

Report analisi dell'aria

David Guzman Piedrahita e Marco Vinciguerra

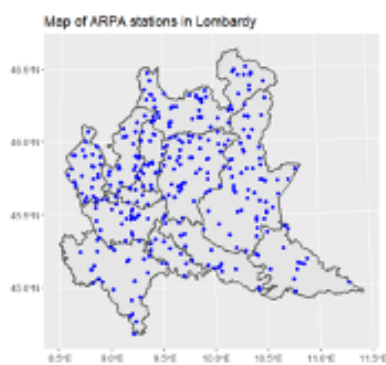
30 settembre 2021



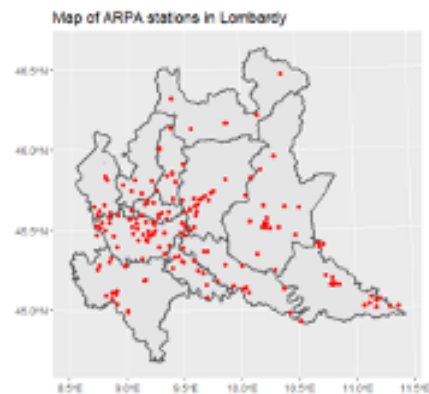
Abstract: Paper preliminare per la tesi di laurea (gestione dei dati inquinanti-meteo)

1 Introduzione

La fase iniziale del progetto consiste nell'analisi nell'arco temporale 2018-2020 dei dati forniti dal sito ARPAL relativi allo studio del NH_3 e dei particolati atmosferici PM_{10} e $PM_{2.5}$ al fine di dimostrare i risultati positivi e la diminuzione dell'inquinamento dell'aria dovuti dal Covid19 in Lombardia. Le mappe della delle stazioni che rilevano gli inquinanti e meteo sono le seguenti:



(a) Stazioni degli inquinanti



(b) Stazioni del meteo

Come si può osservare entrambe le reti di centraline non sono equidistanti tra loro e non formano una rete omogenea ma una rete eterogenea.

2 Analisi preliminare dei dati per gli inquinanti

La fase iniziale del progetto consiste nel cercare le centraline in Lombardia che misurano contemporaneamente NH_3 , PM_{10} e $PM_{2.5}$ oppure solo due di essi (sono ammessi dei dati mancanti sporadicamente per entrambi i casi). In totale ARPA Lombardia mette a disposizione 174 stazioni di qualità dell'aria e 279 stazioni meteorologiche. Le centraline che misurano tutti e 3 i regressori sono solamente 6 e sono le seguenti:

- Cremona via Fatebenefratelli (ID station: 677)
- Schivenoglia (ID station: 703)
- Sannazzaro de Burgondi Agip (ID station: 693)
- Pavia via Folperti (ID station: 642)
- Milano Pascal Citta Studi (ID station: 705)
- Moggio (ID station: 681)

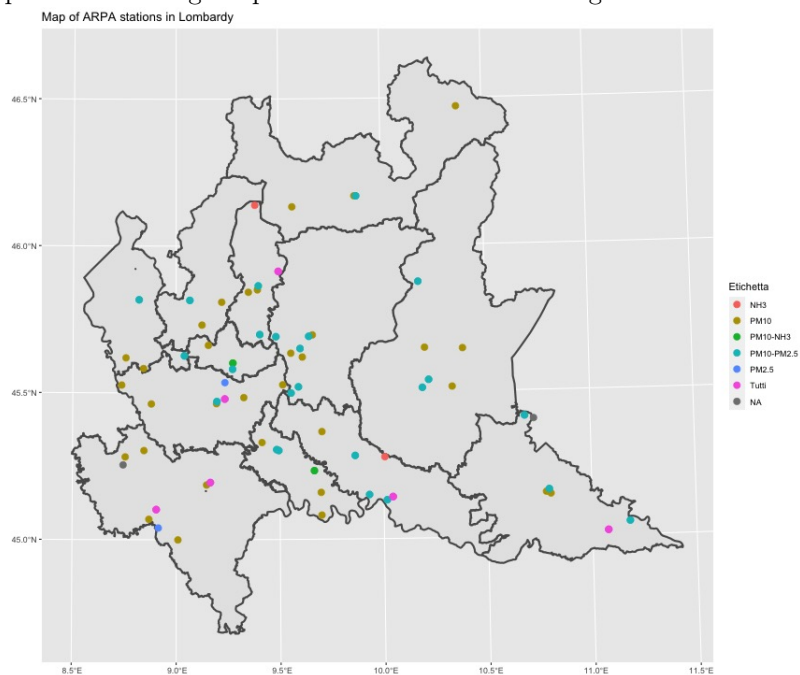
Bisogna sottolineare che tutti i risultati finora ottenuti riguardano il 2018, ma, ripetendo le stesse analisi per il 2019 e il 2020, le centraline che rilevano tutte e tre le variabili risultano essere le stesse 6. Le stazioni che ne misurano solo due regressori di interesse sono in totale 26. Per ognuno di essi è stato calcolato quanti giorni tra il 2018 e il 2020 sono assenti NH_3 , PM_{10} e $PM_{2.5}$ (singolarmente) e quanti giorni sono assenti tutti e 3 contemporaneamente (allegata con il nome `MissingFromTheBeginning.csv`).

In allegato c'è una tabella che descrive cosa viene misurato in ognuna delle centraline prese in considerazione precedentemente (`presencetableRed.csv`), i dati di queste stazioni possono comunque risultare utili qualora fosse

IDStation	NameStation	MissingAmmonia	MissingPM10	MissingPM25	MissingAllThree
1	677 Cremona Via Fatebenefratelli	2	4	7	2
2	642 Pavia Via Folperti	19	18	39	3
3	681 Moggi	61	24	27	15
4	1266 Bertinico	68	4	365	2
5	703 Schivenoglia	119	24	30	15
6	705 Milano Pascal Citta Studi	167	27	34	14
7	693 Sannazzaro De Burgondi Agip	200	19	39	19
8	1374 Monza Parco	260	85	365	54
9	548 Milano Via Senato	365	18	19	15
10	554 Saronno Via Santuario	365	0	21	0
11	560 Varese Via Copelli	365	0	5	0
12	561 Como Viale Cattaneo	365	8	9	8
13	576 Merate	365	24	22	22
14	583 Bergamo Via Meucci	365	12	12	12
15	592 Treviglio	365	17	2	1

IDStation	NameStation	PM25	PM10	Ammonia	Etichetta
63	642 Pavia Via Folperti	1	1	1	Tutti
64	677 Cremona Via Fatebenefratelli	1	1	1	Tutti
65	681 Moggi	1	1	1	Tutti
66	693 Sannazzaro De Burgondi Agip	1	1	1	Tutti
67	703 Schivenoglia	1	1	1	Tutti
68	705 Milano Pascal Citta Studi	1	1	1	Tutti
1	504 Sesto San Giovanni	1	0	0	PM2.5
29	672 Comale Voghera Energia	1	0	0	PM2.5
37	548 Milano Via Senato	1	1	0	PM10-PM2.5
38	554 Saronno Via Santuario	1	1	0	PM10-PM2.5
39	560 Varese Via Copelli	1	1	0	PM10-PM2.5
40	561 Como Viale Cattaneo	1	1	0	PM10-PM2.5
41	576 Merate	1	1	0	PM10-PM2.5
42	583 Bergamo Via Meucci	1	1	0	PM10-PM2.5
43	592 Treviglio	1	1	0	PM10-PM2.5
44	600 Lodi Viale Vignati	1	1	0	PM10-PM2.5
45	609 Casirate D'Adda	1	1	0	PM10-PM2.5
46	627 Cremona P.zza Cadorna	1	1	0	PM10-PM2.5
47	633 Soresina	1	1	0	PM10-PM2.5

necessario un volume di dati più elevato. Per ogni centralina che presenta tutti e 3 i regressori di interesse è stato fatto un plot della serie storica e in presenza di un dato mancante in corrispondenza di uno specifico giorno è stata tracciata una linea verticale blu. Il risultato della mappa della Lombardia in funzione di tutte le centrali che misurano 2 o più inquinanti che vengono presi in considerazione è il seguente:



La mancanza di dati può essere un fattore importante per gestire la fase successiva della tesi e quindi sono state scelte le stazioni con il numero di missing data inferiore. Ecco un esempio di una delle 6 migliori centraline con un numero accettabile di dati mancanti e una con un numero molto alto di dati mancanti (sempre appartenente alla lista delle 6 migliori centrali):

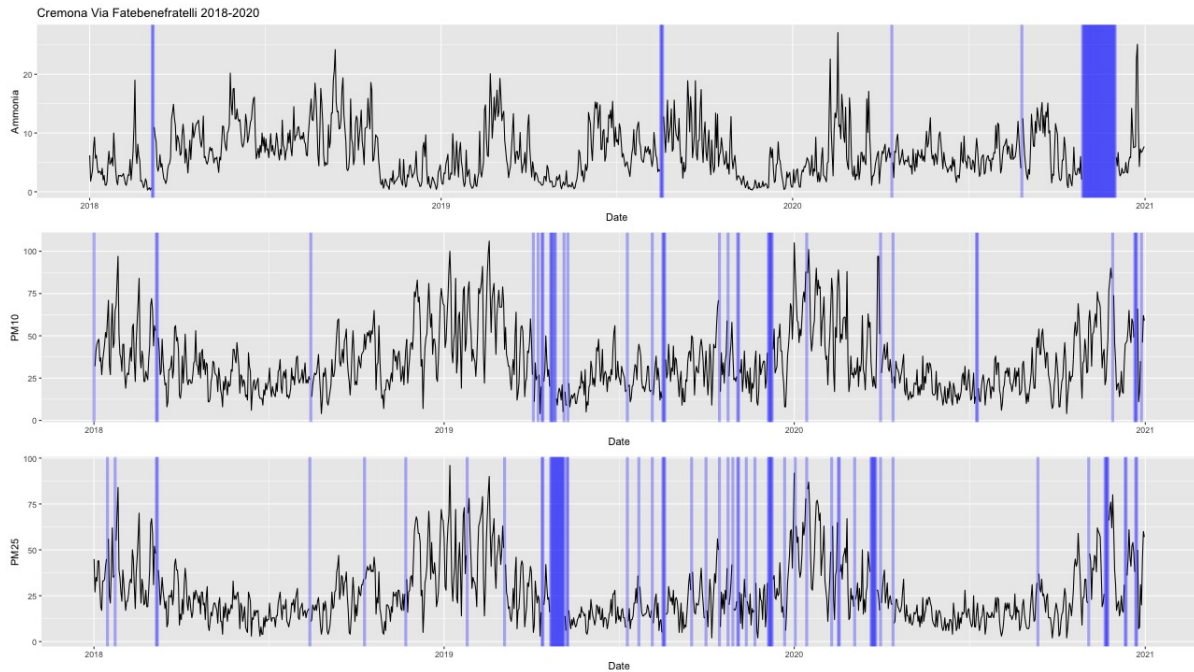


Figura 2: Cremona Via Fatebenefratelli 2018-2020, Mancanti Ammonia: 2, PM10: 4, PM25: 7, tutti e 3 contemporaneamente: 2

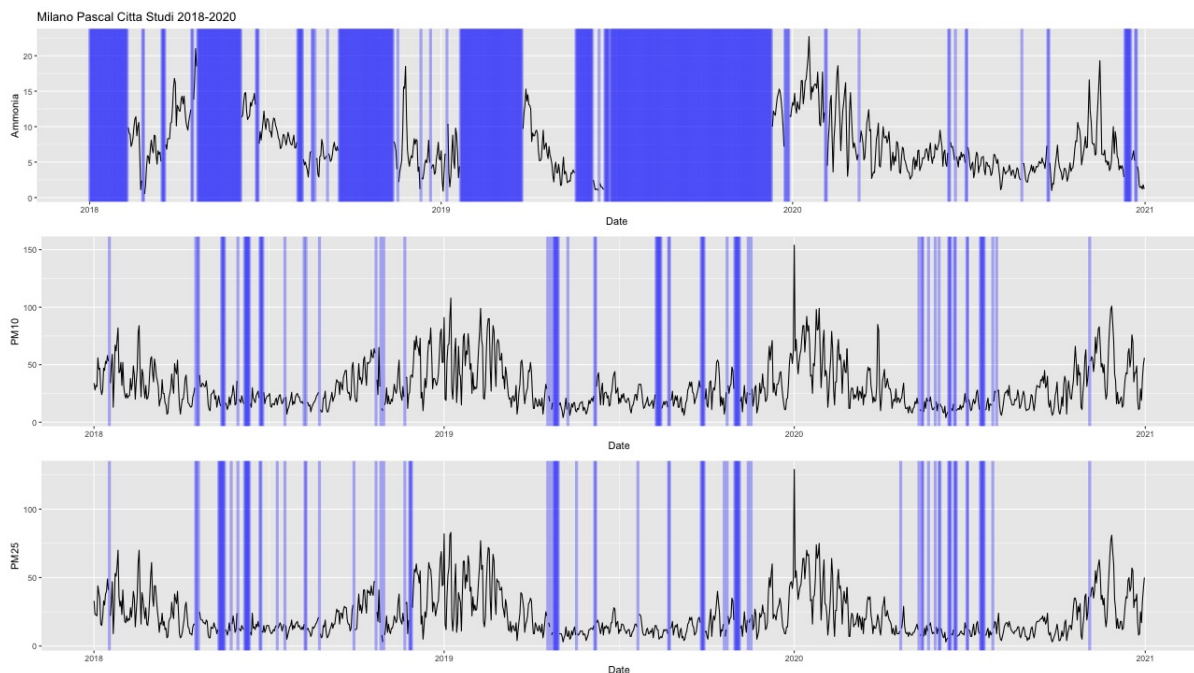


Figura 3: Milano Pascal Citta Studi 2018-2020, Mancanti Ammonia: 167, PM10: 27, PM25: 34, tutti e 3 contemporaneamente: 14

3 Analisi dei dati per il meteo

Nel caso delle centraline meteorologiche, la strategia e l'obiettivo per studiare la qualità dei dati differisce da quella proposta per gli inquinanti. Difatti, mentre nelle variabili della qualità dell'aria l'ammoniaca e il particolato erano, per così dire, i bersagli, le variabili meteorologiche hanno un ruolo meramente ausiliario. Dopotutto, anche se il loro ruolo può essere determinante, non sono le variabili che i modelli cercheranno di predire. Successivamente

è stata cercata per ogni centralina che misura gli inquinanti, le due stazioni meteo con la distanza euclidea inferiore che misurassero contemporaneamente velocità del vento (wind speed), direzione del vento (wind direction), temperatura (temperature) e precipitazioni (rainfall). Infine è stata fatto un join per unire i dati delle 6 migliori stazioni con i 2 nearest neighbor delle stazioni meteo più vicine e i relativi dati meteo (allegato come NNdata.csv).

4 Analisi missing data per il clima