

Breve report sui risultati ottenuti tesi

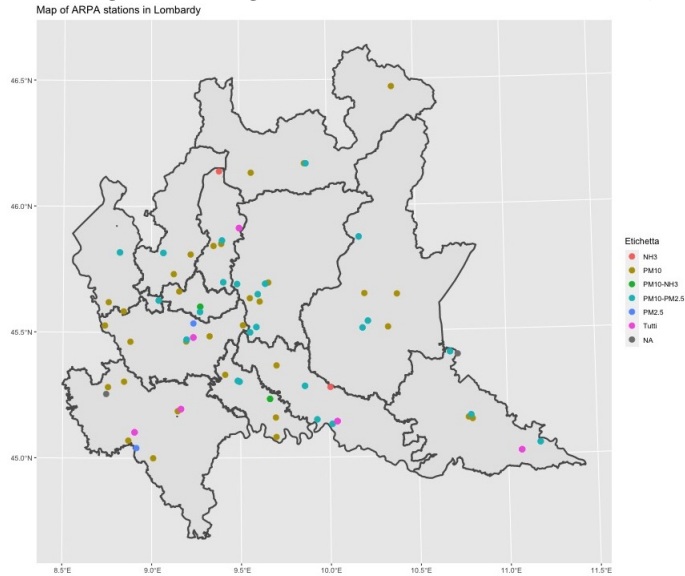
David Guzman Piedrahita and Marco Vinciguerra

19 febbraio 2022

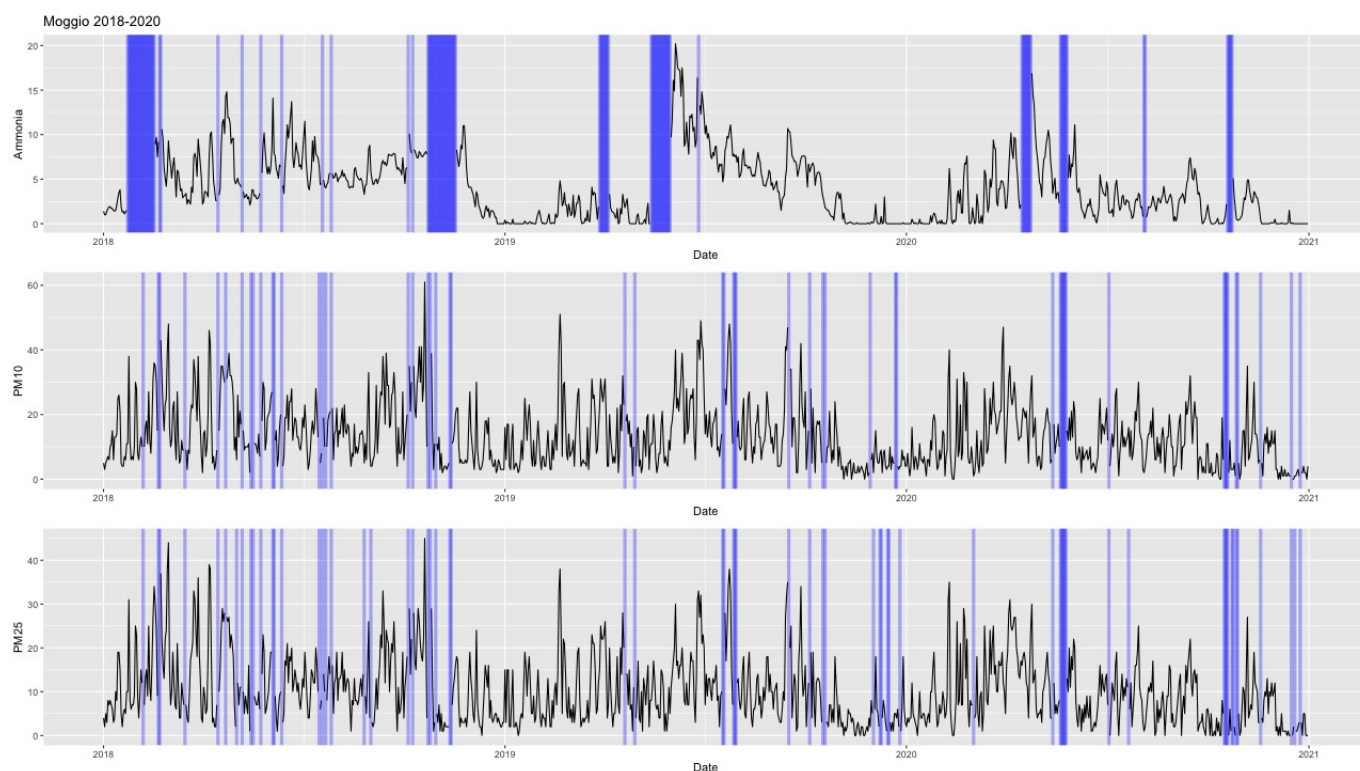
1 Introduzione e reti neurali utilizzate

Per studiare la potenziale relazione tra ammoniaca e particolato in Lombardia, sono state utilizzate diverse strategie di machine learning al fine di trovare un modello in grado di descrivere nel modo più accurato possibile, il comportamento e le interazioni tra le due sostanze in questione.

In particolare, sono state utilizzate diverse tipologie di reti neurali LSTM (long-short term memory), strumenti adatti a modellizzare serie storiche i cui valori dipendono dal passato utilizzando le informazioni degli istanti temporali precedenti. I diversi modelli implementati permettono di gestire dati mancanti nel dataset, e sono in grado di migliorare il rendimento dei modelli, stabilizzandolo e rimuovendo le informazioni superflue.



Il problema dei dati mancanti in questo tipo di analisi riveste un ruolo importante in quanto i sensori di rilevamento degli inquinanti vengono spesso disattivati causando la perdita dei dati. In questo caso, come si può vedere dalla figura seguente, i dati mancanti, sottolineati in blu, sono molto frequenti.



2 Caso studio e dati

Le 3 centraline prese in considerazione sono state Moggio, Schivenoglia e Cremona via Fatebenefratelli. Le prime 2 sono di tipo rurale mentre l'ultima è di tipo urbano. Per il caso studio sono stati utilizzati dati provenienti dalle tre centraline di qualità dell'aria e da centraline meteorologiche dislocate in tutta la Lombardia.

Usando, per esempio, i dati della centralina di Moggio dal 2014 al 2020, possiamo ottenere risultati che sottolineano una forte relazione tra il particolato e l'ammoniacale, tenendo conto anche di altre variabili atmosferiche, quali: la velocità e la direzione (tramite il quadrante cartesiano) del vento, la temperatura e le precipitazioni.

3 Risultati delle reti neurali

Le informazioni sono state utilizzate per creare e allenare il modello, ma anche per valutare le sue prestazioni quando lo stesso deve gestire nuovi dati e periodi temporali. A tal fine sono stati utilizzati gli anni dal 2014 fino al 2019 per costruire la rete neurale e l'anno 2020 per confrontare i dati previsti dal modello con i rispettivi dati reali.

Il modello utilizza l'andamento di cinque giorni consecutivi per calcolare il valore atteso del sesto giorno e, come evidenziato dal grafico, si ottiene un notevole inseguimento della previsione rispetto ai dati reali quando si modella l'andamento del particolato PM10 in funzione dell'ammoniacale. Lo stesso si può dire della relazione inversa, ovvero calcolando la previsione dell'andamento dell'ammoniacale in funzione del PM10.

In più, è da sottolineare il fatto che i dati utilizzati per il modello sono stati rilevati prima della pandemia e i successivi lockdown, ma il modello continua a offrire previsioni valide nonostante la più alta variabilità dell'attività industriale e urbana in tutto il 2020.

4 Regressione tramite reti neurali

La figura qui sotto rappresenta in arancione il modello da noi ottenuto, mentre quello in blu l'andamento reale del PM10.

