoportale della Lombardia [10] e consistono in uno shapefile delle zone individuate dalla matrice precedentemente illustrata.

### 5.1.2 Risultati analisi

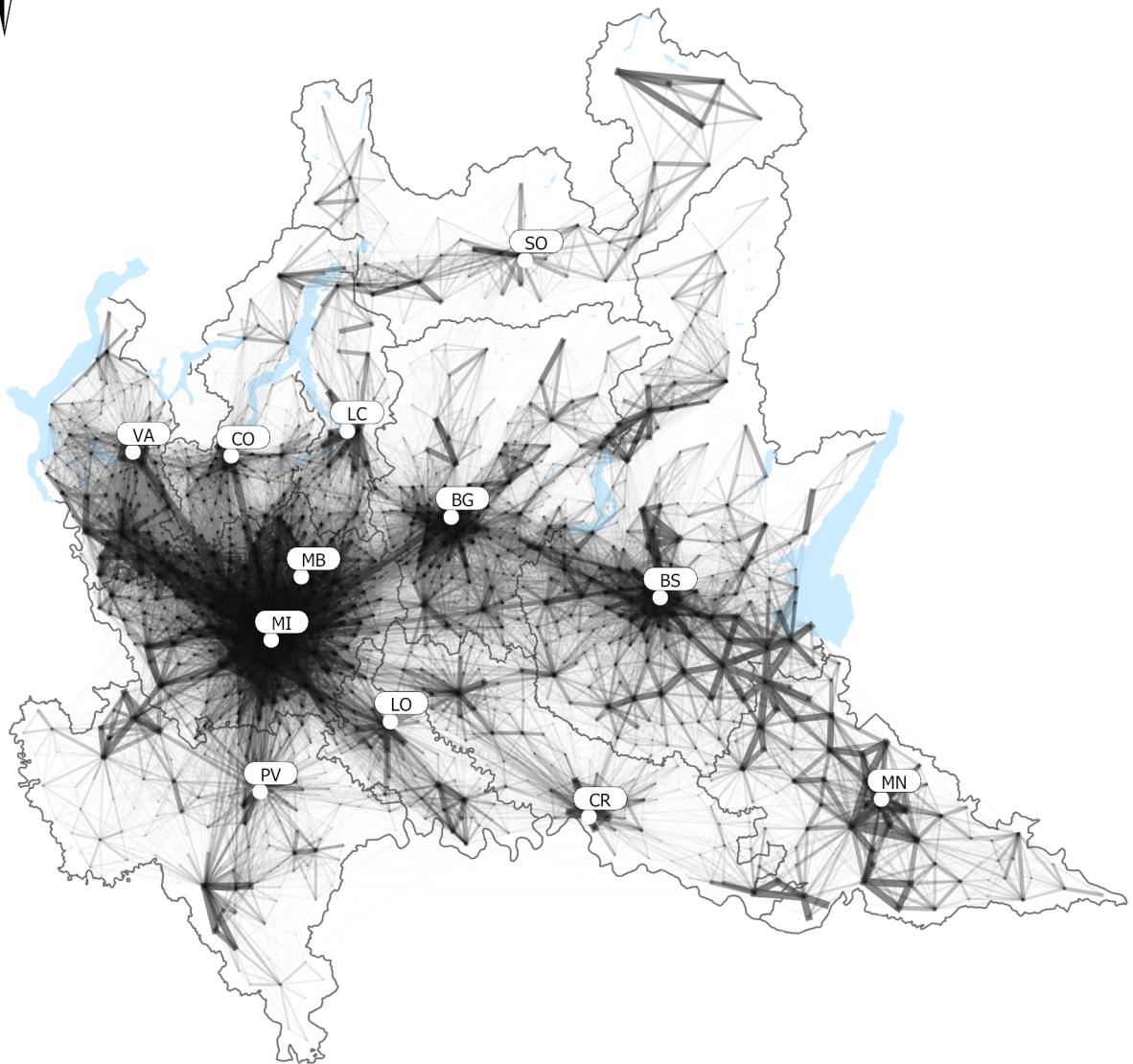
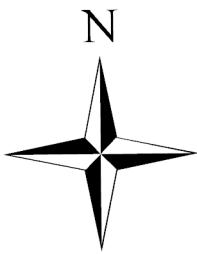
La mappa seguente mostra gli spostamenti tra coppie di comuni. Ogni coppia di comuni presenta un collegamento tanto più scuro tanto più è trafficato, indipendentemente dalla tipologia di spostamento.

Per rendere la rappresentazione più nitida, sono stati rimossi i collegamenti con il numero

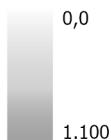
di spostamenti minore, superficiali per questa analisi e che avrebbero reso la mappa poco chiara.

Per poter tracciare i collegamenti, è stato inoltre calcolato il centroide di ogni zona e aggiunta l'informazione di latitudine e longitudine alla tabella contenente le informazioni di trasporto. Con il tool **XT to Line** di ArcGIS Pro quindi è stato possibile collegare i vari comuni e modificare la simbologia delle rette affinchè avessero un colore proporzionato al numero totale di spostamenti.

I valori dei collegamenti variano da 0 (al di sotto dei 1000 spostamenti giornalieri in media) fino a 1.100 (più di 1 milione di spostamenti medi giornalieri). Confrontando gli spostamenti possiamo confermare la grande correlazione tra traffico veicolare e andamento dell' $\text{NO}_2$ . Le zone con maggior rilevazioni di  $\text{NO}_2$  risultano infatti Milano, Monza-Brianza e Brescia in cui si rilevano anche il maggior numero di spostamenti. Allo stesso modo le province che riscontano minor valore di  $\text{NO}_2$  che sono quelle nel nord della regione (Sondrio in primo luogo) sono le stesse che rilevano il minor numero di spostamenti.



Numero di spostamenti medio (auto, moto, trasporto pubblico) in migliaia



Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# Regressore traffico veicolare

## 5.2 Analisi della rete stradale

Proseguendo lo studio sul traffico veicolare come regressore dell' $\text{NO}_2$  abbiamo spostato l'attenzione sulla rete stradale. Questa è sicuramente correlata al traffico veicolare ma ignota invece è la sua correlazione con l'inquinante.

### 5.2.1 Reperimento ed elaborazione dati

I dati "*Rete stradale*" sono stati reperiti dal Geoportale della Lombardia [11]. Si presentano come uno shapefile e le strade sono suddivise tramite l'attributo "*DTIPO\_CTR*" in strada comunale, strada provinciale, strada statale o autostrada/tangenziale. Le strade comunali e provinciali si sono rivelate da subito poco utili per il nostro studio in quanto particolarmente ingombranti sulla mappa, rendendola illeggibile. Si è quindi deciso di escluderle. Di contro, le strade statali e le autostrade hanno una distribuzione non uniforme e possono quindi essere utilizzate per lo studio dell'inquinante.

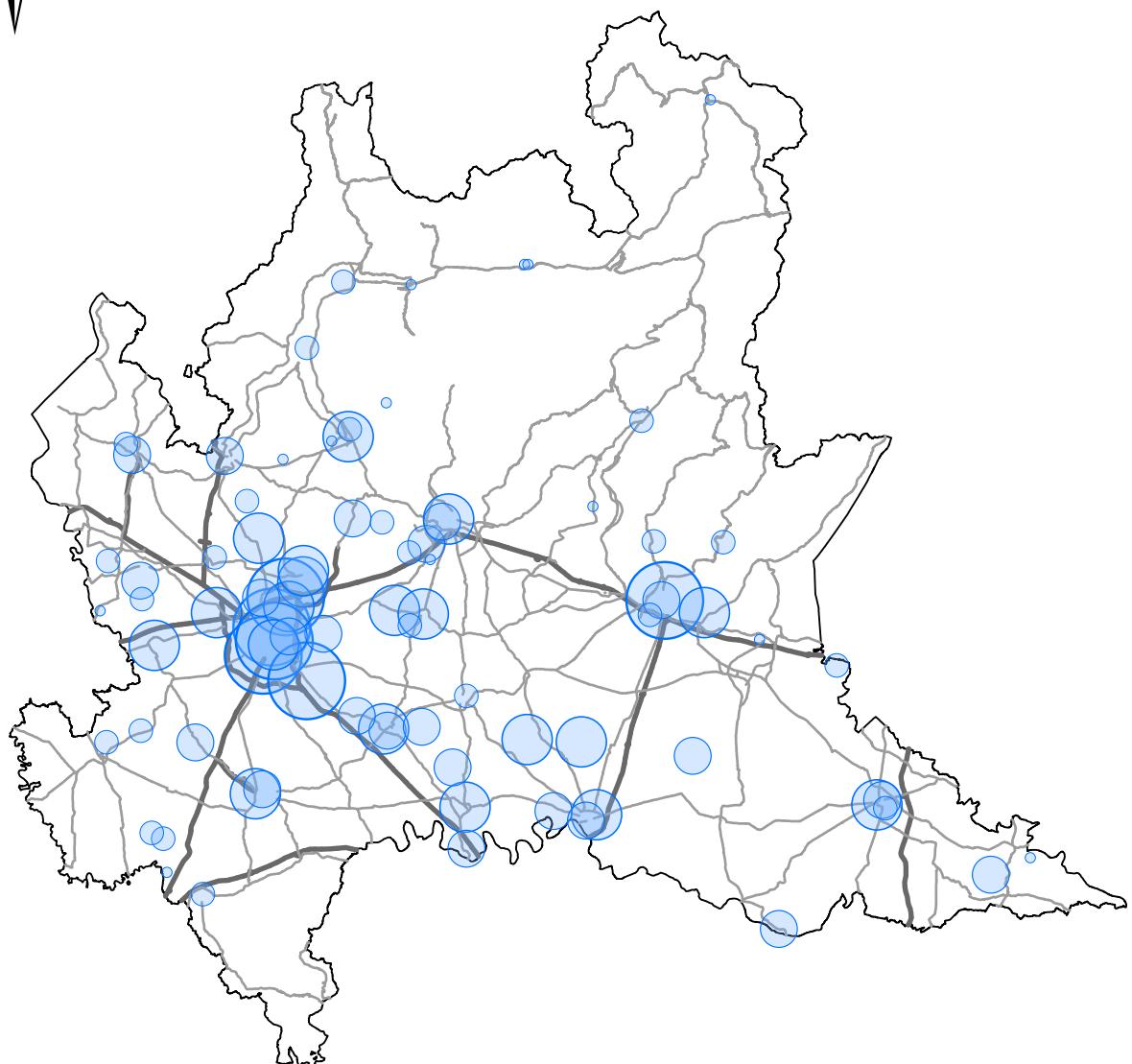
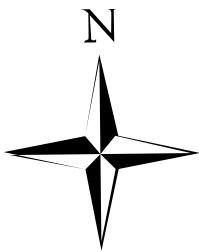
La Tabella 5.2 ci dà informazioni sulla struttura dati dello shapefile:

FID	DTIPO_CTR	NOME	geometry
221	Strada Comunale	STRADE COMUNALI PROVINCIA DI SONDRIO	LINESTRING
222	Strada Comunale	STRADE COMUNALI PROVINCIA DI SONDRIO	LINESTRING
223	Strada Statale	DEL LAGO DI COMO E DELLO SPLUGA	LINESTRING
224	Strada Comunale	STRADE COMUNALI PROVINCIA DI SONDRIO	LINESTRING
225	Strada Comunale	STRADE COMUNALI PROVINCIA DI SONDRIO	LINESTRING

Tabella 5.2: File della rete stradale

### 5.2.2 Risultati analisi

La mappa successiva illustra rispettivamente in nero le autostrade, in grigio le strade statali e con pallini blu di diverse dimensioni l'andamento dell'inquinante  $\text{NO}_2$  nel 2020. Si può notare una decisa correlazione tra la rete stradale e l'inquinante. La forte presenza di  $\text{NO}_2$  nei dintorni di Milano, Monza-Brianza e Brescia è infatti ben spiegata dalla presenza di diverse strade in queste zone, con conseguente probabilità di traffico. Allo stesso modo la bassa presenza di autostrade e strade statali nelle zone del nord della regione (soprattutto in provincia di Sondrio) ben spiega basse concentrazioni di  $\text{NO}_2$ .



NO<sub>2</sub> 2020 ( $\mu\text{g}/\text{m}^3$ )

- 5,25 - 9,75
- 9,75 - 13,24
- 13,24 - 17,64
- 17,64 - 23,56
- 23,56 - 32,16

Rete stradale

- Strade statali
- Autostrade

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>2</sub>: regressore rete stradale

## 5.3 Analisi della popolazione

Come terza possibile cause abbiamo pensato alla distribuzione della popolazione sul territorio Lombardo. Il risultato atteso è che il numero di abitanti sia in relazione alla quantità di riscaldamento abitativo e industriale emesso e che sia quindi in stretta correlazione con l'inquinante.

### 5.3.1 Reperimento ed elaborazione dati

I dati "*Dati statistici ISTAT*" sono stati reperiti dal Geoportale della Lombardia [12]. Si presentano sotto forma di shapefile e contenenti informazioni sulla popolazione residente, ovvero dati rielaborati da Eupolis Lombardia da dati ISTAT relativi alla popolazione residente del 2020. Per ogni comune lombardo si riporta il numero di abitanti. Le città di grandi dimensione, come ad esempio Milano o i principali capoluoghi di provincia, sono suddivise in zone/distretti e per ognuno di essi è indicato il numero di abitanti.

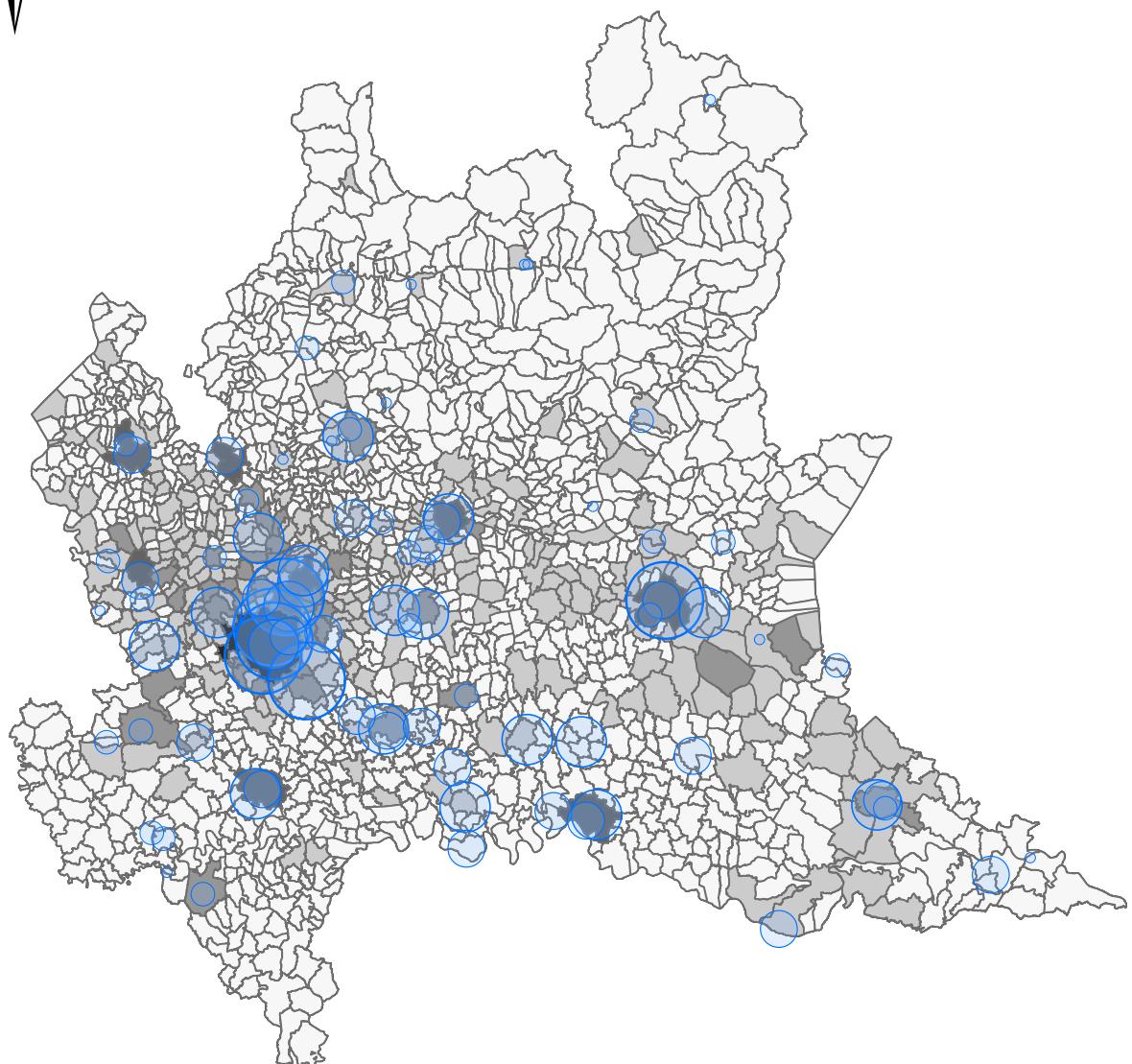
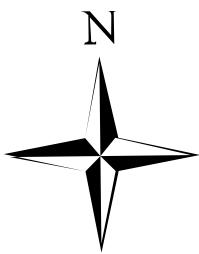
La Tabella 5.3 ci dà informazioni sulla struttura dati dello shapefile:

NOME_COM	NOME_PRO	POP_2020	geometry
CAMPOSPINOSO	PAVIA	1078.0	POLYGON
BRUNELLO	VARESE	936.0	POLYGON
INTROBIO	LECCO	1953.0	POLYGON
PIARIO	BERGAMO	1066.0	POLYGON
GORLA MAGGIORE	VARESE	4954.0	POLYGON

Tabella 5.3: File dei dati statistici ISTAT

### 5.3.2 Risultati analisi

Nella mappa successiva, i comuni presentano una colorazione che va dal bianco al nero in funzione della classe di appartenenza (determinata dal numero di abitanti). Possiamo constatare una evidente correlazione tra la popolosità del comune e il valore dell'inquinante rilevato. I quartieri di Milano e Monza e Brianza che risultano essere i più popolati sono anche gli stessi che registrano il maggior valore di  $NO_2$ . La seconda zona che rileva la maggior quantità di  $NO_2$  è Brescia ed è ben spiegata dal fatto che si concentra un elevato numero di abitanti.



NO<sub>2</sub> 2020 ( $\mu\text{g}/\text{m}^3$ ) Numero abitanti

- 5,25 - 9,75
- 9,75 - 13,24
- 13,24 - 17,64
- 17,64 - 23,56
- 23,56 - 32,16
- 30- 7227
- 7227 - 24082
- 24082 - 62969
- 62969 - 196340
- 196340 - 1406242

30 15 0 30 60 90 120 Kilometers

Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

# NO<sub>2</sub>: regressore popolazione

## 5.4 Analisi della temperatura

Lo studio prosegue con un'analisi del legame tra la temperatura ed il livello di concentrazione dell'inquinante.

### 5.4.1 Reperimento ed elaborazione dati

I dati "*Dati sensori meteo 2020*" sono stati reperiti dal Geoportale della Lombardia [13]. I dati si presentano in formato tabellare, la cui struttura è riassunta nella [Tabella 5.4](#). Nella tabella originale sono presenti altre informazioni di contorno, omesse per semplicità.

<b>IdSensore</b>	<b>Data</b>	<b>Valore</b>	<b>Stato</b>	<b>idOperatore</b>
9	01/01/2020 00:00:00	0.0	VA	4
46	01/01/2020 00:00:00	0.0	VA	4
115	01/01/2020 00:00:00	2.9	VA	1
116	01/01/2020 00:00:00	0.0	VA	4
117	01/01/2020 00:00:00	214.0	VA	1
121	01/01/2020 00:00:00	10.9	VA	1
122	01/01/2020 00:00:00	0.0	VA	4
2001	01/01/2020 00:00:00	0.2	VA	1

Tabella 5.4: File delle rilevazioni dei sensori

La [Tabella 5.5](#) fornisce informazioni sui sensori, quali il loro identificativo, la loro tipologia, la loro unità di misura, l'identificativo della stazione a cui fanno riferimento, la quota, la provincia e le coordinate.

<b>IdSensore</b>	<b>Tipologia</b>	<b>Dim</b>	<b>IdStazione</b>	<b>Quota</b>	<b>Prov</b>	<b>lng</b>	<b>lat</b>
22006	Umidità Relativa	%	1890	1790	BS	10.175407	45.753896
22007	Direzione Vento	°	1890	1790	BS	10.175407	45.753896
22003	Temperatura	°C	1890	1790	BS	10.175407	45.753896
22009	Velocità Vento	m/s	1890	1790	BS	10.175407	45.753896
17612	Altezza Neve	cm	1545	1285	CO	9.238323	46.071947
22011	Precipitazione	mm	1890	1790	BS	10.175407	45.753896
22002	Altezza Neve	cm	1890	1790	BS	10.175407	45.753896
22017	Radiazione Globale	W/m <sup>2</sup>	1890	1790	BS	10.175407	45.753896

Tabella 5.5: File dei sensori

Per riuscire a ottenere la relazione tra l'informazione relativa ad un dato sensore e quella relativa alla temperatura media rilevata, si svolti i seguenti passi sequenziali:

1. Per ogni sensore si calcola la media delle rilevazioni acquisite. Ogni sensore infatti è predisposto per rilevare solamente una tipologia di fenomeno (es. temperatura o velocità vento o umidità relativa).
2. Si selezionano le sole acquisizioni relative alla temperatura, nonché le uniche di interesse.
3. Si intersecano le due tabelle in modo da ottenere la corrispondenza tra sensore (con relative coordinate spaziali) e temperatura media.

Nel [Listing 5.1](#) si è riportato il codice utilizzato per svolgere le operazioni di elaborazione sui dati.

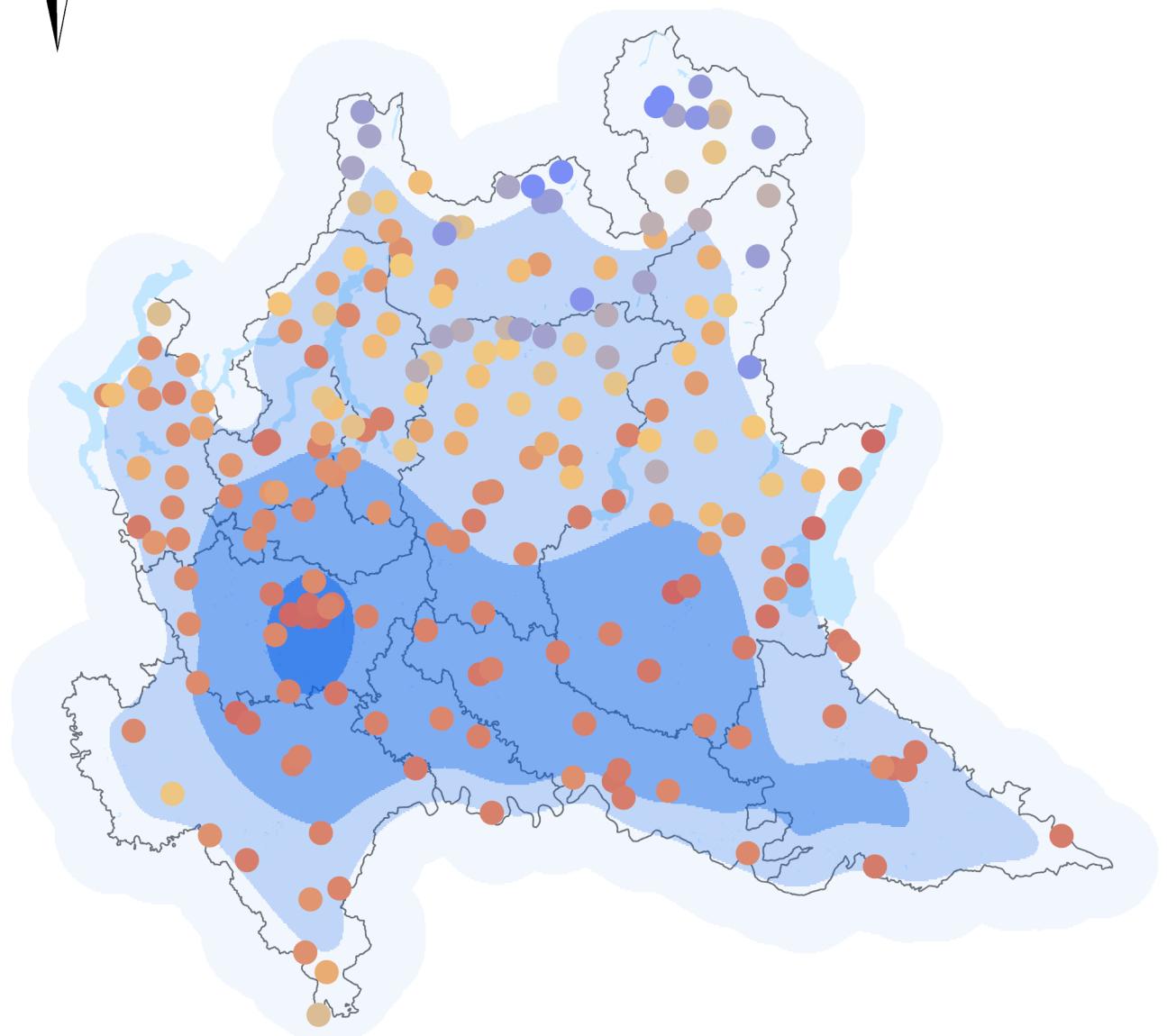
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Lettura csv Id Sensori
5 csv1 = 'resources/id_sensori.csv'
6 data1 = pd.read_csv(csv1)
7
8 # Lettura csv rilevazioni sensori
9 csv2 = 'resources/rilevazioni_sensori.csv'
10 data2 = pd.read_csv(csv2)
11
12 # Media per sensore
13 data1 = data1.groupby("IdSensore").mean()
14
15 # Merge
16 table = pd.merge(data1, data2[data2["Tipologia"]=="Temperatura"], how='inner', on=['IdSensore'])
17
18 # Salvataggio in .csv
19 table.to_csv(r'output\temperatura_2020.csv', index = False)
```

Listing 5.1: Join .csv sensori temperatura

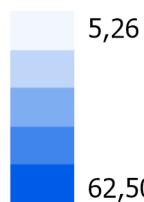
#### 5.4.2 Risultati analisi

Nella mappa seguente si illustrano con cerchi di gradazione dal viola al rosso i valori temperatura, con un'escursione che varia rispettivamente da -3,2° di media a 16° gradi di media, e in tonalità di blu la distribuzione dell'inquinante nel 2020. Come possiamo notare vi è una certa correlazione tra i due fenomeni e le zone con maggior rilevazione dell'inquinante registrano anche i picchi maggiori di temperatura media rilevata. Viceversa

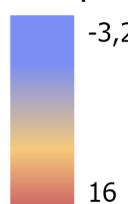
le zone a minor concentrazione di inquinante registrano le minori temperature. Seppur con scarsa frequenze, sono visibili diverse eccezioni che non evidenziano con la stessa chiarezza questa correlazione positiva. Tra le diverse cause analizzate fino ad ora, si può dire che la temperatura sia quella che porta minore informazione.



Inquinante NO<sub>2</sub> 2020



Temperatura (°C)



Scala 1:1.500.000  
WGS 1984 UTM Zone 32N

NO<sub>2</sub>: regressore temperatura

# Capitolo 6

## Interpretazione dei risultati

### 6.1 Confronto tra i regressori

Secondo il nostro studio svolto sul territorio Lombardo sull'inquinante NO<sub>2</sub> nel 2020 possiamo affermare che:

- Il traffico veicolare, rappresentato come numero medio di spostamenti, come da aspettative è un ottimo regressore per l'inquinante.
- La rete stradale, composta da autostrade e strade statali, risulta anch'essa essere un ottimo regressore per l'inquinante. I risultati sono in alcune zone persino migliori di quelli rilevati con il traffico veicolare.
- La popolazione, in particolare il numero di abitanti approssima abbastanza bene il riscaldamento presente nelle strutture abitative, risultando quindi un buon regressore per l'inquinante.
- La temperatura dimostra di avere una buona correlazione ma ottiene risultati peggiori dei precedenti tre regressori che si dimostrano essere più precisi. Il motivo risiede nel fatto che la temperatura utilizzata come approssimazione al riscaldamento delle abitazioni/industrie è sicuramente indicativa ma non risulta essere precisa nel riprodurne l'andamento.



# Capitolo 7

## Conclusioni

Il lavoro si è suddiviso su due fronti:

- Open Data Cube (ODC).
- Analisi degli inquinanti e dell'impatto del lockdown sull'inquinamento dell'aria.

Per quanto riguarda ODC, l'esperienza ci ha permesso di addentrarci nella complessa logica nonché nel funzionamento di questa tecnologia. L'obiettivo finale che ci eravamo prefissati è stato rispettato riuscendo a eseguire la creazione di un Open Data Cube costruito sul territorio lombardo con dati riguardanti sostanze inquinanti. La trascrizione di ogni passaggio svolto avrà anche la funzione di guida per futuri utilizzatori visto l'oneroso dispendio per la parte di setup e la non sempre fedele guida ufficiale presente tutt'ora.

Per quanto riguarda l'analisi degli inquinanti, è stato un progetto mosso dalla domanda che in molti ci siamo posti in questo periodo difficile di pandemia: "il lockdown ha portato con sé anche fattori positivi?" e in caso affermativo "in che misura in termini di riduzione degli inquinanti dell'aria?". Abbiamo constatato che effettivamente il lockdown ha portato ad una decisa diminuzione dei livelli degli inquinanti presenti nell'aria. Fattore decisamente positivo considerato che la pianura padana è una delle zone maggiormente inquinate d'europa. Per proseguire lo studio ci siamo chiesti quali fossero le cause di questi inquinanti e una volta ricercate le fonti di emissione abbiamo studiato come questi regressori spiegassero la distribuzione dell'inquinante. Visto il forte impatto del traffico veicolare tra le cause dell'inquinante, l'avvento delle macchine elettriche e di altri mezzi di trasporto alternativi (quali ad esempio i monopattini elettrici), consentono di vedere in maniera ottimistica il futuro in termini di livelli di inquinamento dell'aria.

## 7.1 Sviluppi futuri

Per quanto riguarda Open Data Cube il nostro lavoro può essere utile come "apriporta" per successivi lavori visto lo scarso utilizzo in Italia rispetto all'estero. L'obiettivo ultimo da raggiungere è la creazione di un Open Data Cube Lombardo in cui vengano caricati i dati di diversi dataset. In questo modo la presenza di diversi dataset radunati in un unico grande database consentirà l'accesso ad una sempre maggiore quantità di dati differenti e ad una maggior efficienza nell'elaborazione degli stessi.

Nello studio degli inquinanti dell'aria invece vi è ancora ampio margine di approfondimento in diverse direzioni. In primo luogo, interessante sarebbe poter estendere lo studio all'intera Italia e svolgere confronti tra le varie regioni, piuttosto che limitarsi allo studio regionale sulla Lombardia. In seconda battuta, visto il protrarsi della pandemia e del conseguente lockdown, sarebbe interessante ripetere lo studio sui dati aggiornati che si potranno rilevare nel 2021. Questo studio su un arco temporale più ampio consentirebbe di trarre conclusioni di carattere più generale.

# Capitolo 8

## Sviluppo

### 8.1 Python

Illustriamo le principali librerie e pacchetti Python usati nel progetto e alcuni confronti con le alternative in MATLAB<sup>1</sup>.

#### 8.1.1 xarray

Il pacchetto *xarray*<sup>2</sup> è un progetto open source utile per lavorare con array multidimensionali in modo semplice ed efficiente.

Grazie a questa libreria si è potuto importare i dati provenienti da files in formato NetCDF o Shapefile.

#### 8.1.2 pandas

Il pacchetto *pandas*<sup>3</sup> è uno strumento open source che permette la manipolazione e l'analisi dei dati in modo veloce, potente e flessibile.

Questa libreria è stata usata per trattare la serie storica dell'ammoniaca e creare il dataframe contenente i regressori (armoniche).

#### 8.1.3 geopandas

Il pacchetto *geopandas* è un progetto open source per semplificare il lavoro con i dati geospaziali in Python. GeoPandas estende i tipi di dati utilizzati dai panda per consentire operazioni spaziali sui tipi geometrici. Le operazioni geometriche vengono eseguite da

---

<sup>1</sup>MATLAB (<https://it.mathworks.com/products/matlab.html>)

<sup>2</sup>xarray (<http://xarray.pydata.org/>)

<sup>3</sup>pandas (<https://pandas.pydata.org/>)

*shapely*. GeoPandas dipende inoltre da *fiona* per l'accesso ai file e da *matplotlib* per i grafici.

### 8.1.4 matplotlib

Il pacchetto *matplotlib*<sup>4</sup> è una libreria per creare grafici statici, animati e interattivi in Python.

Questa libreria è stata usata per generare tutti i grafici presenti nel progetto, riuscendo a gestire un'ampia varietà di dati come input e permettendo la personalizzazione di ogni dettaglio del grafico quale colori, legenda, scala, sovrapposizioni di immagini...

### 8.1.5 numpy

Il pacchetto *numpy*<sup>5</sup> è fondamentale per il calcolo scientifico in Python.

Offre diverse funzioni tra cui: numerose funzioni matematiche di base, generazione casuale di numeri, algoritmi di algebra lineare, trasformate di Fourier.

Jupyter Notebook *Jupyter Notebook*<sup>6</sup> è un'applicazione open source che consente di creare e condividere documenti che contengono live script, equazioni, visualizzazioni e commenti. Gli usi includono: pulizia e trasformazione dei dati, simulazione numerica, modellazione statistica, visualizzazione dei dati, apprendimento automatico e molto altro ancora.

---

<sup>4</sup>matplotlib (<https://matplotlib.org/>)

<sup>5</sup>numpy (<https://numpy.org/>)

<sup>6</sup>Jupyter Notebook (<https://jupyter.org/>)

# Bibliografia & Sitografia

- [1] *Sentinel 5P*. URL: <https://docs.sentinel-hub.com/api/latest/data/sentinel-5p-12/> (cit. a p. 2).
- [2] *ARPA Lombardia*. URL: <https://www.arpalombardia.it/> (cit. a p. 2).
- [3] *Open Data Cube*. URL: <https://www.opendatacube.org/> (cit. a p. 5).
- [4] *ARPA Lombardia NO<sub>2</sub>*. URL: <https://www.arpalombardia.it/Pages/Aria/Inquinanti/Biossido-di-Azoto.aspx?firstlevel=Inquinanti> (cit. a p. 17).
- [5] *Rilevazioni Ancler PM<sub>10</sub>*. URL: <https://ancler.com/pm-10/> (cit. a p. 20).
- [6] *DTM 20 - ESRI GRID*. URL: <https://bit.ly/3ePZdjn> (cit. a p. 27).
- [7] *Lago 10000 CT10*. URL: <https://www.dati.lombardia.it/Territorio/Lago-10000-CT10/qm9t-uzst> (cit. a p. 27).
- [8] *Limiti amministrativi*. URL: <https://bit.ly/30Zfnys> (cit. a p. 27).
- [9] *Matrice OD2020 - Passeggeri*. URL: <https://www.dati.lombardia.it/Mobilit-e-trasporti/Matrice-OD2020-Passeggeri/hyqr-mpe2> (cit. a p. 54).
- [10] *Shape Matrice OD2016 - Passeggeri - Zone interne*. URL: <https://www.dati.lombardia.it/Mobilit-e-trasporti/Shape-Matrice-OD2016-Passeggeri-Zone-interne/pium-6jqb> (cit. a p. 54).
- [11] *Rete stradale*. URL: <https://www.dati.lombardia.it/Territorio/Rete-Stradale/nyxe-8s3y> (cit. a p. 57).
- [12] *Dati statistici ISTAT*. URL: <https://bit.ly/391Pjre> (cit. a p. 59).
- [13] *Dati sensori meteo 2020*. URL: <https://www.dati.lombardia.it/Ambiente/Dati-sensori-meteo-2020/erjn-istm> (cit. a p. 61).



# Elenco delle figure

2.1	ODC come layer tra provider ed applicazioni . . . . .	6
2.2	La posizione delle installazioni attuali di Open Data Cube nel mondo . . . . .	7
2.3	Schema dell'ecosistema Open Data Cube . . . . .	7
2.4	Schema realizzazione ODC . . . . .	9
2.5	Esempio del contenuto di un product . . . . .	16
2.6	Visualizzazione grafica del contenuto del product . . . . .	16
3.1	NO <sub>2</sub> , distribuzione e serie storica . . . . .	18
3.2	NO <sub>x</sub> , distribuzione e serie storica . . . . .	20
3.3	Concentrazioni PM <sub>10</sub> in Europa, 2013 . . . . .	21
3.4	Le fonti di emissioni di PM <sub>10</sub> , 2005-2015 . . . . .	22
3.5	PM <sub>10</sub> , distribuzione e serie storica . . . . .	23
3.6	PM <sub>2.5</sub> , distribuzione e serie storica . . . . .	24
3.7	Correlazioni tra inquinanti . . . . .	25
4.1	Numero di sensori nel territorio lombardo . . . . .	30



# Elenco dei listati

2.1	Creazione ambiente virtuale in Python . . . . .	8
2.2	Installazione dei requisiti . . . . .	8
2.3	Installazione e uso dei Jupyter Notebooks . . . . .	9
2.4	Variabili d'ambiente di sistema di PostgreSQL . . . . .	10
2.5	Creazione database . . . . .	10
2.6	Configurazione database (File <code>.datacube.conf</code> ) . . . . .	10
2.7	Installazione datacube . . . . .	10
2.8	Versione del datacube . . . . .	11
2.9	Versione del datacube . . . . .	11
2.10	Inizializzazione database . . . . .	11
2.11	Aggiunta del prodotto al datacube . . . . .	11
2.12	Esempio del file <code>product.yaml</code> . . . . .	12
2.13	Validazione del <code>dataset.yaml</code> . . . . .	13
2.14	Aggiunta del <code>dataset.yaml</code> al datacube . . . . .	13
2.15	Risolvere problemi con <code>shapely</code> . . . . .	13
2.16	Risolvere problemi con <code>sqlalchemy</code> . . . . .	13
2.17	Esempio del file <code>dataset.yaml</code> . . . . .	13
2.18	Importare un datacube in Python . . . . .	15
2.19	Lista dei product nel datacube . . . . .	15
2.20	Accesso al product . . . . .	15
5.1	Join <code>.csv</code> sensori temperatura . . . . .	62
A.1	Preparazione dataset degli inquinanti . . . . .	79
B.1	Kriging dei valori di concentrazione degli inquinanti . . . . .	81



# Elenco delle tabelle

3.1	Parametri normativi $NO_2$	18
3.2	Parametri normativi $NO_x$	19
3.3	Parametri normativi $PM_{10}$	21
3.4	Parametri normativi $PM_{2.5}$	24
5.1	File della matrice dei trasporti (dopo il pre processing)	54
5.2	File della rete stradale	57
5.3	File dei dati statistici ISTAT	59
5.4	File delle rilevazioni dei sensori	61
5.5	File dei sensori	61



# Appendice A

## Preparazione dataset degli inquinanti

Preparazione dataset degli inquinanti con selezione e media delle concentrazioni nei mesi tra Marzo e Luglio.

```
1 def clean_dataframe(df, registry, start, end, verbose=False):
2
3     registry['DateStop'] = pd.to_datetime(registry['DateStop'])
4     registry = registry[registry['Pollutant'] == pollutant]
5     active = registry[(pd.isna(registry.DateStop)) | (registry.DateStop >
6                         pd.to_datetime(end))]
7
8     df['Date'] = pd.to_datetime(df['Date'])
9
10    data = df[(df.Date > pd.to_datetime(start)) & (df.Date < pd.to_datetime(
11                           end))]
12    data = data.groupby('IDStation', as_index=False).apply(lambda g: g.mean(
13                           skipna=True))
14    data['IDStation'] = data['IDStation'].astype(int)
15
16    active = registry[(pd.isna(registry.DateStop)) | (registry.DateStop >
17                         pd.to_datetime(end))]
18
18    clean_df = pd.merge(data, active[['IDStation', 'Latitude', 'Longitude',
19                                     'Altitude']], on='IDStation', how='inner')
20
21
22    if verbose:
23        print(clean_df.shape)
24        print(clean_df.info())
25
26    return clean_df
```

Listing A.1: Preparazione dataset degli inquinanti



# Appendice B

## Kriging degli inquinanti

Kriging dei valori di concentrazione degli inquinanti.

```
1 class MultitaskGPModel(gpytorch.models.ExactGP):
2     def __init__(self, train_x, train_y, likelihood):
3         super(MultitaskGPModel, self).__init__(train_x, train_y, likelihood)
4         self.mean_module = gpytorch.means.MultitaskMean(
5             gpytorch.means.ConstantMean(), num_tasks=num_tasks
6         )
7         self.covar_module = gpytorch.kernels.MultitaskKernel(
8             gpytorch.kernels.MaternKernel(), num_tasks=num_tasks, rank=
9             num_tasks
10            )
11
12     def forward(self, x):
13         mean_x = self.mean_module(x)
14         covar_x = self.covar_module(x)
15         return gpytorch.distributions.MultitaskMultivariateNormal(mean_x,
16 covar_x)
17
18 def fit_and_predict(train_x, train_y, num_tasks, training_iterations, lats,
19 lons):
20
21     likelihood = gpytorch.likelihoods.MultitaskGaussianLikelihood(num_tasks
22 =num_tasks)
23     model = MultitaskGPModel(train_x, train_y, likelihood)
24     model.train()
25     likelihood.train()
26     optimizer = torch.optim.Adam([{'params': model.parameters()}, ], lr
27 =0.05)
28     mll = gpytorch.mlls.ExactMarginalLogLikelihood(likelihood, model)
```

```

25     for i in range(training_iterations):
26         optimizer.zero_grad()
27         output = model(train_x)
28         loss = -mll(output, train_y)
29         loss.backward()
30         optimizer.step()
31
32     grid = [ (lat, lon) for lat in lats for lon in lons]
33     grid_x = torch.from_numpy(np.array(grid)).float()
34     model.eval()
35     likelihood.eval()
36     with torch.no_grad(), gpytorch.settings.fast_pred_var():
37         predictions = likelihood(model(grid_x))
38         mean = predictions.mean
39         lower, upper = predictions.confidence_region()
40     return mean
41
42 data = pd.read_csv(in_path + pollutant + '.csv')
43 coords = [(lat, lon) for (lat, lon) in zip(data['Latitude'], data['Longitude'])]
44 train_x = torch.from_numpy(np.asarray(coords)).float()
45 lats = np.linspace(lat_bounds[0], lat_bounds[1], resolution)
46 lons = np.linspace(lon_bounds[0], lon_bounds[1], resolution)
47 grid = [ (lat, lon) for lat in lats for lon in lons]
48 grid_x = torch.from_numpy(np.array(grid)).float()
49
50 hm = pd.DataFrame({'Latitude': np.array(grid_x[:, 0]), 'Longitude': np.
51                     array(grid_x[:, 1]) })
52
53 for ds in features(pollutant):
54     train_y = torch.from_numpy(np.asarray(data[ds])).unsqueeze(1).float()
55     predictions = fit_and_predict(train_x, train_y, 1, training_iterations,
56                                   lats=lats, lons=lons)
57     hm[ds] = np.array(predictions.squeeze())

```

Listing B.1: Kriging dei valori di concentrazione degli inquinanti