



UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE
Corso di Laurea Triennale in Informatica

Studio degli open data su meteo,
aria e COVID19 in Lombardia ed
analisi delle possibili relazioni con
Random Forest

Tesi di Laurea di:
Marco Belotti
Matr. 871440

Relatore:
Andrea Trentini

Anno Accademico 2019/2020

Indice

1	Introduzione	2
2	Normalizzazione meteorologica utilizzando Random Forest	5
3	La situazione in Lombardia	10
3.1	NO_x e NO_2	11
3.2	PM10 e PM2.5	15
3.3	CO	18
3.4	Ozono	20
3.5	Ammoniaca	22
3.6	Benzene	23
3.7	SO_2	25
4	Effetti del traffico sui principali inquinanti atmosferici	27
4.1	NO_x e NO_2	28
4.1.1	NO_x	28
4.1.2	NO_2	29
4.2	CO	30
4.3	PM10	31
4.4	Utilizzo di dati relativi al traffico	32
4.4.1	NO_x	32
4.4.2	NO_2	34
4.4.3	CO	34
4.4.4	PM10	36
5	Effetti del lockdown per il Covid-19 sulle concentrazioni degli inquinanti	37
5.1	NO_x e NO_2	37
5.1.1	NO_x	37
5.1.2	NO_2	38
5.2	PM10 e PM2.5	39
5.3	CO	41
5.4	Benzene	42
5.5	SO_2	43
5.6	Ozono	44
5.7	Ammoniaca	45
6	Appendice sulla validità dei modelli nel corso degli anni	47
7	Conclusioni	49

Capitolo 1

Introduzione

L'inquinamento atmosferico è un tema che viene spesso discusso e dibattuto, visto che respirare un'aria pulita è considerato un requisito basilare per la buona salute umana. Da anni il WHO [25] si occupa di studiare i potenziali effetti dell'inquinamento sulla salute umana, cercando di stimare che tipo di impatto possa avere un'esposizione prolungata ad alte concentrazioni.

Chiaramente non è facile riuscire a stabilire con certezza il collegamento tra causa (esposizione agli inquinanti) ed effetto (malattie e morte) poiché spesso i danni si manifestano solo a seguito di ripetute e prolungate esposizioni, così come non è facile riuscire a stabilire con certezza che una morte sia stata causata esplicitamente dall'inquinamento atmosferico.

L'inquinamento atmosferico è un fenomeno molto complicato da analizzare, poiché sia per quanto riguarda la produzione e l'emissione in atmosfera che per la dispersione entrano in gioco molti fenomeni e fattori, che hanno scale spaziali molto differenti e che è difficile riuscire ad analizzare correttamente.

Se in alcune zone del mondo, soprattutto quelle dei paesi in forte via di sviluppo come il sud-est asiatico o l'India, l'inquinamento rappresenta ancora un problema molto grave causato dall'aumento del fabbisogno energetico che è stato soddisfatto incrementando l'uso di combustibili fossili, in Europa (come in altre parti del mondo) negli ultimi 25 anni si è registrato un disaccoppiamento tra la crescita economica e le emissioni dei principali inquinanti, causato da un maggiore impegno nel loro contenimento da parte dei governi, che hanno imposto limitazioni e misure sempre più stringenti [8].

È quindi normale che sia la comunità scientifica che le autorità incaricate di decidere le misure di contenimento da mettere in campo siano interessate a capire quale sia il reale andamento delle concentrazioni di inquinanti, cercando di identificare quali possano essere i metodi più efficaci e quanto le variazioni registrate siano attribuibili specificatamente ai cambiamenti delle emissioni umane rispetto agli effetti di altri fattori come il clima che possono avere un'influenza molto maggiore sulle concentrazioni registrate [27].

Quasi tutti gli inquinanti di maggior interesse, come gli ossidi di azoto, il biossido di zolfo, il monossido di carbonio, il particolato e l'ozono sono infatti caratterizzati da una forte variabilità stagionale che, per esempio, non ci permette di confrontare valori registrati durante la stagione estiva con quelli invernali, così come l'azione di agenti atmosferici come il vento e la pioggia possono portare ad abbattimenti delle concentrazioni. Nella pianura padana, per esempio, è infatti frequente assistere, specialmente nei mesi invernali che sono già quelli più critici per le concentrazioni di molti inquinanti, a lunghi periodi senza precipitazioni che portano a grossi aumenti delle concentrazioni, che però vengono rapidamente abbattute all'arrivo della pioggia.

Per poter quindi trarre delle conclusioni oggettive sullo stato della qualità dell'aria, riuscendo ad identificare le cause delle variazioni che si registrano e quindi a capire anche l'efficacia delle misure prese, nel corso degli ultimi anni molti studi si sono avvalsi di diverse tecniche statistiche/probabilistiche per poter fare un'analisi oggettiva sugli andamenti dell'inquinamento atmosferico [8, 29, 20, 1]. In tutti gli studi si è evidenziato come considerare anche la meteorologia durante l'analisi delle concentrazioni degli inquinanti, cercando di eliminarne l'influenza, sia fondamentale per poter capire come dei cambiamenti nel quadro emissivo vengano riflessi sulle

concentrazioni misurate. Questo poi ci permette di poter fare delle analisi oggettive e quantitative più precise sui trend delle serie storiche, qualsiasi sia la località considerata, visto che i metodi utilizzati sono facilmente generalizzabili allo scenario che si vuole considerare in ogni analisi specifica.

Riuscire ad avere a disposizione tecniche e modelli che ci permettono di analizzare l'andamento delle concentrazioni e ad identificare le cause che possono cambiare questo andamento ci permette di capire quali siano le aree più critiche su cui andare ad intervenire ma ci può anche permettere di identificare gli interventi più efficaci e sui quali conviene investire maggiormente. La crescita economica e l'aumento del fabbisogno energetico rendono quindi fondamentale il controllo delle emissioni, per poter mantenere la qualità dell'aria a livelli non pericolosi per la salute umana e al tempo stesso non bloccare le attività economiche.

Qualsiasi studio sui trend degli inquinanti, così come qualsiasi conclusione che ne può derivare, non possono quindi non tener conto degli effetti della meteorologia sulle concentrazioni, poiché i risultati che si otterrebbero correrebbero il rischio di non rappresentare il cambiamento nelle emissioni antropiche ma anche una variazione nei fenomeni atmosferici e negli agenti climatici.

Il nostro obiettivo è quindi stato quello di ricercare una tecnica per normalizzare i dati sull'inquinamento rispetto alla meteorologia e al clima, ovvero che fosse in grado di eliminare l'influenza degli agenti atmosferici come pioggia, vento e radiazione solare e della stagionalità che porta ad avere condizioni più o meno favorevoli alla dispersione dalle concentrazioni registrate degli inquinanti, in modo da arrivare poi ad avere una serie storica su cui poter fare analisi sull'efficacia delle misure prese nel corso degli anni per il contenimento delle concentrazioni.

Un metodo comune per riuscire a fare questa "pulizia" dei dati si basa sulla creazione di modelli statistici che tramite l'uso di diverse variabili *predittive*, come le misure meteorologiche su vento, precipitazioni e radiazione solare e variabili temporali legate al giorno dell'anno e alla data, siano capaci di prevedere i valori delle concentrazioni che si sarebbero misurate per un inquinante in base ai valori di ciascuna. Col termine variabile predittiva, infatti, si intende una variabile in grado di spiegare, anche solo parzialmente, cambiamenti nel valore di un'altra variabile, permettendo quindi di fare previsioni su di esso basandosi sul valore della prima. Se i modelli che riusciamo a creare si rivelano abbastanza precisi nell'attività predittiva e in grado di spiegare buona parte della varianza delle concentrazioni allora possiamo usarli con una certa confidenza per eliminare gli effetti della meteorologia dalle concentrazioni di inquinanti.

Questo procedimento viene però complicato dal fatto che le dinamiche che portano la meteorologia a modificare le concentrazioni varino molto a seconda della località e che quindi differenti contesti, come ad esempio possono essere una città rispetto ad una località montana o situata a fondo valle, necessitino trattazione specifiche a seconda del caso. Inoltre è chiaro come le variabili meteorologiche tra di loro siano collegate e correlate, cioè che cambiamenti nei valori di una vengano riflessi anche sulle altre, e questo porta all'insorgere di una serie di problemi come la normalità, la multi collinearità e l'indipendenza che rendono la trattazione con diversi metodi molto difficile e complicata [14].

Nel corso degli ultimi tre decenni nel mondo dell'informatica ha preso campo un'area detta *machine learning* (ML), che fa uso dei computer e della statistica per creare dei modelli predittori che potessero porsi come alternativa ai classici metodi statistici già esistenti [17]. Sono quindi state sviluppate tecniche di ML non parametriche, ovvero che non necessitano di fare alcun tipo di assunzione statistica sulle variabili considerate, che, grazie all'uso dei computer per analizzare grosse moli di dati, permettono di arrivare alla creazione di modelli con buone capacità predittive anche senza occuparsi dei problemi che invece caratterizzano gli approcci puramente statistici. Questi ultimi richiedono infatti, prima di poterli applicare, di fare una serie di assunzioni sui dati considerati, come la linearità della relazione con i valori della variabile da predire o l'assenza di autocorrelazione¹, e su quali siano le relazioni che li legano ed inoltre molte volte anche la matematica necessaria è complicata e richiede particolare attenzione nella trattazione per essere sicuri di ottenere dei

¹Per autocorrelazione si intende quel fenomeno per cui i valori di una variabile sono dipendenti dai valori assunti in precedenza dalla stessa

risultati attendibili. Gli algoritmi di machine learning, invece, sfruttando la capacità computazionale degli elaboratori, permettono di arrivare ad ottenere dei modelli altrettanto affidabili senza però doversi preoccupare dei diversi aspetti richiesti dalla statistica e che spesso ne limitano l'applicabilità. Inoltre, molte volte, l'obiettivo dei modelli statistici è quello di inferire il modello dei dati per il caso considerato, ovvero capire come diversi fattori influenzino quantitativamente i valori di una variabile di interesse, mentre nel machine learning l'obiettivo principale solitamente è quello di essere capaci di fare previsioni precise sui valori di tale variabile in base a quelli dei diversi fattori considerati [7].

ARPA Lombardia ² possiede, sul territorio regionale, una rete di 85 stazioni fisse che, per mezzo di analizzatori automatici, sono in grado di fornire rilevazioni ad intervalli di tempo regolari, permettendoci di creare quindi un dettagliato catalogo dei valori delle concentrazioni nel corso del tempo. Tutte queste rilevazioni vengono poi rese accessibili liberamente tramite archivi online, per permetterne la libera consultazione da parte di tutti i cittadini.

Nel nostro studio abbiamo quindi fatto uso di questi dati, unitamente a quelli delle variabili meteorologiche sempre resi disponibili da ARPA Lombardia, per creare tramite l'uso di una tecnica di machine learning chiamata *Random Forest* [6] dei modelli predittivi che ci permettano di arrivare a fare questa normalizzazione delle concentrazioni di inquinanti rispetto alla meteorologia e quindi di analizzare come si sia evoluta la situazione nella nostra regione nel corso degli anni dal 2012 ad oggi, cercando di capire quali siano stati i provvedimenti e gli eventi più efficaci nell'abbattimento delle concentrazioni.

Nel capitolo 2 viene introdotto Random Forest e viene mostrato come questa tecnica possa essere utilizzata per eliminare gli effetti della meteorologia e della stagionalità dall'andamento delle concentrazioni.

Nel capitolo 3 viene analizzata, utilizzando la tecnica appena discussa, la situazione regionale andando a controllare l'andamento delle serie dei principali inquinanti nei capoluoghi di provincia lombardi.

Nel capitolo 4 verranno presentate le analisi svolte sull'influenza del traffico sulle concentrazioni di tre inquinanti: ossidi di azoto, particolato atmosferico (PM10 e PM2.5) e monossido di carbonio. Nel capitolo 5 verrà presentata un'analisi del periodo di diffusione dell'epidemia di Covid-19 e di quali siano stati gli effetti del conseguente *lockdown* svoltosi nei mesi di marzo ed aprile sulle concentrazioni.

Nell'appendice finale, inserito a seguito della lettura di un articolo [3] che ha utilizzato alcune tecniche di machine learning, tra cui anche Random Forest, ed i nostri stessi dataset per provare a fare previsioni sulla qualità dell'aria della città di Milano, è presentata una verifica fatta sulla validità nel corso degli anni dei modelli creati.

²https://www.arpalombardia.it/Pages/ARPA_Home_Page.aspx

Capitolo 2

Normalizzazione meteorologica utilizzando Random Forest

L'atmosfera è lo strato gassoso che ricopre esternamente la terra come un involucro ed è costituita principalmente da azoto e ossigeno, mescolati con piccole quantità di altri gas e pulviscolo. La porzione più bassa, che sta a diretto contatto col suolo, è definita strato limite planetario (Planetary Boundary Layer, o PBL, in inglese) e sia la sua struttura che l'evoluzione hanno influenze fondamentali sul comportamento degli inquinanti atmosferici di maggior interesse [9].

A causa della sua struttura, infatti, gli inquinanti emessi in atmosfera rimangono bloccati al suo interno e variazioni sull'altezza, che può variare da qualche centinaio di metri a qualche chilometro, causate da agenti atmosferici come il vento, la temperatura e la pressione atmosferica, hanno l'effetto di disperdere o aumentare le concentrazioni di tali inquinanti [35].

La forte variabilità stagionale che si riscontra su tanti inquinanti è infatti direttamente collegata al fatto che nei mesi invernali (così come nelle ore notturne) l'altezza dello strato di rimescolamento sia ridotta e porti quindi ad un aumento delle concentrazioni.

Per poter analizzare i dati sull'inquinamento atmosferico volendo capire come le misure messe in atto e le attività antropiche abbiano influenzato l'andamento delle concentrazioni nel corso degli anni non possiamo trascurare questa grossa influenza da parte della meteorologia e del clima, altrimenti non sarebbe possibile stabilire con certezza quanta influenza possa avere avuto un determinato provvedimento rispetto a quella data dalla variazione delle condizioni atmosferiche.

Nel corso degli anni diversi studi hanno provato ad affrontare la questione usando metodi che portassero ad ottenere dei modelli statistici in grado di capire e modellare le relazioni tra inquinanti e condizioni atmosferiche e, indipendentemente dal metodo usato, si è visto come variabili legate alla meteorologia, all'altezza dello strato di rimescolamento ed alla stagionalità siano in grado di spiegare con una certa precisione gli andamenti delle serie di inquinanti. Segal ha dimostrato come fosse possibile applicare un modello numerico per simulare la circolazione di aria nella zona della baia di Chesapeake (USA) per poterne predire la qualità dell'aria [32]. Eagleman nel suo libro illustra diversi metodi matematici applicabili per lo studio delle concentrazioni di inquinanti e degli effetti che ha su di esse la meteorologia [10]. Aldrin e Haff propongono un modello per la stima delle concentrazioni dei principali inquinanti utilizzando i valori delle variabili meteorologiche e dei volumi di traffico [2]. Un altro studio si è invece concentrato sull'influenza della meteorologia sul PM10 e su come essa complichino l'analisi dei reali trend di un inquinante [4]. Cattani, in uno studio del 2014, ha proposto un metodo matematico per il calcolo dei trend delle concentrazioni di un inquinante, eliminando l'influenza della stagionalità [8].

I processi atmosferici, però, sono molto complessi, non lineari e spesso si riscontra collinearità tra le osservazioni di variabili diverse, una situazione in cui i valori di una possono essere approssimati in modo preciso da quelli di un'altra, e questo rende la modellazione grazie all'uso di metodi parametrici, cioè quelli basati su metodi

puramente matematici come la regressione lineare, molto complicata poiché questi fattori influiscono molto sulla precisione e sull'affidabilità dei risultati ottenuti [14].

Il machine learning, grazie all'uso di metodi non parametrici, ovvero metodi che non hanno bisogno di fare nessun tipo di assunzione sui dati considerati o sulle relazioni che intercorrono tra le diverse variabili, che quindi si adattano perfettamente ad un'analisi così complessa come quella degli inquinanti e della loro dispersione in atmosfera, ci permette di arrivare ad estrarre dei modelli in grado di simulare in modo abbastanza preciso le relazioni tra le diverse variabili in gioco. Grange e Carslaw, tramite l'uso di un metodo di ML non parametrico per affrontare il problema della normalizzazione meteorologica, sono giunti alla conclusione che la libertà sui valori di input per i modelli, l'efficiente gestione della collinearità delle variabili e della loro interazione rendano il loro uso efficace per trattare tale problema e che la loro applicazione sia abbastanza flessibile per un buon range di analisi sull'efficacia dei provvedimenti presi in materia di qualità dell'aria [12].

Una volta in possesso di uno strumento che sappia spiegarci come variazioni dei valori di una variabile influenzino quelli di un'altra, che nel nostro caso significa arrivare ad ottenere un modello capace di dirci come cambiamenti nelle condizioni atmosferiche e nella stagionalità portino a cambiare le concentrazioni registrate per gli inquinanti, diventa semplice anche arrivare ad eliminare la loro influenza. Le variabili predittrici, ovvero quelle delle quali vogliamo capire ed eliminare l'influenza dalle concentrazioni degli inquinanti, utilizzate nei nostri modelli sono state: la temperatura, le precipitazioni registrate, il numero di giorni passati dalle ultime precipitazioni, l'umidità relativa, la radiazione globale e direzione e velocità del vento. Inoltre sono state inserite anche il giorno dell'anno ed il giorno della settimana, due variabili temporali che potessero fare da proxy per gli andamenti tipici stagionali e settimanali (dovute alle maggiori attività antropiche, soprattutto nelle città, durante i giorni feriali) mostrati dalle concentrazioni di molti inquinanti.

Analizzando la serie di ogni inquinante e riportandola, grazie all'uso di un modello appositamente costruito basandosi sui dati di tale serie, a condizioni meteorologiche e di stagionalità uguali per tutti i giorni possiamo quindi arrivare ad ottenere una serie il cui andamento sia indipendente dall'influenza data da questi fattori.

I metodi non parametrici esistenti nel campo del machine learning adatti ad affrontare problemi di regressione, ovvero problemi in cui vogliamo fare previsioni su una quantità numerica, come nel nostro caso vogliamo fare per le concentrazioni degli inquinanti, partendo dai valori di altre grandezze (le variabili meteorologiche e stagionali nel nostro caso), sono diversi ed ognuno è più adatto ad affrontare determinati problemi rispetto ad altri. Noi abbiamo scelto di utilizzare Random Forest, un metodo d'insieme basato sull'uso di molteplici alberi di decisione, inventato da Leo Breiman nel 2001 [6].

Gli alberi di decisione sono strutture ad albero, composte quindi da nodi e rami, che nel ML sono usate per creare dei modelli capaci di prevedere il valore di una variabile in base a quello assunto da una o più variabili predittrici. Questi alberi sono creati in modo che tutti i nodi che non siano foglie, compresa quindi anche la radice, contengano dei test da effettuare sui valori dei predittori, mentre nelle foglie siano presenti i risultati delle previsioni fatte da tali strutture. Per ottenere la previsione del modello relativa ad un determinato set di valori delle variabili predittrici in uso si parte quindi dalla radice, in cui troveremo appunto un test da eseguire su tali valori. I rami in uscita dal nodo rappresentano i possibili esiti di quel test e ci indicano la direzione in cui proseguire la nostra discesa all'interno della struttura. Se giungiamo ad un nodo intermedio, contenente nuovamente una condizione, ripeteremo ricorsivamente lo stesso passaggio svolto in precedenza fino ad arrivare ad una foglia, la quale conterrà il valore della previsione fatta dall'albero per il set di dati ricevuto in input. In figura 2.1 è riportato un esempio di albero di decisione che può essere usato per fare previsioni sul sesso di una persona in base alla sua altezza ed al suo peso. Chiaramente non è detto che l'albero faccia sempre la previsione corretta ed infatti esistono poi una serie di test per verificare la precisione della struttura creata.

Gli alberi di decisione sono una tecnica valida poiché non fanno nessun tipo di assunzione sulla struttura delle relazioni che sono presenti tra le diverse variabili utilizzate, così come permettono collinearità e sono capaci di scartare le misure che risultano irrilevanti rispetto alla variabile da predire. In fase di costruzione, grazie all'uso di specifici algoritmi, questi alberi vengono creati in modo che le condizio-

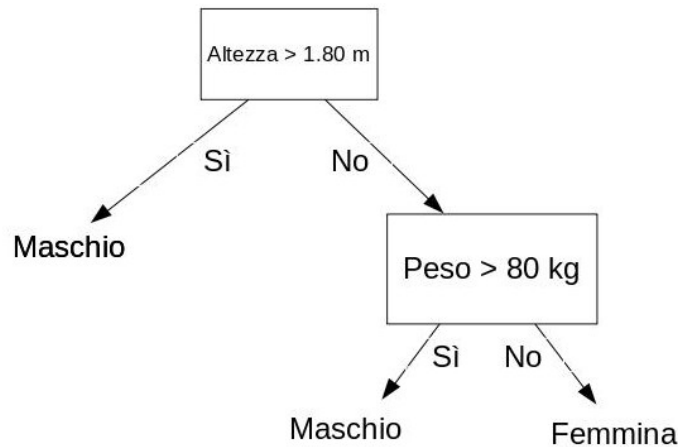


Figura 2.1: Albero di decisione per prevedere il sesso di una persona in base alla sua altezza ed al suo peso

ni presenti nei nodi intermedi dell'albero, e che ci guidano verso il risultato della previsione, riescano a minimizzare il più possibile l'errore delle previsioni fatte dal modello sui dati usati in tale fase.

Queste strutture, specialmente se le si lascia crescere liberamente in profondità, soffrono molto di un problema detto *overfitting*, ovvero che i modelli ottenuti si adattano perfettamente ai dati utilizzati in fase di costruzione, ma quando poi si prova a verificare le loro prestazioni su dati nuovi risultano scarse, poiché le relazioni individuate dal modello risultano essere troppo contaminate da rumore presente in tali dati [5]. Questo accade perché, se non si pone un limite alle dimensioni degli alberi, gli algoritmi usati per crearli riescono a costruire strutture in grado di trattare in modo preciso ogni singolo caso contenuto nel set di dati usato in fase di costruzione, ma non è detto che poi tali strutture siano in grado di rilevare correttamente le relazioni presenti nei dati non permettendoci quindi di ottenere risultati attendibili.

Random Forest, grazie all'uso di un numero elevato di alberi di decisione (da qui il termine forest), ciascuno costruito a partire da un sottoinsieme estratto casualmente dai dati iniziali, riesce a controllare il problema dell'*overfitting* mantenendo però tutti i vantaggi già presenti per gli alberi di decisione e che quindi lo rendono un algoritmo che si adatta perfettamente al tipo di lavoro che vogliamo fare. Random Forest, infatti, non fa nessun tipo di assunzione sulla struttura dei dati e delle relazioni che intercorrono tra ciascuna variabile, che è un grosso vantaggio quando stiamo trattando un problema complesso come quello della dispersione di inquinanti in atmosfera. È inoltre capace di selezionare in modo automatico le variabili più importanti, che poi userà maggiormente nell'attività predittoria, senza però porre vincoli né sul numero di grandezze utilizzate né su eventuali relazioni che possono intercorrere tra ciascuna. Random Forest ha inoltre il vantaggio di poter utilizzare contemporaneamente sia variabili numeriche che categoriche, senza richiedere particolari pre-trattazioni come invece avviene con altri metodi.

Random Forest ci permette di analizzare come ogni modello operi per fare delle scelte, tramite strumenti come l'importanza delle variabili, che ci indica quali risultino più utilizzate ed influenti sulla precisione dei risultati ottenuti, o i *partial dependence plots*, che ci mostrano come cambiamenti nei valori di una singola variabile influenzino le previsioni fatte dal modello. Questi sono strumenti molto utili perché ci permettono di studiare, oltre ai risultati forniti dal modello, anche come ci si sia arrivati potendo così approfondire le relazioni e le dinamiche individuate.

Di ogni modello sono state valutate le prestazioni grazie all'uso della tecnica del *validation set*, una tecnica molto conosciuta che grazie all'utilizzo di una parte dei dati iniziali, che viene sottratta casualmente ed in modo automatico a quelli usati per creare il modello e la cui dimensione è liberamente impostabile, permette di verificare quanto siano precise le previsioni fatte da un modello confrontandole con valori reali. Per la costruzione dei nostri modelli è stato usato l'80% di tutte le

osservazioni disponibili, mentre il restante 20 è stato usato per la validazione.

L'algoritmo Random Forest permette di personalizzare alcuni aspetti del suo funzionamento per cercare di ottimizzarne le performance grazie all'uso di alcuni iperparametri, che vanno settati prima della creazione del modello e ne determinano le regole da applicare in tale fase. Una delle prime attività svolte, dopo aver verificato che fosse effettivamente possibile utilizzare Random Forest per ottenere la normalizzazione meteorologica, è stata proprio una ricerca per individuare i valori migliori da assegnare ai diversi iperparametri in modo da ottenere il miglior compromesso possibile tra performance e affidabilità. Il numero di alberi creato in ciascuna foresta è stato di 300, la profondità massima di ogni singolo albero è stata fissata a 20 ed il numero di variabili da utilizzare per creare le condizioni presenti nei nodi intermedi a 5.

Per ottenere modelli che fossero i più fedeli possibile rispetto alla situazione reale da descrivere si è scelto di trattare ogni località ed ogni inquinante creando di volta in volta un modello apposito, costruito sfruttando sempre il procedimento descritto ed utilizzando le serie storiche registrate nella località considerata, che potesse comprendere al meglio le dinamiche caratteristiche del luogo.

Quello che abbiamo quindi ottenuto da Random Forest sono modelli che, ricevuta in ingresso una data ed i valori delle condizioni meteorologiche e atmosferiche, sono in grado di prevedere con una certa precisione quale sarebbe stata la concentrazione dell'inquinante misurata per quel giorno se si fossero verificate tali condizioni meteorologiche.

Una volta in possesso di uno strumento di questo tipo diventa facile capire come eliminare l'influenza della meteorologia e delle condizioni atmosferiche dalle concentrazioni degli inquinanti, poiché, per ogni giorno della serie considerata, basta far fare al nostro modello molte previsioni, variando di volta in volta le condizioni atmosferiche, e poi calcolarne la media, che rappresenta quindi la concentrazione dell'inquinante per tale data se si fossero verificate condizioni medie. Per ottenere le serie utilizzate nelle nostre analisi per ogni giornata sono state fatte 1000 previsioni, variando di volta in volta i valori delle variabili predittive, poiché si è visto che pur aumentando ulteriormente questo numero non si riuscisse ad ottenere risultati più precisi.

Matematicamente possiamo definire le variabili predittive usate per la costruzione dei modelli come X_1, \dots, X_p e la concentrazione dell'inquinante trattato come Y . Possiamo quindi dire, visto che le variabili predittive sono variabili che influenzano l'andamento di quella che dobbiamo predire, che

$$Y = f(Data, X_1, \dots, X_n) + \epsilon \quad (2.1)$$

dove f è la funzione, chiaramente sconosciuta, che descrive come cambiamenti nei valori delle variabili predittive e nella data considerata influenzino il valore di Y ed ϵ un termine d'errore, indipendente da Y e con valore medio 0 [17]. Un modello che usi tali variabili predittive per cercare di predire il valore di Y è quindi definibile come $\hat{f}(Data, X_1, \dots, X_n)$, ovvero una stima della funzione f , i cui valori sono definiti \hat{Y} e rappresentano i valori di Y predetti dal modello in base alla data considerata ed ai valori delle variabili predittive. Una volta in possesso di un modello, per il calcolo della concentrazione normalizzata di una data D , che indicheremo con c_{normD} , è stata quindi applicata la formula seguente:

$$c_{normD} = \frac{\sum_{i=1}^{1000} \hat{f}(D, x_{1_i}, \dots, x_{p_i})}{1000} \quad (2.2)$$

in cui D rappresenta la data per la quale si vuole calcolare la concentrazione, che viene fissata allo stesso valore per tutte le previsioni, e x_{1_i}, \dots, x_{p_i} rappresentano i valori degli altri predittori che ogni volta sono estratti casualmente e con reimmissione dalle loro serie storiche.

Ripetendo questo passaggio per tutti i giorni di una serie e per diversi inquinanti e località siamo così arrivati ad ottenere delle serie il cui andamento fosse stato "pulito" dall'influenza della meteorologia, permettendoci quindi di analizzarne meglio sia il trend che di verificare in modo più preciso quale sia stato l'impatto delle misure di contenimento prese.

Fondamentale per il funzionamento della nostra tecnica era aver a disposizione molti dati, sia per quanto riguarda le serie delle concentrazioni di inquinante che per le variabili meteorologiche scelte. ARPA Lombardia, che da anni si occupa di monitoraggio dell'aria su tutto il territorio regionale, grazie alla sua vasta rete di rilevazione raccoglie tutte queste misure, che sono poi rese disponibili per il libero download da parte di chiunque [21, 22]. Mettere questi dati a disposizione dei cittadini, favorendone la loro fruizione, permette ad ognuno di poter fare delle valutazioni sia sulla reale qualità dell'aria che si respira tutti i giorni che sull'operato degli enti governativi, acquisendo una miglior conoscenza su un tema molto delicato e discusso. In tal senso l'apertura di ARPA Lombardia, che dal 2017 ha semplificato l'accesso ed il download dei dataset, è fondamentale e ha segnato una svolta rispetto ad un passato in cui la fruizione non era certamente facilitata [45]. In assenza di questi dati non sarebbe possibile fare analisi di questo tipo, poiché per imparare le diverse relazioni tra le variabili in gioco i nostri modelli hanno per forza bisogno di grandi quantità di informazioni, che possano descrivere tutte le diverse situazioni riscontrabili nella realtà. L'Italia, secondo un report del 2019 in cui sono presentati i risultati di un sondaggio europeo sulla percezione della qualità dell'aria e dei relativi problemi da parte dei cittadini [11], è il secondo stato per maggioranza di persone che non si sentono adeguatamente informate sulla questione (69% del totale degli intervistati). Inoltre, sempre nello stesso studio, l'Italia risulta essere il quinto stato per maggioranza di persone che pensano che la situazione sia peggiorata nel corso degli ultimi anni (74% degli intervistati). Considerato che oltre il 90% della popolazione ritiene seri i problemi salutarie legati all'inquinamento atmosferico, si capisce come sia importante informare maggiormente la popolazione sull'argomento, così da permettere di poter prendere maggiore coscienza dei reali problemi e della situazione, ed avere libero accesso ai dati è sicuramente uno dei punti di partenza fondamentali per raggiungere questo obiettivo.

Il processo completo tramite il quale si è arrivati ai risultati presentati nelle prossime sezioni è iniziato con una fase di raccolta e pre-elaborazione dei dati messi a disposizione da ARPA Lombardia relativi all'inquinamento atmosferico e alla meteorologia, creando per ciascun sensore, che sia esso relativo al meteo o all'inquinamento, un file CSV specifico contenente le medie giornaliere dei valori registrati nel corso degli anni. Dopodiché, per ogni località considerata nel corso delle analisi, si è creato un modello specifico utilizzando i dati delle concentrazioni di inquinante, della temperatura, dell'umidità, delle precipitazioni registrate, della radiazione globale e di direzione e velocità del vento dei sensori presenti in tale località, così che i dati utilizzati per ciascun modello rispecchiassero il più fedelmente possibile le reali condizioni atmosferiche e meteorologiche del luogo in esame. I buchi presenti nei dati, causati dalla mancanza di rilevazioni per determinate date, sono stati riempiti con il valore medio di ciascuna variabile, tranne che per la concentrazione di inquinante, in cui assenza si è scelto di non considerare la data nel calcolo della serie normalizzata. Per le foreste dei nostri modelli il numero di alberi presenti è stato fissato a 300, la profondità massima del singolo albero a 20 ed il numero di variabili utilizzate per creare le condizioni presenti nei nodi interni di ciascuna struttura a 5, poiché tramite una serie di test si è stabilito che questi fossero i valori che offrono il miglior compromesso prestazioni/precisione possibile nel nostro caso. Una volta entrati in possesso del modello si è poi passati al calcolo della serie normalizzata, facendo fare, per ciascuna data del periodo considerato, 1000 previsioni sulla concentrazione dell'inquinante, variando di volta in volta le condizioni meteorologiche considerate estraendole casualmente dall'insieme dei dati di partenza, per poi calcolarne la media che rappresenta il valore della concentrazione normalizzata calcolato dalla nostra tecnica.

Tutto il lavoro di raccolta ed analisi dei dati è stato fatto utilizzando il linguaggio open source Python. Per l'utilizzo dell'algoritmo Random Forest ci si è appoggiati all'implementazione presente nella libreria Scikit-learn [26]. Tutto il codice usato per il recupero dei dati, la loro elaborazione, la creazione dei modelli con Random Forest e le analisi sui risultati ottenuti è liberamente accessibile ed utilizzabile nel repository pubblico presente su GitHub [44].

Capitolo 3

La situazione in Lombardia

Inizialmente sono state affrontate una serie di analisi in cui, tramite l'applicazione di Random Forest per ottenere la normalizzazione dei dati sull'inquinamento relativamente alle condizioni meteorologiche, si è cercato di delineare il quadro della situazione per quanto riguarda i principali inquinanti atmosferici in Lombardia sfruttando i vantaggi derivanti dall'aver eliminato dalle concentrazioni l'influenza della meteorologia.

Sono state analizzate le serie storiche normalizzate dei principali inquinanti atmosferici, usando i dati registrati nei dodici capoluoghi di provincia lombardi. Per ogni inquinante di ogni capoluogo è stato creato un modello apposito, così da poter analizzare ogni situazione con un modello specifico creato appositamente per cercare di stimare al meglio le dinamiche locali degli eventi climatici e la loro relazione con gli inquinanti considerati.

Il requisito di iniziare a considerare le serie storiche con cui creare i modelli solo a partire dalla data di messa in funzione del sensore più recente tra tutti quelli delle variabili coinvolte nella loro creazione ha purtroppo limitato il periodo considerato dalle nostre analisi, poiché per la maggior parte delle località le rilevazioni della direzione e velocità del vento, che è sicuramente uno dei fattori più influenti sull'andamento delle concentrazioni, sono iniziate solo nell'ultimo decennio. Una possibile soluzione a questo problema poteva essere quella di usare i valori medi di ciascuna variabile per il periodo in cui non fossero disponibili le registrazioni, ma questo, specialmente se fatto su periodi estesi, portava di fatto ad alterare le prestazioni e l'affidabilità del nostro modello poiché causava l'introduzione di un numero consistente di dati "fittizi" che compromettevano l'individuazione delle diverse dinamiche che governano gli andamenti delle concentrazioni.

I modelli ottenuti nel corso delle analisi hanno avuto tutte prestazioni molto buone, specialmente quelli relativi all'ozono, con le polveri sottili che invece si sono rivelate l'inquinante con cui fanno più fatica ad adattarsi, evidenziando come per questo tipo di inquinanti, che arrivano a percorrere anche grosse distanze, le sole variabili delle rilevazioni meteorologiche di superficie e temporali considerate non siano sufficienti per spiegarne completamente l'andamento, che è governato da eventi su scale molto più grandi di quelle descritte dalle nostre variabili [13].

Come si può osservare in figura 3.1, tra tutti i modelli ottenuti la variabile più importante è risultata essere la temperatura, seguita dalla data e dal giorno dell'anno. Questa considerazione però non dovrebbe meravigliarci se pensiamo al collegamento che esiste tra ciascuna e l'inquinamento: la temperatura serve sia ad identificare i cicli stagionali che eventuali periodi più caldi o freddi della media in cui le emissioni di determinati settori, come ad esempio i riscaldamento domestici, potrebbero aumentare considerevolmente, la data ci permette di identificare l'evoluzione nel corso degli anni dei valori registrati, e quindi la presenza di eventuali trend o andamenti particolari, e il giorno dell'anno serve per identificare la stagionalità caratteristica degli inquinanti. Alle variabili legate alle precipitazioni, ovvero la quantità registrata e il numero di giorni trascorsi dall'ultimo evento, è sempre stata assegnata un'importanza abbastanza bassa. È bene ricordare, però, che l'importanza delle variabili è calcolata come il totale della diminuzione dell'impurità dei nodi della foresta, pesata per la probabilità di raggiungere ciascun nodo. Questa è quindi una misura che cerca di stimare come il nostro modello lavori per fare le previsioni, ma non indica

assolutamente che tipo di relazioni ci siano tra la variabile considerata e quella da predire. Quindi, anche se nei nostri modelli alle precipitazioni viene associata una bassa importanza, non significa che non abbiano influenza sulle concentrazioni registrate, ma semplicemente che la variabile per lui è meno utile nel fare previsioni più precise.

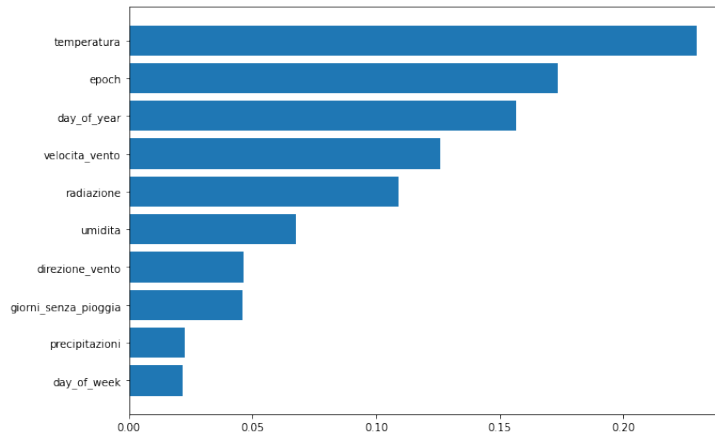


Figura 3.1: Media delle importanze delle variabili predittive nei modelli generati durante le nostre analisi

3.1 NO_x e NO_2

Nella categoria degli ossidi di azoto sono considerati due inquinanti di diverso interesse e che quindi è bene analizzare separatamente. Il monossido di azoto, infatti, alle concentrazioni tipiche misurate non risulta pericoloso né per la salute umana né per la vegetazione; il biossido di azoto, invece, desta sicuramente una maggior preoccupazione per i suoi effetti sulla salute. Per questo motivo sono state considerate ed analizzate separatamente le serie dei sensori misuratori di NO_x e di quelli dedicati solo al biossido [25].

Inizialmente sono state analizzate le serie degli ossidi di azoto. Quando si considerano questi inquinanti è bene ricordare che il monossido ha principalmente origine primaria, mentre il biossido si genera in atmosfera grazie all'ossidazione del primo e quindi ha origine principalmente secondaria. All'emissione, infatti, si stima che il biossido sia circa il 5-10% del totale per questa categoria [23]. Entrambi sono emessi in atmosfera da processi di combustione ad alta temperatura (impianti di riscaldamento, motori dei veicoli, combustioni industriali, centrali di potenza, ecc..) e per ossidazione dell'azoto presente in atmosfera (o dei suoi composti contenuti nei combustibili utilizzati). Il biossido di azoto, oltre che alla pericolosità per quanto riguarda la salute di persone e vegetazione, ha un ruolo fondamentale nella formazione dello smog fotochimico in quanto è l'intermediario per la produzione di inquinanti secondari come l'ozono.

Secondo l'inventario regionale INEMAR 2017 [15] la categoria maggiormente responsabile per le emissioni di questi inquinanti è il trasporto su strada, con un contributo pari al 50% del totale delle emissioni, seguita dalle combustioni industriali e civili (con circa il 15% di contributo ciascuna). Sempre secondo l'inventario il combustibile maggiormente responsabile delle emissioni di questa categoria è il diesel, con oltre il 50% del totale delle emissioni, seguito dal gas naturale (20%).

Le prime normative introdotte per contenerne le concentrazioni sono arrivate nei primi anni 90, imponendo limiti alle emissioni degli impianti e favorendo l'uso del gas naturale al posto di gasolio e cherosene come combustibili. Negli anni più recenti i limiti normativi sono stati ulteriormente abbassati. Per quanto riguarda il traffico, visto che è il settore maggiormente responsabile, i provvedimenti più importanti sono sicuramente l'introduzione delle marmitte catalitiche e l'introduzione di limiti alle emissioni dei veicoli, con le famose categorie EuroX. Proprio questo ultimo aspetto ha fatto sì che, sebbene nel corso degli anni il numero dei veicoli circolanti e dei

kilometri percorsi sia aumentato, l'inquinante abbia comunque fatto registrare un trend in forte calo nel corso degli anni [16].

Variabile	Importanza
Temperatura	0.29
Giorno dell'anno	0.19
Velocità del vento	0.15
Radiazione globale	0.10
Data	0.07
Umidità	0.05
Direzione del vento	0.04
Giorni senza pioggia	0.03
Giorno della settimana	0.02
Precipitazioni registrate	0.01

Tabella 3.1: Importanza media delle variabili nei modelli creati per analizzare gli ossidi di azoto

Nei modelli ottenuti le variabili più importanti sono state la temperatura, che è risultata quella col valore decisamente maggiore rispetto alle altre, il giorno dell'anno e la velocità del vento. Sicuramente non siamo sorpresi dalle prime due, visto il classico andamento stagionale dell'inquinante, legato sia alle maggiori emissioni antropiche durante il periodo invernale che alle condizioni atmosferiche più sfavorevoli alla dispersione. Anche il vento, specialmente la sua velocità, è risultato un buon predittore per le concentrazioni di ossidi di azoto, non sorprendentemente vista la forte azione dispersiva legata a questo fattore 3.1.

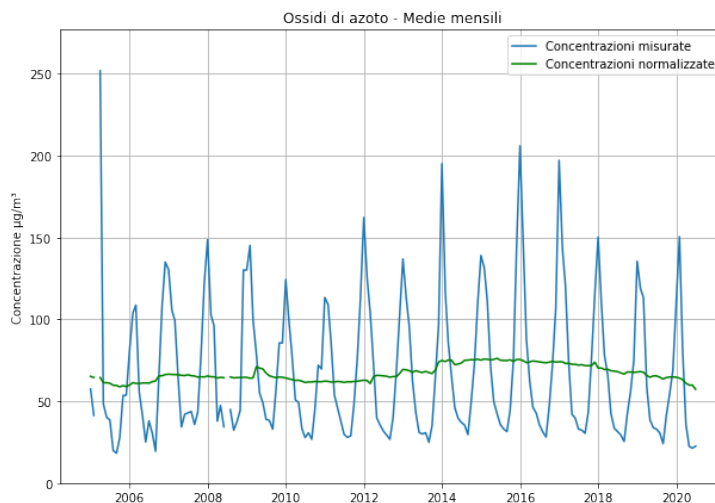


Figura 3.2: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Controllando il grafico in figura 3.2 delle medie mensili si nota come la serie normalizzata riesca a non essere influenzata dall'andamento stagionale, mantenendo sempre un livello "medio" costante, che è proprio il risultato che ci aspettavamo di ottenere dall'applicazione della nostra tecnica. Noi volevamo riuscire ad avere delle serie che fossero indipendenti da tutti questi fattori che possono incidere sulle concentrazioni e visto l'andamento mostrato da quella ottenuta, che rimane approssimativamente intorno ad un valore "medio" indipendentemente dalla giornata, possiamo essere abbastanza fiduciosi di essere riusciti a stimarla con una buona precisione. Sulle medie annuali, riportate in figura 3.3 vengono sicuramente eliminati i picchi, probabilmente causati da condizioni meteorologiche più favorevoli/sfavorevoli all'accumulo avute da alcuni anni in particolare (ad esempio 2006 o 2011), ma l'andamento delle due serie risulta comunque compatibile, ancora una volta a dimostrazione della validità del risultato ottenuto, che sarebbe sicuramente stata messa in discussione se si fossero ottenuti dei risultati completamente diversi.

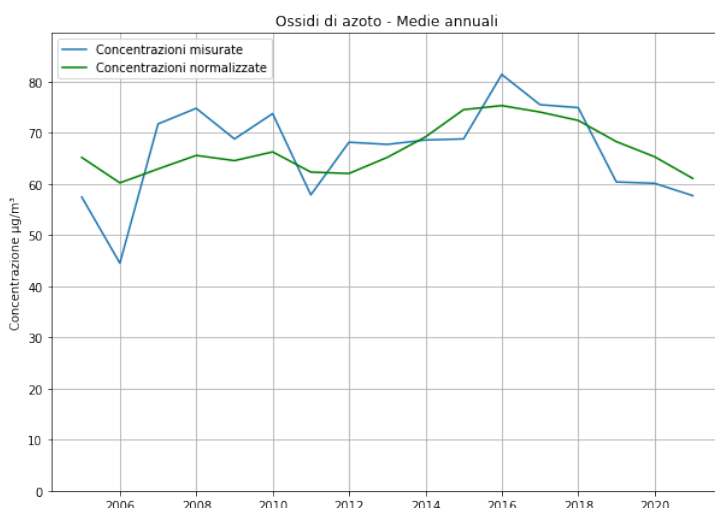


Figura 3.3: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Guardando i grafici 3.2, 3.3 si vede chiaramente che l'andamento delle concentrazioni dal 2006 ad oggi sia abbastanza costante, con un aumento di $12\mu\text{g}/\text{m}^3$ (corrispondente al 18%) registrato sulla serie normalizzata tra gli anni 2012 e 2016, seguito però poi da un costante calo che ormai le ha riportate ai livelli di 8 anni fa. Se avessimo avuto a disposizione le serie complete a partire dai primi anni 90 sicuramente avremmo visto un trend negativo nelle concentrazioni, poiché è stato proprio in quel decennio che le emissioni sono state maggiormente ridotte. Negli ultimi anni non sono state introdotte nuove importanti normative a riguardo, se non aggiornamenti dei limiti imposte da quelle già esistenti (sia per quanto riguarda le emissioni degli impianti che dei veicoli). Questo sicuramente ha aiutato a mantenere le concentrazioni su un livello abbastanza costante, nonostante il numero di veicoli circolanti e le attività produttive siano in continuo aumento.

Successivamente ci siamo occupati del biossido di azoto misurato singolarmente. Per le concentrazioni di questo inquinante il D.Lgs 155/2010 ha stabilito le seguenti soglie: $200\mu\text{g}/\text{m}^3$ sulla media oraria, con un massimo di 18 superamenti annui concessi, e di $40\mu\text{g}/\text{m}^3$ per la media annua. Se il limite orario non viene praticamente mai superato, quello della media annua non sempre viene rispettato. Nel rapporto di ARPA Lombardia [23] per quanto riguarda la provincia milanese, nel 2018, 7 stazioni su 14 hanno fatto registrare una media annuale superiore al limite imposto.

Questo fa sì che il biossido di azoto sia un inquinante di interesse, soprattutto per quando si vuole analizzare l'impatto del traffico sulla qualità dell'aria, essendone così fortemente responsabile.

Variabile	Importanza
Temperatura	0.28
Giorno dell'anno	0.16
Velocità del vento	0.14
Data	0.13
Radiazione globale	0.08
Direzione del vento	0.05
Giorni senza pioggia	0.04
Umidità relativa	0.03
Giorno della settimana	0.03
Precipitazioni registrate	0.02

Tabella 3.2: Importanza media delle variabili nei modelli creati per analizzare il biossido di azoto

Anche per il biossido considerato separatamente le serie ottenute presentano un andamento abbastanza costante e anche in questo caso sembra che dal 2016 le con-

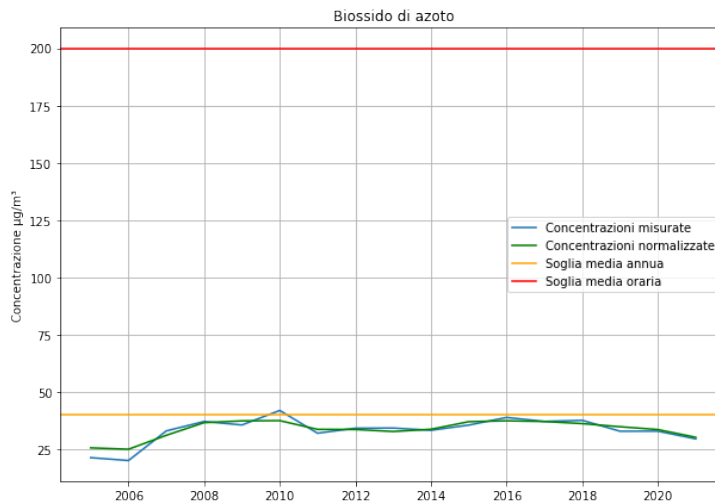


Figura 3.4: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

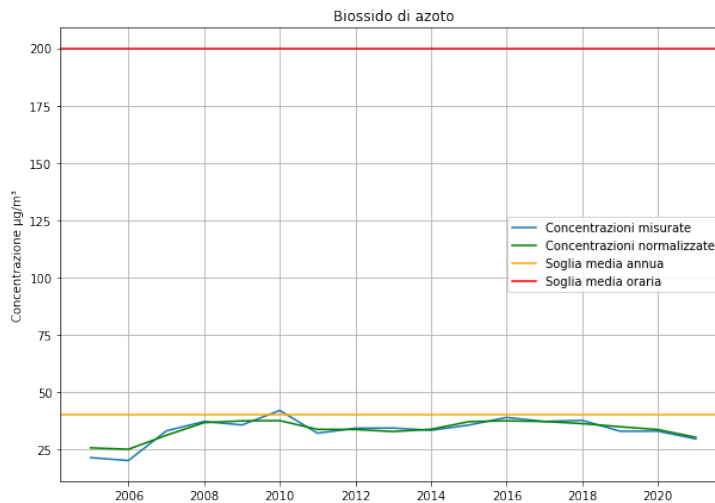


Figura 3.5: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

centrazioni siano leggermente in calo, anche se già in passato (2010-2013) si erano verificati andamenti simili, salvo poi essere tornate ad aumentare negli anni successivi. Sicuramente è importante notare che le medie annuali delle serie normalizzate rimangano sempre sotto alla soglia imposta dalla normativa vigente, a segno che comunque la situazione per quanto riguarda questo inquinante non sia gravemente critica. Questo non significa che non sia importante monitorarla ed occuparsene, visto che comunque in casi di particolari condizioni atmosferiche può capitare che si abbiano delle situazioni leggermente problematiche, ma di certo le concentrazioni di questo inquinante non possono essere considerate un pericolo attualmente.

C'è inoltre da dire che nel corso dei prossimi anni questo inquinante potrebbe ulteriormente calare, poiché l'innovazione tecnologica, soprattutto per quanto riguarda il mondo delle auto e il rinnovo della flotta circolante porteranno ad avere ulteriori abbassamenti delle emissioni. Chiaramente ci si può aspettare di vedere un processo simile anche per quanto riguarda le combustioni industriali e civili, che sono gli altri due settori importanti per quanto riguarda questo inquinante e che potrebbero dare un buon contributo nel contenimento delle concentrazioni.

3.2 PM10 e PM2.5

Le polveri sottili sono sicuramente l'inquinante più discusso degli ultimi anni essendo quello le cui concentrazioni fanno registrare il maggior numero di superamenti della soglia imposta per legge e del quale è noto l'elevato impatto ambientale e sulla salute degli esseri viventi. Questo fa sì che si accendano spesso dibattiti sulla qualità reale dell'aria che respiriamo e su quali misure siano necessarie per riuscire a mantenere le concentrazioni sotto al valore limite oltre il quale il WHO ha riconosciuto la possibilità di avere danni alla salute [25]. Prima di andare ad analizzare i dati è però bene aver chiaro il contesto in cui si svolgeranno queste analisi. Il particolato infatti è un inquinante molto legato alla stagionalità, sia perché le condizioni meteorologiche invernali sono più favorevoli all'accumulo rispetto a quelle estive che per le maggiori emissioni antropiche tipiche della stagione, causate, ad esempio, dall'uso dei riscaldamenti. Va inoltre ricordato che la Lombardia, come tutto il bacino padano in generale, si trova in una posizione geografica sfavorevole, che porta alla maggior formazione di accumuli, soprattutto durante il periodo invernale.

Il particolato classicamente viene diviso in due categorie: PM10 e PM2.5, a seconda del diametro aerodinamico delle particelle esaminate. Quindi particelle con questo diametro, che non si riferisce alla dimensione della particella ma alla sue caratteristiche aerodinamiche ¹, al di sotto dei $10\mu\text{m}$ sono classificate come PM10 e il discorso analogo vale anche per il PM2.5. Essendo appunto particelle molto piccole risultano quindi pericolose per la salute poiché riescono a penetrare più a fondo nell'apparato respiratorio, arrivando quindi ad arrecare danni maggiori. La pericolosità del particolato non deriva solamente dalla sua composizione, che è molto varia per le diverse origini che può avere, ma anche dal fatto che fa da veicolante per altri inquinanti più pericolosi, che si legano in atmosfera e vengono poi trasportate all'interno del corpo dalla particella. Questo inquinante ha appunto origini molto varie, sia di tipo primario da attività come le industrie, i riscaldamenti, il traffico e le combustioni in generale, ma anche di tipo secondario, poiché in atmosfera può formarsi a seguito di trasformazioni chimico-fisiche di altre sostanze.

Secondo INEMAR 2017 [15] la categoria maggiormente responsabile per le emissioni di questi inquinanti sono le combustioni non industriali, quindi, ad esempio, i riscaldamenti domestici, specialmente quelli a biomasse. Il contributo di questa categoria si aggira intorno al 45% per quanto riguarda il PM10 ed al 50% per il PM2.5. Al secondo posto troviamo il traffico, con un contributo che arriva al 25% per quanto riguarda la frazione più grossa del particolato e del 20% per la più piccola. Per quanto riguarda i combustibili, il legno è di gran lunga il maggior responsabile delle emissioni di particolato, con oltre il 50% del totale. Il diesel si ferma invece a solo circa il 10%.

Il PM10 è regolato dall'inizio degli anni 90, con i decreti già citati per gli ossidi di azoto che hanno introdotto limiti e nuove regolamentazioni per le emissioni degli impianti. Per quanto riguarda il settore del traffico l'introduzione più importante è stata quella del filtro antiparticolato nelle marmitte, che permette di ridurre drasticamente le emissioni, insieme chiaramente alle progressive limitazioni alle emissioni imposte tramite le categorie EuroX. Per quanto riguarda il traffico i motori a diesel sono sempre stati riconosciuti come produttori di particolato ed infatti sin dalle prime norme europee ne venivano limitate le emissioni, cosa che non avveniva invece per i motori a benzina, per i quali solo recentemente (a partire dalla categoria Euro 5a, approvata nel settembre 2009) sono state introdotte limitazioni per i motori a iniezione diretta, visto che negli ultimi anni c'è un sempre maggior interesse verso la loro produzione di particolato [30].

Per quanto riguarda il PM10 il D.Lgs 155/2010 impone i limiti di $50\mu\text{g}/\text{m}^3$ come media giornaliera, che non deve essere superata in più di 35 occasioni all'anno, e $40\mu\text{g}/\text{m}^3$ per la media annuale.

Nei modelli ottenuti per trattare questo inquinante la variabile più importante è risultata, non sorprendentemente vista la sua azione dispersiva, la velocità del vento con un valore di 0.20. Al secondo posto troviamo la temperatura (0.18) ed al terzo il giorno dell'anno (0.15), che sono collegate al classico andamento stagionale che fanno

¹Il diametro aerodinamico di una particella è, per definizione, il diametro di una particella perfettamente sferica, di densità unitaria ($1\text{g}/\text{cm}^3$), che ha le stesse caratteristiche inerziali, ovvero che si deposita su un supporto alla stessa velocità.

Variabile	Importanza
Velocità del vento	0.20
Temperatura	0.18
Giorno dell'anno	0.15
Giorni senza pioggia	0.09
Umidità relativa	0.08
Data	0.08
Radiazione globale	0.07
Direzione del vento	0.05
Precipitazioni registrate	0.04
Giorno della settimana	0.02

Tabella 3.3: Importanza media delle variabili nei modelli creati per analizzare il PM10

registrare le concentrazioni di questo inquinante. Anche al numero di giorni dalle ultime precipitazioni registrate è stata data un'importanza discreta (0.1) anche se nemmeno in questo dovrebbe sorprenderci visto che è noto come la pioggia abbatta le concentrazioni portando al suolo parte del particolato presente in atmosfera. La sua assenza prolungata, soprattutto in periodi come quello invernale dove la dispersione naturale è resa più difficile dalle condizioni climatiche, causa sempre innalzamenti delle concentrazioni registrate che sul territorio lombardo sono frequente causa dei superamenti ai limiti imposti dalla normativa vigente. Grazie all'importanza ottenuta possiamo quindi capire che i modelli costruiti sono stati capaci di anche rilevare questo tipo di dinamica e perciò riescono anche ad eliminarne gli effetti più efficacemente 3.3.

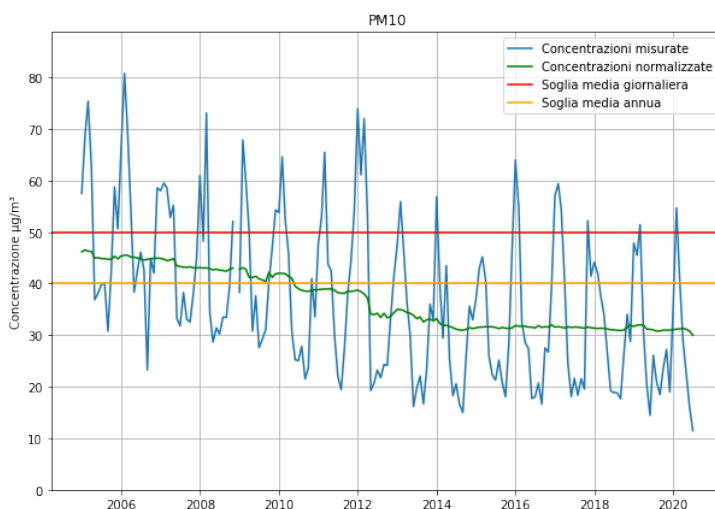


Figura 3.6: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Nei grafici ottenuti è evidente la presenza di un trend in costante calo, diversamente da quanto visto per gli ossidi di azoto, anche se negli ultimi anni sembra aver rallentato molto ed i valori sono rimasti piuttosto costanti. Importante notare come successivamente al 2012 le concentrazioni normalizzate siano rimaste sempre sotto al limite per la media annuale di $40\mu\text{g}/\text{m}^3$, sia per quanto riguarda le medie annuali che per quelle mensili. Ancora una volta vediamo come le serie normalizzate ottenute mantengano un andamento medio rispetto ai dati grezzi, a conferma nuovamente della validità dei risultati ottenuti tramite questa tecnica. Guardando le serie normalizzate nei grafici 3.6, 3.7 si nota come nel corso degli ultimi anni la curva calante che era presente inizialmente sia andata praticamente ad appiattirsi,

Infatti, se a partire dal 2004 la media annuale della serie normalizzata è calata costantemente di anno in anno, dal 2014 in poi i valori si sono stabilizzati sui $30\mu\text{g}/\text{m}^3$. Dalla serie ottenuta si capisce che mediamente la situazione per quanto

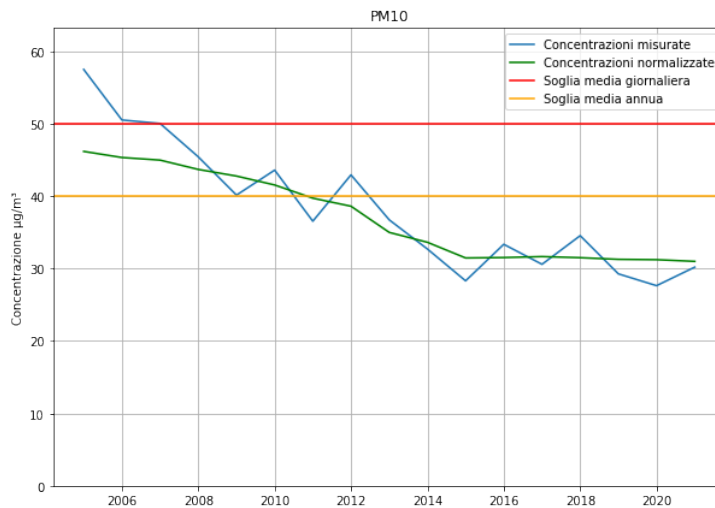


Figura 3.7: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

riguarda questo inquinante non è critica, poiché con condizioni climatiche medie le concentrazioni sarebbero sempre sotto ai limiti legislativi. A causare i superamenti, che avvengono quasi tutti nei mesi invernali, sono invece le condizioni climatiche sfavorevoli di quei mesi che, unitamente all'uso dei riscaldamenti domestici ed in generale ad un maggior consumo energetico, portano il particolato ad accumularsi in atmosfera senza riuscire a disperderlo.

Allo stesso modo abbiamo trattato il PM2.5, per il quale ricordiamo che il limite di legge è fissato a $25\mu\text{g}/\text{m}^3$ sulla media annua, introdotto per la prima volta a partire dal D.Lgs 155/2010.

Variabile	Importanza
Temperatura	0.23
Velocità del vento	0.20
Giorno dell'anno	0.15
Radiazione globale	0.08
Umidità relativa	0.08
Data	0.07
Giorni senza pioggia	0.07
Direzione del vento	0.05
Precipitazioni registrate	0.03
Giorno della settimana	0.01

Tabella 3.4: Importanza media delle variabili nei modelli creati per analizzare il PM2.5

Anche per questo inquinante viene evidenziato un trend negativo leggermente minore di quanto riscontrato col PM10 e che negli ultimi anni presenta la stessa tendenza al rallentamento, rinforzando l'ipotesi di aver raggiunto un limite che per essere passato necessita di forti innovazioni che riescano ad eliminare le emissioni già alla fonte. Notiamo nuovamente come negli anni precedenti al 2014 le medie siano calate di anno in anno, per poi stabilizzarsi su un valore di $22\mu\text{g}/\text{m}^3$ negli anni successivi. Abbiamo notato come dal 2015 le concentrazioni normalizzate siano riuscite anche a scendere sotto al limite di legge imposto e come, anche in questo caso, siano fortemente influenzate dalle condizioni sfavorevoli che poi portano spesso ad avere episodi problematici.

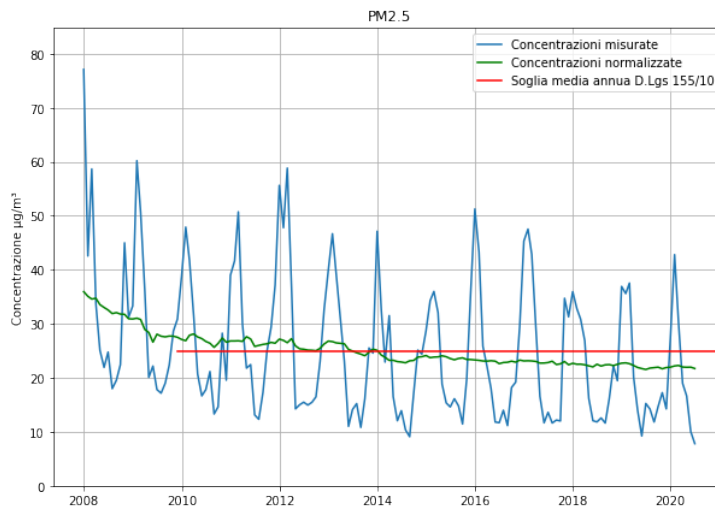


Figura 3.8: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

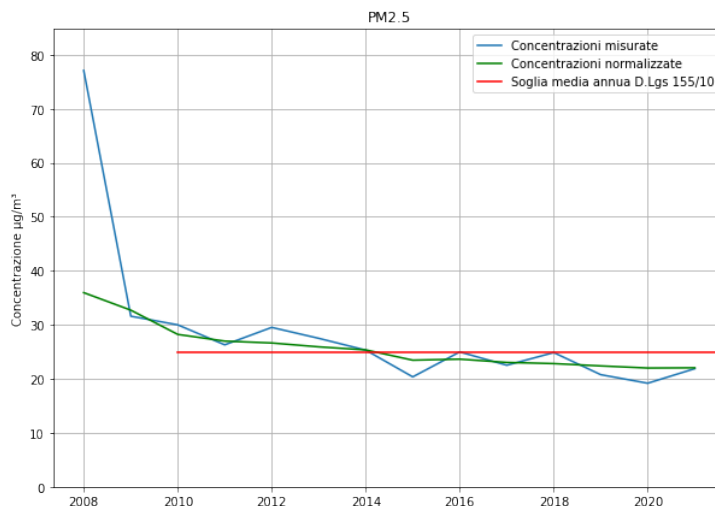


Figura 3.9: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

3.3 CO

Il monossido di carbonio è un inquinante molto pericoloso per la salute, in quanto è nota la sua migliore capacità di legarsi all'emoglobina rispetto all'ossigeno, causando notevoli danni per l'uomo [19]. È un inquinante che viene prodotto da combustioni in difetto di ossigeno ed ha origine prevalentemente primaria, essendo emesso direttamente da tutti i processi di combustione incompleta dei composti carboniosi (gas naturali, propano, carburanti, benzine, carbone, legna, ecc.).

Secondo quanto riportato da INEMAR 2017 i due settori maggiormente responsabili delle emissioni di questo inquinante sono il traffico ed i riscaldamento domestici e le combustioni non industriali, con un contributo pari rispettivamente al 38% e 27% del totale delle emissioni annue. Le emissioni collegate al settore dei trasporti hanno fatto registrare forti cali a partire dagli anni 90, con una riduzione vicina addirittura al 90% favorita da innovazioni quali le marmitte catalitiche e in generale il progresso tecnologico che, per esempio, ha permesso di passare dai 12,66gCO/km emessi dai motori a benzina classificati come Euro 0 (quelli antecedenti all'istituzione delle normative europee) a 1gCO/km che viene imposto come limite per la categoria Euro 6. Per quanto riguarda invece il settore dei riscaldamento domestici nel corso degli ultimi anni le emissioni hanno seguito un trend sempre crescente, causato soprattutto dall'uso della legna come combustibile, che però porta effetti riscontrabili

solo localmente dove è più utilizzata (e solitamente le aree cittadine non sono questo tipo di zone, anche perché il loro uso è fortemente limitato e regolamentato dalla normativa attuale).

Il limite di legge alle concentrazioni di questa sostanza in atmosfera è fissato a $10 \mu\text{g}/\text{m}^3$ a partire dal 2005, anche se ormai sono più di vent'anni che non vengono registrati valori maggiori. Il monossido, quindi, ormai non rappresenta più un problema e le sue concentrazioni in atmosfera sono molto vicine al fondo naturale [23] e spesso si arriva anche ai limiti della rilevabilità da parte dei sensori. Pur non essendo più un problema abbiamo ritenuto che potesse comunque valer la pena di analizzare la situazione anche per quanto riguarda questo inquinante, ripetendo quanto fatto in precedenza per gli altri.

Variabile	Importanza
Temperatura	0.26
Giorno dell'anno	0.18
Data	0.17
Velocità del vento	0.12
Radiazione globale	0.09
Umidità relativa	0.05
Direzione del vento	0.04
Giorni senza pioggia	0.02
Precipitazioni registrate	0.01
Giorno della settimana	0.01

Tabella 3.5: Importanza media delle variabili nei modelli creati per analizzare il monossido di carbonio

Nei modelli la temperatura è risultata essere la variabile con l'importanza maggiore (0.26), in quanto funziona da proxy per la stagionalità e quindi anche per l'uso dei riscaldamenti domestici. Al secondo posto troviamo il giorno dell'anno (0.18) ed al terzo la data, che può essere un indizio di come l'andamento della serie reale possa essere caratterizzato da variazioni non dovute ai fattori meteorologici considerati come potrebbero essere le oscillazioni nel fondo naturale 3.5.

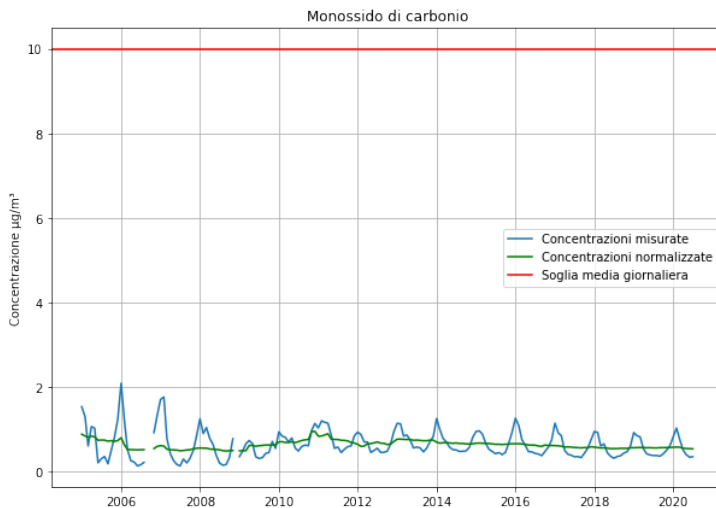


Figura 3.10: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Dal 2006 ad oggi le concentrazioni di monossido di carbonio sono rimaste piuttosto stabili, mostrando un andamento abbastanza costante, con un leggero aumento tra il 2010 e il 2014, seguito poi però da una fase di leggero calo. Questo andamento, che risulta praticamente costante e caratterizzato da variazioni piuttosto casuali, può sicuramente farci pensare di essere vicini alla misura del fondo naturale per questo inquinante, che ormai da anni non rappresenta più un problema. Va infatti notato come nel periodo preso in esame i valori normalizzati siano sempre rimasti sotto al

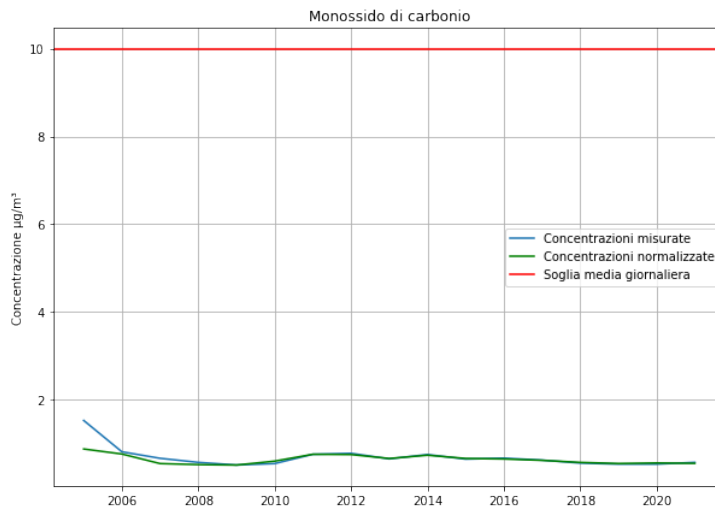


Figura 3.11: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

$1 \mu\text{g}/\text{m}^3$, ben più bassi del limite imposto dalla legge. Inoltre le medie delle serie normalizzate seguono abbastanza bene quelle delle concentrazioni reali, facendoci quindi pensare che le fluttuazioni possano essere più oscillazioni nel valore del fondo naturale che cambiamenti dovuti ad attività umane o altri fattori come quelli considerati dal nostro modello.

3.4 Ozono

L'ozono troposferico è un inquinante di origine secondaria, che si forma in atmosfera quando, favoriti da alte temperature e forte irraggiamento, ossidi di azoto e composti organici volatili subiscono trasformazioni chimico-fisiche che portano alla sua formazione. Per questo motivo l'ozono è considerato smog fotochimico ed il periodo critico per le sue concentrazioni, a differenza degli altri inquinanti, è quindi l'estate [33]. I precursori dell'ozono arrivano generalmente da combustioni civili ed industriali e da processi che usano o producono sostanze chimiche volatili, ma avendo una formazione più complessa è chiaramente più difficile capire quali siano le sue origini e quindi dove intervenire per controllarne efficacemente le concentrazioni.

L'ozono è noto per la sua pericolosità per l'apparato respiratorio, al quale può causare problemi temporanei anche a seguito di un'esposizione a basse concentrazioni; chiaramente se l'esposizione avviene in modo ripetuto o prolungato il rischio di avere dei danni permanenti aumenta. In generale comunque, soprattutto ad alte concentrazioni, l'ozono diventa pericoloso sia per gli esseri umani che per la vegetazione, sulla quale può avere effetti negativi.

La normativa attuale fissa a $120 \mu\text{g}/\text{m}^3$ il valore obiettivo per la media mobile calcolata su 8 ore, con un massimo di 25 superamenti annui calcolati sulla media di 3 anni. Attualmente questo limite viene rispettato a fatica (nel 2018 nella provincia di Milano tutte le stazioni l'hanno abbondantemente superato), ma ciò comunque non costituisce una criticità per la regione, in quanto questi superamenti sono tutti collegati al classico andamento stagionale [23]. Nei grafici 3.13 questi superamenti non sono però così evidenti poiché sono presentati gli andamenti delle medie mensili, mentre la soglia di legge fa riferimento alla media mobile calcolata sulle 8 ore, che, in determinati momenti e con determinate condizioni, può raggiungere facilmente picchi più alti rispetto al resto del mese.

Nei modelli ottenuti per la trattazione dell'ozono le due variabili nettamente più importanti si sono rivelate essere la radiazione solare (0.33) e la temperatura (0.32), dimostrando quindi che sono stati capaci di interpretare perfettamente le dinamiche che portano alla maggior formazione di questo inquinante 3.6.

L'andamento delle concentrazioni a partire dal 2013 è leggermente in aumento con una crescita del 10%. L'ozono, per la sua origine complessa, è molto difficile

Variabile	Importanza
Radiazione globale	0.33
Temperatura	0.32
Giorno dell'anno	0.12
Umidità relativa	0.08
Velocità del vento	0.06
Data	0.02
Direzione del vento	0.01
Giorni senza pioggia	0.01
Precipitazioni registrate	0.00
Giorno della settimana	0.00

Tabella 3.6: Importanza media delle variabili nei modelli creati per analizzare l'ozono

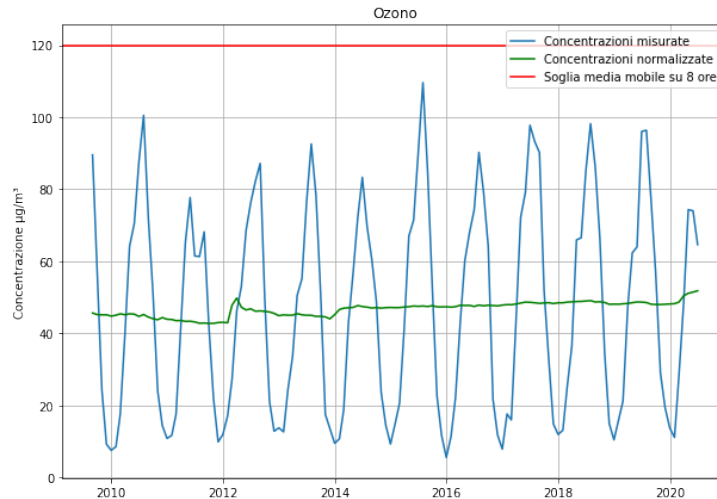


Figura 3.12: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

da tracciare e quindi diventa complicato anche intervenire per contenerne le concentrazioni in modo efficace. Attualmente la situazione in Lombardia non risulta problematica, ma il trend in aumento richiede di porre attenzione alla situazione di questo inquinante.

Un aspetto da notare è che l'ozono, tra tutti gli inquinanti, è quello che la nostra tecnica riesce a trattare con la maggiore precisione ed infatti i modelli ottenuti usati per ricavare le serie normalizzate sono quelli che in fase di costruzione hanno mostrato le performance migliori. Questo probabilmente è collegato alla forte relazione che questo inquinante ha con la temperatura e, soprattutto, la radiazione solare, che i nostri modelli hanno ben individuato (come si è visto con l'importanza delle variabili) e che quindi gli permetteva di essere più precisi nell'attività predittoria.

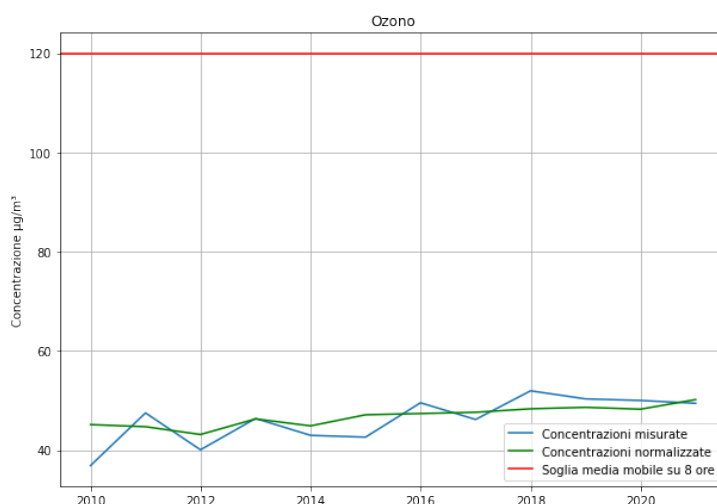


Figura 3.13: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

3.5 Ammoniaca

L'ammoniaca è un inquinante prodotto da processi degradativi di sostanza organica, le cui principali sorgenti sono infatti attività agricole e, in misura minore, il trasporto su strada e la combustione di legna e combustibili fossili. È un gas molto solubile in acqua e questo lo rende pericoloso poiché può essere causa dell'acidificazione dei suoli. Inoltre negli ultimi anni l'interesse per l'ammoniaca è sempre crescente a causa della sua partecipazione nella formazione di particolato secondario quando presente in atmosfera.

Secondo INEMAR2017 in Lombardia il 96% delle emissioni di questa sostanza sono collegabili al settore dell'agricoltura, che infatti è il principale responsabile per quanto riguarda questo inquinante. Il traffico, invece, arriva a dare un contributo pari solamente al 1% del totale annuo.

La normativa attuale non impone limiti per le concentrazioni registrate, anche se una serie di direttive europee per lo sviluppo rurale ha mirato, nel corso degli anni, a favorire la diffusione di buone pratiche per contenerne le emissioni, come ad esempio il divieto dello spargimento di liquami (DGR X/7095 del 18 Settembre 2017).

Variabile	Importanza
Data	0.35
Giorno dell'anno	0.14
Temperatura	0.11
Umidità relativa	0.08
Radiazione globale	0.07
Direzione del vento	0.06
Velocità del vento	0.05
Giorni senza pioggia	0.04
Giorno della settimana	0.02
Precipitazioni registrate	0.02

Tabella 3.7: Importanza media delle variabili nei modelli creati per analizzare l'ammoniaca

Nei modelli ottenuti la variabile più importante è risultata essere la data (0.35), seguita dal giorno dell'anno (0.14) e dalla temperatura (0.11). Anche in questo caso, in cui alla data viene data un'importanza molto alta, potremmo trovarci nella stessa situazione che già parzialmente si era vista per il monossido di carbonio. L'andamento di questo inquinante, infatti, probabilmente è regolato da fenomeni diversi di quelli considerati dalle nostre variabili e quindi la data stessa risulta essere l'informazione migliore per fare previsioni sulle concentrazioni di ammoniaca in tale data 3.7.

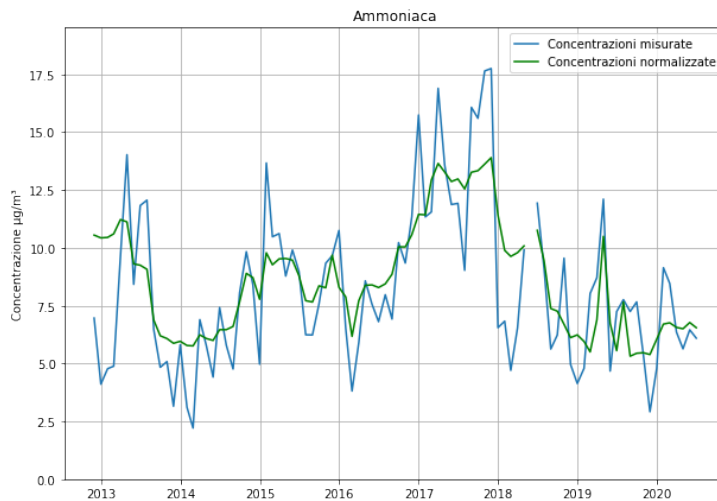


Figura 3.14: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

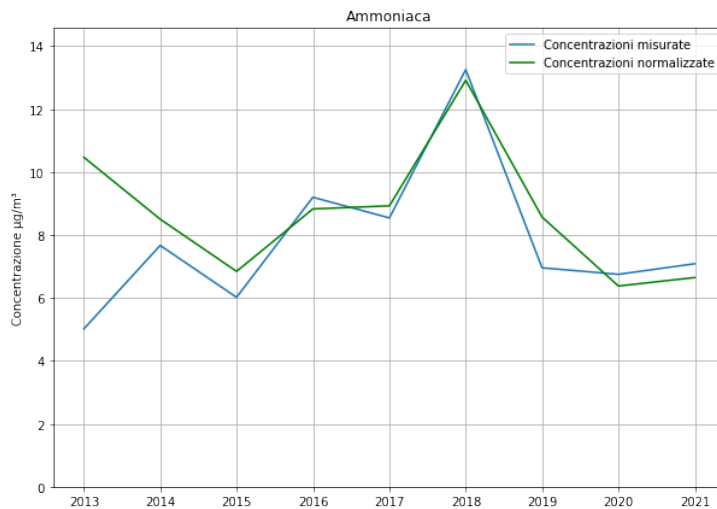


Figura 3.15: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Come si vede nei grafici delle figure 3.14, 3.15 l'andamento delle serie normalizzate ottenute rispecchia abbastanza fedelmente quello delle concentrazioni realmente misurate. Questa è quindi un'ulteriore dimostrazione di come l'influenza della meteorologia su questo inquinante sia praticamente nulla, visto che dopo averne teoricamente eliminato gli effetti si ottiene ancora una serie molto simile.

3.6 Benzene

Il benzene viene sintetizzato dal petrolio e viene usato per produrre materie plastiche e soprattutto come sostanza antidetonante nella benzina. Il benzene, infatti, ad inizio anni 90 ha sostituito il piombo come additivo presente in essa, permettendo il passaggio da quella rossa a quella verde. Quello presente in atmosfera deriva da processi di combustione incompleta di combustibili fossili, quindi da attività come traffico (soprattutto dai veicoli a benzina) e processi di combustione industriale.

La sua pericolosità per la salute umana varia molto a seconda della concentrazione e della durata dell'esposizione, ma anche alle basse quantità a cui viene misurato in atmosfera può comunque essere pericoloso: lo IARC (agenzia internazionale per la ricerca sul cancro) l'ha infatti inserito tra le sostanze per le quali esiste una sufficiente evidenza di cancerogenicità per l'uomo [24].

Il D.Lgs 155/2010 stabilisce per questo inquinante un valore limite di $5\mu g/m^3$ per la media oraria, che viene ormai ampiamente rispettato in tutta la regione. Questo inquinante, che in passato era più problematico, è stato drasticamente ridotto grazie a diverse misure come la riduzione del suo tenore nelle benzine e l'adozione del ciclo chiuso e dei catalizzatori nelle marmitte.

Variabile	Importanza
Temperatura	0.26
Giorno dell'anno	0.16
Data	0.16
Velocità del vento	0.15
Radiazione globale	0.07
Umidità relativa	0.06
Direzione del vento	0.03
Giorni senza pioggia	0.02
Giorno della settimana	0.01
Precipitazioni registrate	0.01

Tabella 3.8: Importanza media delle variabili nei modelli creati per analizzare il benzene

Nei modelli ottenuti l'importanza delle variabili si è rivelata essere abbastanza in linea con quanto visto per gli altri inquinanti, con temperatura, giorno dell'anno e data che risultano essere le più importanti 3.8.

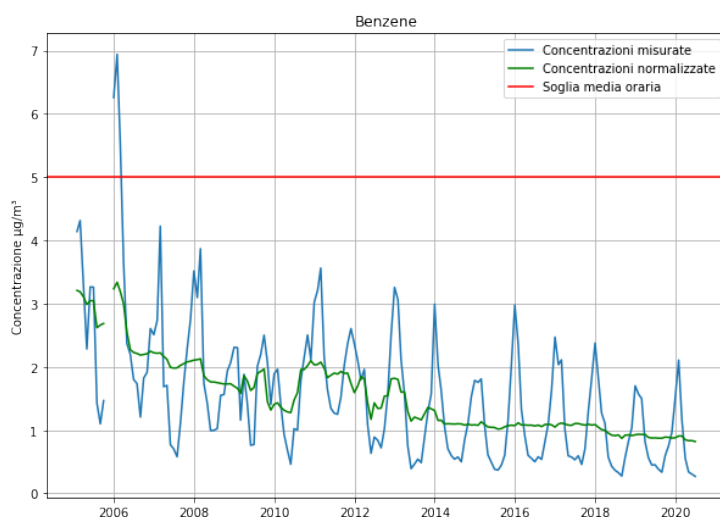


Figura 3.16: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

Nei grafici in figure 3.16, 3.17 sono riportati gli andamenti delle serie reali e normalizzate di questo inquinante. Si nota chiaramente un costante trend in calo, dovuto alle misure sopracitate, che hanno permesso, nonostante l'aumento dei volumi di traffico, di abbattere le concentrazioni ad un livello che ormai non risulta più pericoloso per la salute umana.

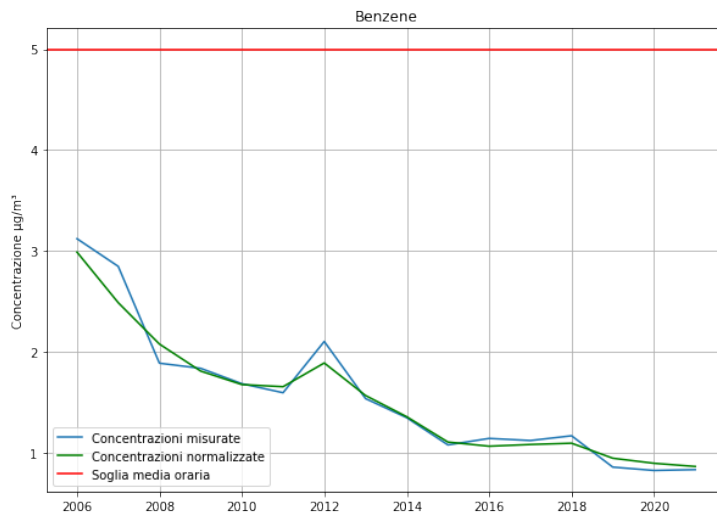


Figura 3.17: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

3.7 SO_2

Sebbene una volta fosse sicuramente uno degli inquinanti a destare più preoccupazioni, così come è stato uno dei primi ad essere monitorato e limitato per legge, è già ormai diversi anni che le concentrazioni di biossido di zolfo sono ampiamente sotto al limite legislativo per questo inquinante ($350 \mu g/m^3$ media oraria), visto che a partire dagli anni 90 sono state introdotte una serie di normative atte a limitarne le emissioni. Le più importanti sono sicuramente quelle riguardanti gli usi di gasolio e nafta come forme di riscaldamento, che ormai sono stati quasi completamente sostituiti dal metano, e quelle che nel corso degli anni hanno progressivamente abbassato il limite di zolfo che può essere contenuto nei carburanti (da 0.8% nel 1980 a 0.2% nel 1995 e infine a 0.1% nel 2008).

Sebbene non vi siano più preoccupazioni per quanto riguarda le concentrazioni di questa sostanza in atmosfera abbiamo ritenuto che fosse utile verificare anche con essa i risultati ottenuti tramite la nostra tecnica.

Variabile	Importanza
Data	0.41
Giorno dell'anno	0.14
Temperatura	0.11
Velocità del vento	0.06
Radiazione globale	0.06
Umidità relativa	0.05
Direzione del vento	0.05
Giorni senza pioggia	0.05
Precipitazioni	0.02
Giorno della settimana	0.01

Tabella 3.9: Importanza media delle variabili nei modelli creati per analizzare il biossido di zolfo

La data è risultata di gran lunga la variabile più importante (0.41), con giorno dell'anno (0.14) e temperatura (0.11) che sono le sole due variabili a cui veniva attribuita un minimo di importanza. Trovare la data come variabile più importante per un inquinante che ormai è arrivato al fondo naturale ci conferma esattamente come le fluttuazioni presenti nelle sue serie siano dovute a fattori esterni a quelli considerati nei nostri modelli 3.9.

Come si può vedere nei grafici 3.18, 3.19 le concentrazioni di SO_2 registrate sono ormai a livelli bassissimi, tanto da non destare più preoccupazioni. Come già ampiamente riconosciuto da diversi studi in materia [16, 31] questo inquinante

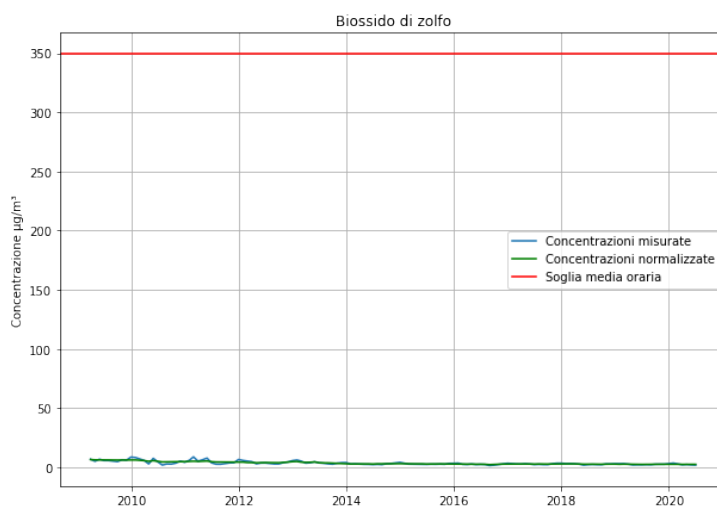


Figura 3.18: Confronto tra le medie mensili delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

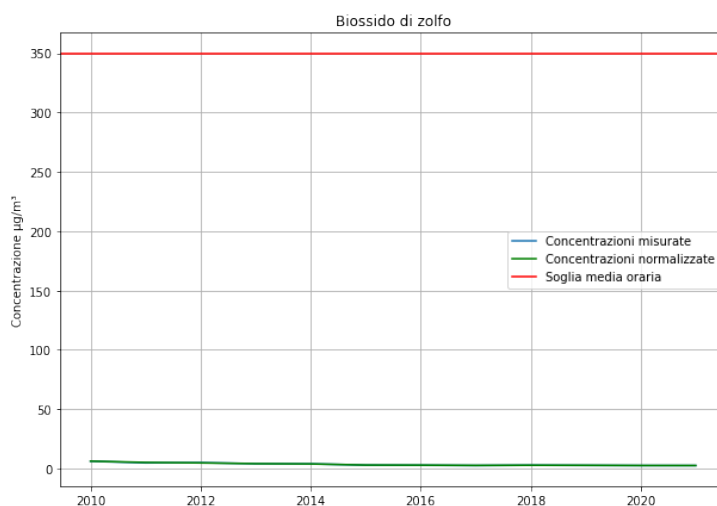


Figura 3.19: Confronto tra le medie annuali delle serie storiche rilevate dai sensori e quelle delle serie normalizzate

potrebbe essere arrivato al fondo naturale e non c'è motivo di preoccupazione per quanto riguarda la sua situazione attuale.

Capitolo 4

Effetti del traffico sui principali inquinanti atmosferici

Avere a disposizione una tecnica che ci permette di eliminare l'influenza delle condizioni meteorologiche e della variabilità stagionale dalle concentrazioni degli inquinanti atmosferici ci permette di poter fare delle analisi più precise sull'efficacia di misure prese per contrastare l'inquinamento. L'andamento delle serie normalizzate, infatti, deve per forza essere causato da fattori non considerati nella loro generazione. Sicuramente i due più importanti sono: l'eventuale trend rilevato per le concentrazioni, che noi ovviamente non andiamo ad eliminare poiché è quello che ci interessa per monitorarne l'andamento, ed altri fattori che possono avere un'influenza e che non sono stati considerati, quindi ad esempio particolari situazioni in cui le emissioni antropiche sono state ridotte (COVID) o nuove normative messe in campo.

Quando si parla di inquinamento atmosferico uno degli aspetti su cui sicuramente il dibattito è più acceso e che attira più interesse è l'influenza del traffico veicolare sulla qualità dell'aria. Spesso infatti si sente additare il traffico come il principale responsabile delle concentrazioni degli inquinanti più preoccupanti (e, salvo gli ossidi di azoto, questa cosa non è vera) e sempre maggiori nel corso degli anni sono stati i provvedimenti presi nelle grandi città atti a ridurre i volumi, in nome di una migliore qualità dell'aria. Questi provvedimenti sono stati sempre molto contestati e dibattuti, poiché molti ritengono che il traffico non sia la causa principale dell'inquinamento e che la soluzione non sia bloccarlo ma studiare alternative per renderlo più sostenibile ed efficiente, poiché gli spostamenti in automobile sono comunque una necessità per molte persone.

Il nostro obiettivo è stato quello di provare ad usare la nostra tecnica per verificare quali siano stati i risultati ottenuti in seguito all'applicazione di alcuni provvedimenti sul traffico presi nella città di Milano nel corso degli ultimi anni, in modo da mostrare come l'applicazione di tecniche per normalizzare i dati dell'inquinamento rispetto alla meteorologia permetta uno studio migliore dell'efficacia delle misure prese nell'abbattimento delle concentrazioni. Purtroppo la mancanza di dati relativi a velocità e direzione del vento, due misure (specialmente la prima) a cui di solito viene assegnata una discreta importanza dai nostri modelli, ci costringe a limitare le nostre analisi al periodo successivo al 2012, ovvero solo successivamente all'introduzione di Area C. Sarebbe stato interessante, ma purtroppo non è stato possibile, avere a disposizione anche dati degli anni precedenti, in modo da poter fare un confronto più esteso che coinvolgesse anche il periodo precedente all'entrata in vigore del provvedimento. Per verificare eventuali effetti derivanti dall'introduzione di questi provvedimenti siamo andati a confrontare l'andamento della serie normalizzata ottenuta per la stazione di Milano Via Senato, coinvolta appunto dal provvedimento Area C, con quelli delle serie di altre due stazioni: quella di Pioltello Limito, situata nell'hinterland milanese, e quella di Bormio, che si trova in un ambiente molto diverso da quello cittadino. La prima è stata scelta poiché, essendo comunque vicina alla città, la qualità dell'aria e le condizioni che la determinano dovrebbero essere abbastanza simili. Non essendo stata colpita da nessun provvedimento, inoltre,

ci permette di avere una base di confronto per verificare se appunto l'applicazione di misure come Area C possa effettivamente aver portato a qualche risultato. La stazione di Bormio invece presenta caratteristiche completamente differenti, ma è comunque utile provare a verificare eventuali analogie o discrepanze tra le serie, anche per vedere quali siano le differenze tra due località così diverse.

4.1 NO_x e NO_2

4.1.1 NO_x

Quando si parla di traffico e si vuole analizzare il suo impatto sulla qualità dell'aria gli inquinanti di maggiore interesse sono sicuramente gli ossidi di azoto, di cui questa categoria risulta responsabile per il 50% delle emissioni secondo quanto riportato da INEMAR 2017 [15].

Quando si parla degli ossidi di azoto relativamente al traffico l'attenzione viene posta maggiormente sul diesel, che è il maggior responsabile delle emissioni di questo inquinante. Nel corso degli ultimi anni il numero dei veicoli con motorizzazione diesel circolanti in Italia è andato sempre ad aumentare, tanto che IIR2020 [16] stima un consumo quasi triplo di carburante per quanto riguarda veicoli diesel rispetto a quelli a benzina. Questo aumento di veicoli circolanti avrebbe dovuto aver l'effetto di innalzare le concentrazioni di ossidi di azoto misurate, ma è stato contrastato dall'innovazione tecnologica che è riuscita sempre di più a ridurre le emissioni prodotte (basti pensare alle categorie EuroX), introducendo importanti invenzioni come ad esempio la marmitta catalitica.

In precedenza, analizzando l'andamento della media tra le serie dei capoluoghi di provincia lombardi, avevamo ottenuto un andamento abbastanza costante nel corso degli ultimi 15 anni, con un leggero trend calante negli anni più recenti. Sembra quindi che anche per la Lombardia l'aumento dei veicoli circolanti motorizzati a diesel sia stato contrastato dal progresso tecnologico, che ha permesso di mantenere le concentrazioni sotto controllo.

Abbiamo quindi creato tre modelli per questo inquinante, uno per ogni località scelta per il nostro esperimento, e provato a verificare come negli anni si siano evolute le diverse situazioni.

Innanzitutto abbiamo notato come per tutte e tre le località la velocità del vento sia risultata una variabile a cui è sempre stata assegnata una buona importanza (0.21 a Milano, 0.28 a Limite e 0.34 a Bormio). Per quest'ultima stazione, addirittura, l'importanza risulta di gran lunga più alta rispetto a tutte le altre variabili. Pensando alle sue caratteristiche geografiche la cosa però non dovrebbe sorprenderci più di tanto, poiché è evidente come in una valle montana l'azione del vento possa avere una funzione dispersiva ancora più importante che in altre località. L'altra variabile che in tutti e tre i casi è stata riconosciuta come buon predittore risulta essere la temperatura, che può essere utile per verificare sia l'andamento stagionale che possibili eventi meteorologici particolari che possono portare ad un aumento delle emissioni antropiche (ad esempio periodi particolarmente freddi in cui vengono maggiormente usati riscaldamenti e automezzi rispetto a quando si registrano temperature più miti).

Dal grafico 4.1 si nota innanzitutto come la serie di Bormio abbia davvero dei valori ridotti che ci mostrano come in località di questo tipo l'inquinamento da ossidi di azoto non sia sicuramente un problema. Un altro aspetto che risalta subito all'occhio è l'andamento pressoché identico delle serie relative a Milano e Limite, che mostrano entrambe come a partire dal 2016/2017 si sia registrato un trend decrescente praticamente compatibile tra le due stazioni. Analizzando solo la serie di Milano avremmo potuto pensare che questo calo potesse essere collegato alla riduzione del traffico causata dai provvedimenti presi, ma osservare lo stesso andamento anche su una stazione diversa e non colpita da tali limitazioni ci indica che la causa deve per forza essere un'altra. Una possibile idea è che l'origine di questo trend sia proprio nell'innovazione tecnologica, i cui risultati stanno finalmente mettendosi in mostra nonostante le sempre maggiori attività umane, portando a registrare un calo generale e non specifico di alcune località. A Bormio, dove sicuramente la densità di attività antropiche inquinanti è nettamente ridotta rispetto alle altre due stazioni,

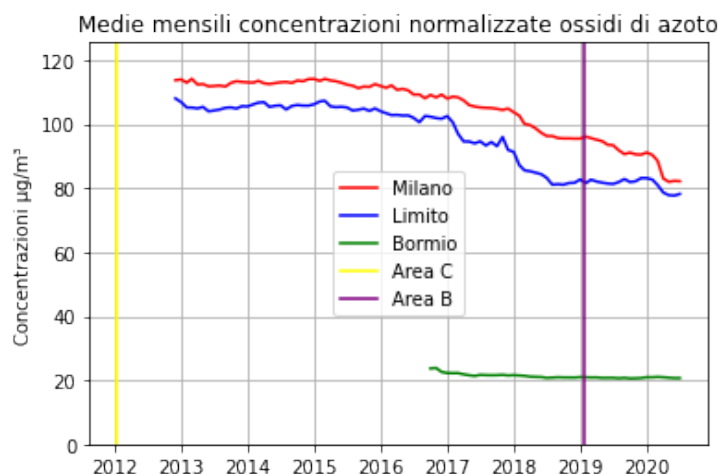


Figura 4.1: Confronto degli andamenti delle serie normalizzate calcolate per le stazioni di Milano via Senato, Limite di Pioltello e Bormio

questo calo risulta chiaramente molto minore, sia appunto perché gli effetti dell'innovazione possono essere meno evidenti a causa della minor applicazione che perché le concentrazioni sono già molto ridotte e quindi è ancora più difficile far registrare dei cali consistenti.

4.1.2 NO_2

Per completezza è stato trattato anche il biossido di azoto, inquinante sempre molto collegato al traffico e attorno al quale c'è molto interesse per i possibili effetti sulla salute umana.

I modelli creati hanno mantenuto le buone performance viste con quelli creati per trattare gli ossidi ed anche l'importanza delle variabili è risultata praticamente invariata, non sorprendentemente vista la relazione tra le due misure.

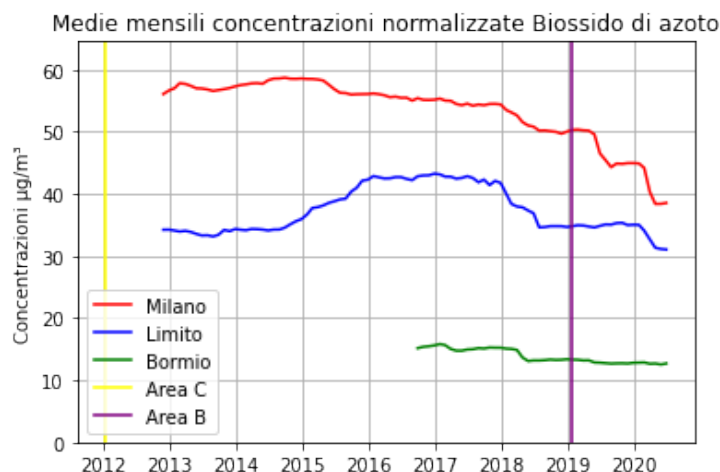


Figura 4.2: Confronto degli andamenti delle serie normalizzate calcolate per le stazioni di Milano via Senato, Limite di Pioltello e Bormio

In questo caso notiamo una sostanziale differenza tra l'andamento delle serie di Milano e Limite. La prima, in modo abbastanza compatibile con quanto già visto per gli ossidi di azoto, mostra un trend decrescente a partire dagli ultimi cinque anni 4.2. Per la seconda, invece, tra gli anni 2015 e 2018 si è assistito ad un aumento delle concentrazioni, che prima si mantenevano su un livello abbastanza stabile, che sono tornate a riscendere nel corso degli ultimi due anni. È proprio negli stessi anni in cui la serie normalizzata fa registrare questo aumento che proprio in quella zona è stata inaugurata l'autostrada BreBeMi, che potrebbe aver portato ad un aumento del

flusso di traffico e conseguentemente delle concentrazioni registrate. Andrebbe però verificato per quale motivo a seguito del 2018 sembra che le concentrazioni tornino sul livello più o meno costante visto tra il 2012 ed il 2014. Per quanto riguarda la situazione di Milano sembra che effettivamente la qualità dell'aria sia migliorata nel corso degli anni, potenzialmente anche grazie alla riduzione del traffico causata da Area C prima ed Area B poi, visto che proprio qualche mese dopo l'introduzione di questo provvedimento il nostro modello ha effettivamente rilevato un calo intorno al 10% nelle concentrazioni di biossido. Se questa è sicuramente una potenziale causa è anche vero che probabilmente a contribuire ai ribassamenti c'è un trend di fondo, dettato da fattori come l'innovazione tecnologica, che infatti sembra parzialmente verificato anche per Pioltello (almeno a seguito del 2018) che per Bormio.

Per quanto riguarda gli ossidi di azoto i nostri modelli sembra che effettivamente mostrino un calo intorno al 10% a seguito dell'introduzione di Area B. Sarebbe interessante poter comparare questo calo col numero di veicoli che a causa di questa misura non hanno più circolato, per capire se possa esistere una corrispondenza tra i due dati. Dall'altra parte, però, questo calo risulta comunque minimo e sicuramente minore di quello dovuto al trend mostrato dalle concentrazioni, che dal 2012 ad oggi hanno mostrato un calo intorno al 33% sia per gli ossidi che per il biossido considerato singolarmente, e che è sicuramente collegato col progresso tecnologico che ha portato alla riduzione delle emissioni.

4.2 CO

Per quanto riguarda il monossido di carbonio INEMAR 2017 individua trasporto su strada e combustioni non industriali (quindi i riscaldamenti, in particolare quelli alimentati a biomasse) come le due principali fonti emissive, con un percentuale di circa il 60% del totale.

La situazione per quanto riguarda questo inquinante non risulta essere critica, visto che le concentrazioni registrate sono ormai prossime a valori riconducibili al fondo naturale. Nonostante ciò abbiamo ritenuto fosse comunque utile fare un'analisi di questo tipo, sia per verificare eventuali effetti dei provvedimenti di limitazione al traffico presi, che per provare a cercare eventuali differenze tra le serie delle tre località. In questo caso risulta particolarmente interessante indagare anche su quella di Bormio, poiché è una zona in cui la legna viene ancora molto usata come forma di riscaldamento e quindi si possono indagare su quali siano gli effetti di questi "comportamenti".

La variabile con importanza maggiore nei tre modelli è stata la temperatura (0.20 a Milano, 0.34 a Limite e 0.19 a Bormio), che gli permette proprio di cogliere l'andamento tipicamente stagionale di questo inquinante. Inoltre, essendo derivante in buona parte dai riscaldamenti domestici, è abbastanza evidente come tale misura possa essere utilizzata per stimarne un loro utilizzo e quindi riuscire a fare previsioni più precise e che sappiano fare da proxy anche questo aspetto. La velocità del vento, chiaramente, ha ottenuto ancora una buona importanza, che ci conferma nuovamente la forte azione dispersiva di questo elemento. Per quanto riguarda il modello di Bormio abbiamo visto che la radiazione globale sia risultata essere la variabile con importanza maggiore (0.19), al pari della temperatura. Anche per essa e per il collegamento con la presenza di sole (e quindi il possibile uso di riscaldamenti) e la stagionalità è naturale che valga lo stesso discorso fatto per la temperatura.

Se le stazioni di Limite e Bormio presentavano valori già molto bassi e con un andamento costante, per la stazione di Milano, nel grafico in figura 4.3 possiamo vedere che nel corso degli ultimi anni si sia registrato un calo abbastanza importante delle concentrazioni. Attualmente i valori risultano ancora più alti rispetto a quelli delle altre due località, potenzialmente anche a causa della maggior densità abitativa (e quindi un maggiore uso di veicoli e riscaldamenti) della città rispetto alle altre due zone. L'origine di questo calo potrebbe sicuramente essere in parte collegata ai provvedimenti di limitazione del traffico, come potrebbe venire il sospetto guardando il grafico, poiché viene registrato solo per la stazione di Milano e non sulle altre due (a differenza di quanto visto in precedenza dove il calo risultava essere compatibile su tutte le località). Altre possibili ragioni potrebbero essere i provvedimenti di limitazione all'uso di riscaldamenti a biomasse nelle città presi nel corso degli ultimi

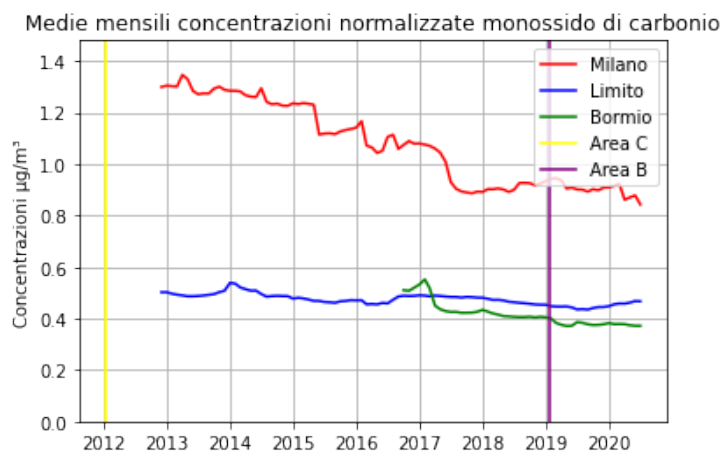


Figura 4.3: Confronto degli andamenti delle serie normalizzate calcolate per le stazioni di Milano via Senato, Limite di Pioltello e Bormio

anni, anche se il loro effetto probabilmente si sarebbe dovuto vedere, anche solo parzialmente, anche sulle serie delle altre località, in particolare quella di Limite che è stata soggetta alle stesse limitazioni.

4.3 PM10

Le polveri sottili sono sempre un inquinante molto discusso e che spesso viene (anche erroneamente) collegato al traffico. Secondo INEMAR 2017 [15] il traffico risulta essere responsabile di meno del 25% delle emissioni di PM10, mentre il settore principale risultano ancora essere le combustioni non industriali, specialmente per quanto riguarda i riscaldamenti a biomasse. Avendo una composizione molto varia è comunque difficile stabilire in modo preciso l'origine del particolato atmosferico e quindi quali siano le fonti di maggiori emissioni.

Negli ultimi anni c'è stata molta attenzione riguardo a questo inquinante, specialmente per gli effetti che ha sulla salute umana. Questo ha quindi portato anche molto interesse nella ricerca di soluzioni efficaci per il loro contenimento, che per quanto riguarda il traffico sono state individuate nell'uso di filtri antiparticolato che riducano le emissioni al tubo di scappamento.

Anche in queste trattazioni si è riscontrato come la nostra tecnica abbia sempre prestazioni leggermente peggiori quando si vanno a trattare le polveri sottili, probabilmente a causa del fatto che le variabili da noi considerate non siano sufficientemente in grado di spiegare i fenomeni su più larga scala che influenzano le loro concentrazioni.

Dal grafico 4.4 notiamo come l'andamento delle tre serie, sebbene quella di Bormio stia su valori molto minori, sia assolutamente compatibile con grande similitudine tra quella di Milano e quella di Limite e vista la natura del particolato, che viene trasportato anche per grandi distanze, la cosa comunque non ci stupisce. In questo caso si vede come non si possa notare alcuna differenza tra gli andamenti delle serie che ci possa suggerire una reale efficacia dei provvedimenti presi in termini di limitazione del traffico. Anzi, guardando il grafico ottenuto, la serie di Limite (non coinvolto da nessuna misura) risulta calata leggermente di più rispetto a quella di Milano. Il trend ricavato è calante ma i valori sono sicuramente ancora molto vicini al limite imposto per legge e questa potrebbe essere un'indicazione della necessità di misure di diverso tipo per contrastare questo inquinante, che dovranno necessariamente coinvolgere i settori maggiormente responsabili delle emissioni e non solamente il traffico.

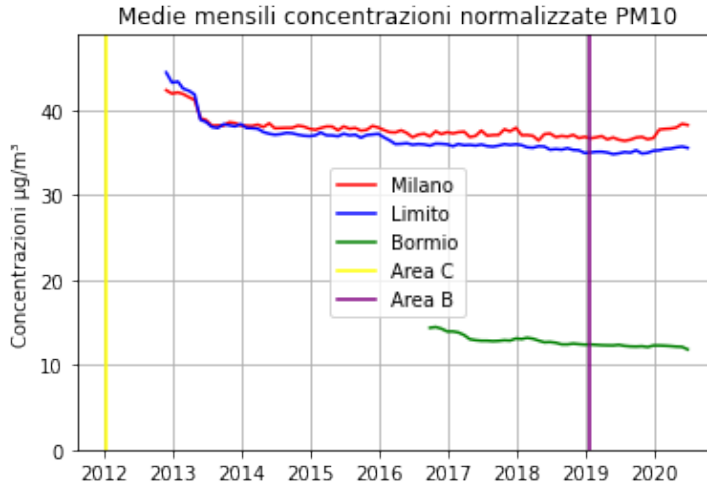


Figura 4.4: Confronto degli andamenti delle serie normalizzate calcolate per le stazioni di Milano via Senato, Limite di Pioltello e Bormio

4.4 Utilizzo di dati relativi al traffico

L'algoritmo Random Forest ha il grande vantaggio di poter sempre provare ad introdurre nuove variabili predittive per i nostri modelli senza doverci preoccupare di problemi come la correlazione o la collinearità, come invece succede con metodi come la regressione lineare. Infatti, in fase di costruzione, sarà proprio il modello stesso a scegliere quali variabili usare maggiormente per fare previsioni, scegliendo di volta in volta quelle che portano ad avere risultati più accurati.

Per indagare ulteriormente sugli effetti del traffico sulle concentrazioni degli inquinanti abbiamo quindi provato ad introdurre una variabile che tracci appunto il numero di veicoli circolanti, in modo da costruire dei modelli che siano in grado di eliminarne l'influenza dalle concentrazioni, portando tutte le giornate della serie storica ad una condizione di traffico medio in maniera analoga a quanto fatto con tutte le altre misure predittive scelte. Per fare questo sono venuti in nostro soccorso i dataset [36, 37, 38, 39, 40, 41, 42, 43] pubblicati dal comune di Milano con le registrazioni degli ingressi in AreaC. Tramite un apposito script abbiamo quindi recuperato i dati anche di tale dataset, organizzati in appositi file e usati poi nella preparazione dei dati su cui basare la costruzione del nostro modello. Così facendo abbiamo ottenuto un modello in grado di eliminare l'effetto del traffico dalle concentrazioni, restituendoci una serie che rappresenti quale sarebbe stata la concentrazione in una giornata con condizioni meteorologiche, stagionali e del traffico medie. Questa serie potrà poi essere confrontata con quelle generate dai modelli creati senza la variabile del traffico per cercare possibili discrepanze, che potrebbero appunto indicarci come l'effetto del traffico abbia pesato sulle concentrazioni registrate in tali periodi.

Nel grafico 4.5 si nota come nel corso degli anni l'andamento del volume di traffico sia sempre stato piuttosto costante, mostrando un andamento ciclico caratterizzato da un calo, prevedibile, nei mesi estivi. Si può chiaramente notare l'effetto lockdown, che ha portato ad avere il traffico a livelli molto più bassi di quanto mai visto in precedenza. Nel corso degli anni, fatta eccezione per la particolare primavera di quest'anno, l'influenza del traffico sulle concentrazioni dovrebbe essere rimasta piuttosto invariata, anche se l'ammodernamento della flotta circolante potrebbe sicuramente aver ridotto le emissioni totali, anche a parità di numero di veicoli circolanti.

4.4.1 NO_x

Le prestazioni del modello ottenuto sono risultate in linea con quelle ottenute senza l'inclusione della variabile sul traffico. Questa misura, quindi, non ci permette di fare previsioni più precise di quanto non riuscissimo già a fare, ma ciò non significa che non ci possano essere delle differenze nelle serie normalizzate ottenute.

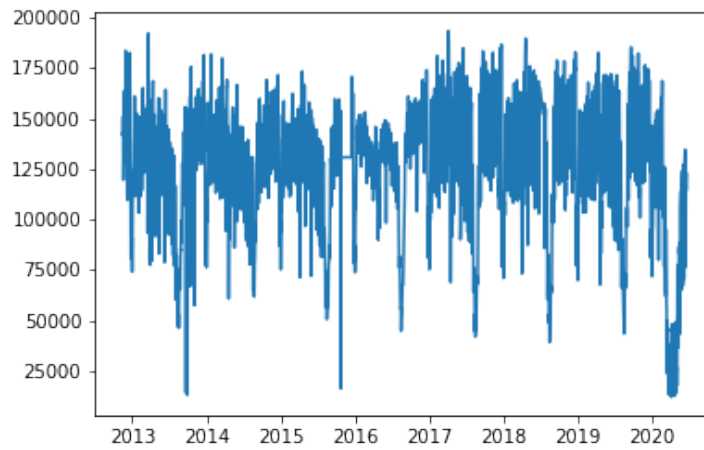


Figura 4.5: Andamento del numero di ingressi in Area C registrati ai varchi

Alla variabile è stata attribuita un'importanza di 0.07, la quinta maggiore tra le undici considerate, che significa che comunque un po' sia utilizzata per fare previsioni. L'importanza delle altre variabili del modello risulta invece praticamente invariata, con temperatura e velocità del vento che hanno continuato ad essere le due più importanti.

Uno strumento utile per verificare come il nostro modello utilizzi le variabili predittive per fare previsioni sono i partial dependence plots. Questi grafici ci mostrano come vari il valore delle previsioni fatte dal nostro modello al variare del valore di una singola variabile, con tutte le altre fissate al loro valore medio.

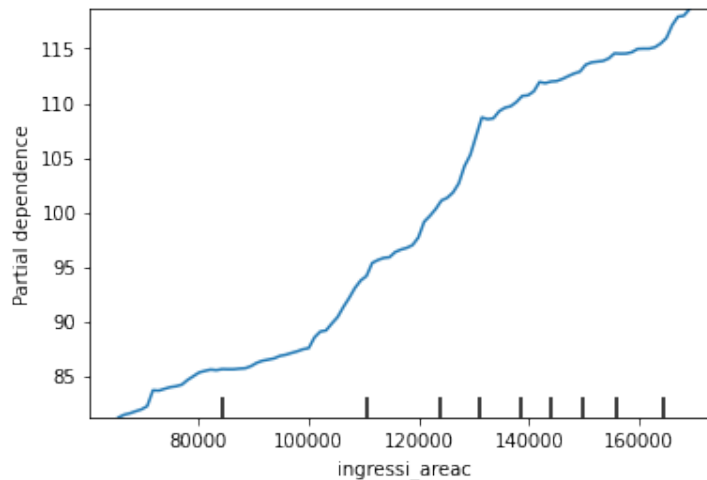


Figura 4.6: Partial dependence plot per il numero degli ingressi in Area C registrati ai varchi

In figura 4.6 è riportato il partial dependence plot della variabile degli ingressi in Area C. Non sorprendentemente, visto il forte collegamento tra questi inquinanti ed il traffico, si vede come le previsioni del nostro modello arrivino ad avere valori decisamente maggiori al crescere del numero di ingressi registrati, con un aumento di oltre $30 \mu\text{g}/\text{m}^3$ se si raddoppia il numero di ingressi.

A primo impatto si nota subito come le due serie siano sempre praticamente equivalenti, mantenendo lo stesso andamento. Si nota anche, però, come per il 2020 le due serie mostrino andamenti molto diversi, con una differenza di circa $10 \mu\text{g}/\text{m}^3$ sulle concentrazioni previste. Come avevamo visto in precedenza è proprio nel 2020 che il traffico ha fatto registrare un importante calo, dovuto alle misure per il contenimento dei contagi da COVID, e si nota chiaramente come in corrispondenza di questo calo le previsioni dei due modelli siano discordanti. Da un lato si vede come il modello costruito senza l'uso del numero di ingressi registrati in Area C

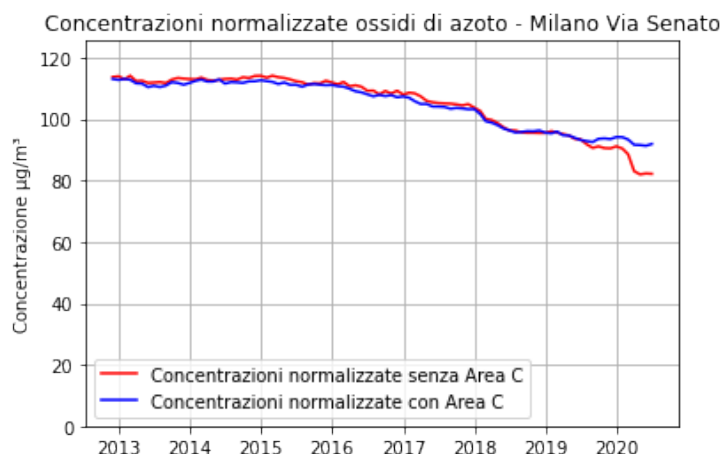


Figura 4.7: Confronto degli andamenti delle serie normalizzate ottenute usando o meno il numero di ingressi in Area C come variabile predittiva

abbia effettivamente rilevato un calo abbastanza importante nelle concentrazioni misurate, dovuto proprio alla diminuzione del traffico. Dall'altro si vede invece come il modello costruito usando anche i dataset di Area C non rilevi questa differenza, continuando a mantenere un andamento in linea col trend decrescente di questi ultimi anni. Questo modello, infatti, essendo stato costruito utilizzando anche i dati sul traffico è capace di eliminarne l'influenza (e ciò avviene sia in condizioni di traffico al di sopra che al di sotto della media) e riportando i dati dei mesi della primavera 2020 a condizioni di traffico medie ha eliminato il calo rilevato in precedenza. Questo ci dà ulteriormente conferma dell'origine da attribuire a tale calo ed inoltre ci dà una stima di come mediamente potrebbero cambiare le concentrazioni se venisse completamente (o quasi, proprio come è successo nei mesi di lockdown) eliminato il traffico veicolare privato. Si è visto quindi come la sua quasi totale eliminazione porterebbe ad avere un miglioramento delle concentrazioni di ossidi di azoto di circa il 10%, che è un valore molto ridotto se si considera il drastico impatto sull'economia dei provvedimenti presi.

Per quanto riguarda gli NO_x abbiamo quindi dimostrato come effettivamente anche i nostri modelli siano in grado di individuare l'influenza del traffico sulle concentrazioni e abbiamo provato anche a quantificare il possibile miglioramento della qualità dell'aria che si otterrebbe eliminandolo completamente. Un'altra nota da fare è che i modelli creati utilizzando anche i dati degli ingressi in Area C individuano comunque lo stesso trend decrescente nelle concentrazioni già visto in precedenza, mostrandoci quindi come questo andamento non sia associabile al volume del traffico, ma che le sue cause siano da ricercare altrove. Questa è un'ulteriore conferma di come l'innovazione tecnologica sia il maggior responsabile del miglioramento delle concentrazioni di questo inquinante, oltre che al modo più efficace per la loro riduzione.

4.4.2 NO_2

Per completezza è stato trattato anche il biossido di azoto.

Anche per questo inquinante valgono le stesse considerazioni fatte per gli ossidi. Importante notare come anche questa volta ci sia una differenza di andamento per l'anno 2020 associabile proprio al calo del traffico circolante dovuto al lockdown.

4.4.3 CO

Le stesse prove, effettuate con la creazione di un modello per la trattazione che faccia uso anche del numero degli ingressi in Area C, sono state fatte sul monossido di carbonio.

Anche se l'aumento rilevato al raddoppiare del traffico non è sicuramente consistente come quelli riguardanti gli ossidi di azoto, anche in questo caso possiamo

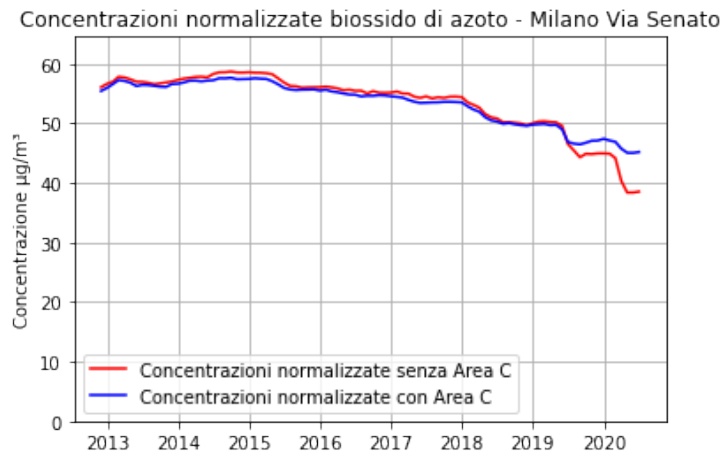


Figura 4.8: Confronto degli andamenti delle serie normalizzate ottenute usando o meno il numero di ingressi in Area C come variabile predittiva

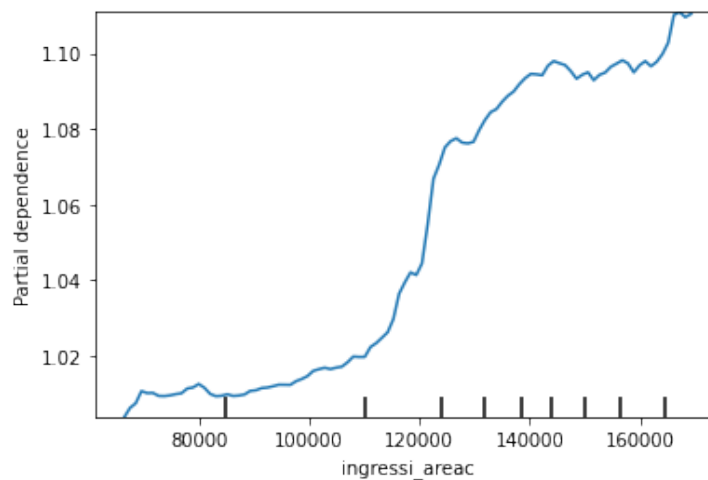


Figura 4.9: Partial dependence plot per il numero degli ingressi in Area C registrati ai varchi

notare come il nostro modello associ una crescita delle concentrazioni al crescere del traffico.

Guardando il grafico 4.10 si nota nuovamente come le serie ottenute siano sempre piuttosto equivalenti. Ci sono due piccole discrepanze sui risultati ottenuti, una riguardante il periodo dell'epidemia di COVID ed una a cavallo tra gli anni 2016 e 2017. Per il periodo riguardante l'epidemia di COVID si nota la stessa situazione vista per gli ossidi di azoto, con il modello costruito senza i dati di Area C che fa previsioni leggermente più basse rispetto all'altro, potenzialmente collegabili alla diminuzione del traffico. Per quanto riguarda il periodo tra il 2016 ed il 2017 si vede invece come le previsioni di tale modello siano più alte, come se l'influenza del traffico in tale periodo sia stata maggiore. Controllando il grafico del numero di ingressi visto in precedenza, però, non si notano aumenti nel numero di veicoli circolanti che possano giustificare tale differenza, quindi le cause potrebbero essere altre oppure si potrebbe semplicemente trattare di una differenza delle previsioni dei due modelli che comunque può esserci, trattandosi di previsioni basate su modelli statistici e che quindi hanno una certa variabilità. Trattandosi comunque entrambe di differenze minime è difficile stabilire con certezza le cause che stanno alla loro origine.

Per il monossido di carbonio, quindi, l'influenza del traffico sembra essere davvero minima, anche perché si sta comunque parlando di concentrazioni molto basse. L'innovazione tecnologica del mondo delle auto ha fatto sì che questo inquinante ormai non sia più un effettivo problema.

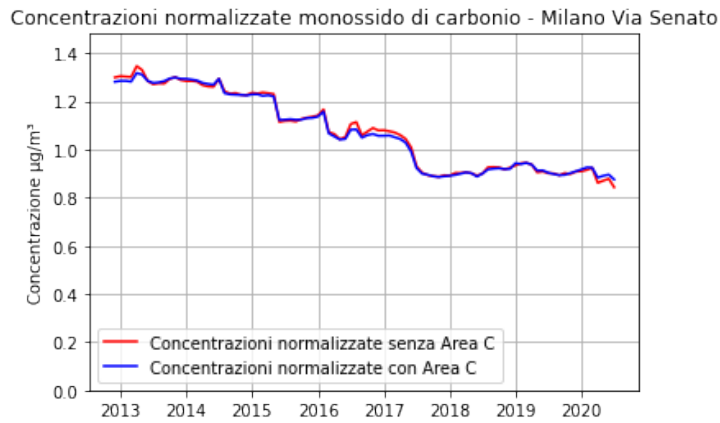


Figura 4.10: Confronto degli andamenti delle serie normalizzate ottenute usando o meno il numero di ingressi in Area C come variabile predittiva

4.4.4 PM10

L'esperimento è stato ripetuto anche per le polveri sottili (in particolare il PM10 - per il PM2.5 si ha comunque una situazione praticamente analoga).

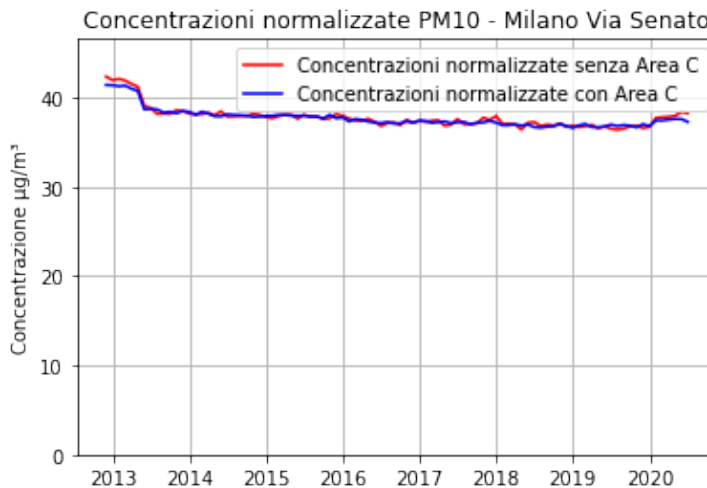


Figura 4.11: Confronto degli andamenti delle serie normalizzate ottenute usando o meno il numero di ingressi in Area C come variabile predittiva

Vediamo infatti come le due serie rimangano praticamente sempre equivalenti, non mostrando particolari scostamenti neanche nel periodo dell'epidemia di COVID.

Per quanto riguarda le polveri sottili, quindi, il nostro metodo non rileva una particolare influenza dell'andamento del traffico sulle concentrazioni registrate, nemmeno a seguito di un calo del volume come quello che è avvenuto nel lockdown, mostrandoci ancora una volta come le fonti emissive più influenti per questi inquinanti non siano i veicoli a motore.

Capitolo 5

Effetti del lockdown per il Covid-19 sulle concentrazioni degli inquinanti

I mesi primaverili del 2020 sono stati caratterizzati dalla diffusione dell'epidemia di Covid-19 e dalle importanti misure prese per il suo contenimento. Queste limitazioni hanno portato ad avere un quadro emissivo eccezionale per gli inquinanti atmosferici, che difficilmente si sarebbe potuto verificare in condizioni normali. Questi mesi, infatti, sono stati caratterizzati da importanti limitazioni sugli spostamenti e alle attività produttive, quindi rappresentano un importante banco di prove per verificare come siano cambiate le concentrazioni degli inquinanti in atmosfera con una così drastica riduzione delle attività antropiche, dandoci la possibilità di valutare come effettivamente si potrebbe intervenire per ottenere ulteriori miglioramenti della qualità dell'aria.

L'obiettivo delle nostre analisi è stato mettere alla prova la nostra tecnica di normalizzazione per vedere come siano effettivamente cambiate le concentrazioni degli inquinanti durante i mesi dell'epidemia rispetto agli anni precedenti, dopo aver eliminato la variabilità derivata dalle condizioni meteorologiche e stagionali. È infatti importante considerare che l'epidemia è arrivata durante dei mesi che già normalmente sono abbastanza favorevoli per la maggior parte degli inquinanti di interesse, poiché le condizioni atmosferiche e meteorologiche dei mesi primaverili portano sempre ad un abbassamento delle concentrazioni misurate rispetto ai mesi invernali. Potendo eliminare la variabilità data dalla stagionalità e da queste condizioni meteorologiche che influenzano così tanto i valori misurati è quindi sicuramente interessante cercare di capire quanto i classici cali delle concentrazioni registrati nei mesi primaverili siano stati influenzati dal blocco generale avuto dalle attività umane in tali mesi.

Abbiamo quindi trattato tutti gli inquinanti, confrontando l'andamento dei mesi tra febbraio e maggio del 2020 con quello degli anni tra il 2015 e il 2019 della media delle serie, normalizzate e non, dei capoluoghi di provincia già usate per le analisi precedenti. In questo modo potremo verificare se per i mesi dell'epidemia il nostro modello ha rilevato dei cali riconducibili proprio alle limitazioni imposte, in modo da capire quale tipo di interventi possa risultare più efficace per il contenimento delle concentrazioni di ciascuno.

5.1 NO_x e NO_2

5.1.1 NO_x

Abbiamo iniziato le nostre analisi trattando gli ossidi di azoto che, viste le limitazioni agli spostamenti e considerato il loro collegamento con l'andamento del traffico, potrebbero essere potenzialmente gli inquinanti più colpiti durante questi mesi di lockdown. Ad influire, inoltre, potrebbe sicuramente essere stato anche il parziale calo delle emissioni industriali, dovuto al fermo delle attività non ritenute essenziali.

Guardando il grafico delle concentrazioni misurate 5.1 si nota chiaramente il calo delle concentrazioni classico dei mesi primaverili, causato proprio dalle condizioni più

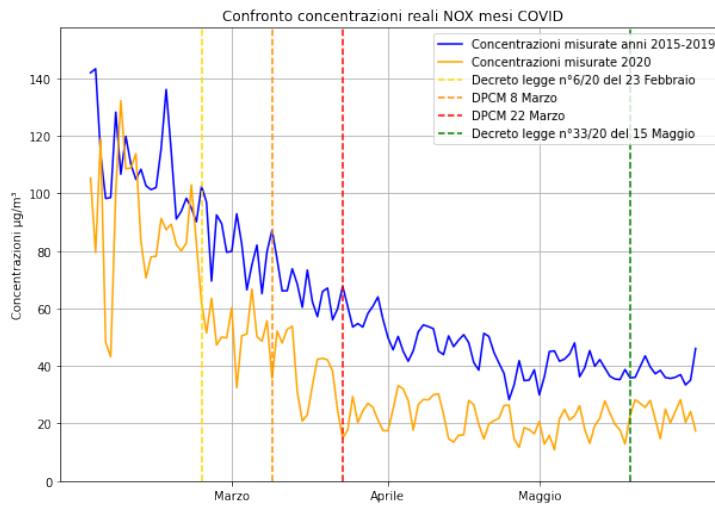


Figura 5.1: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

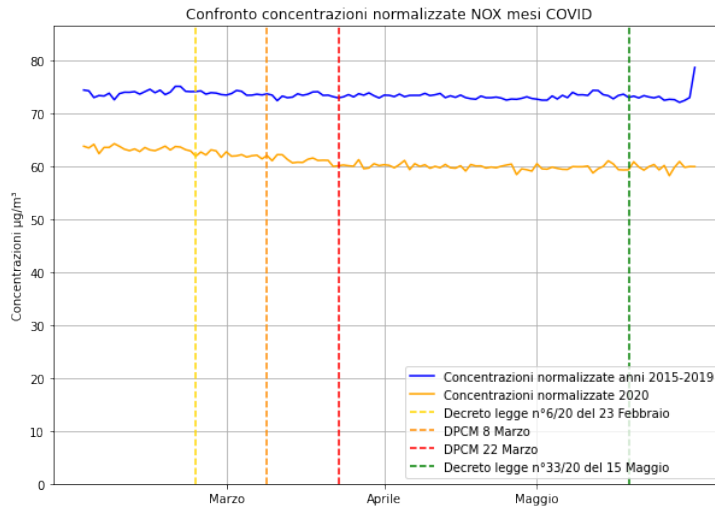


Figura 5.2: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

favorevoli alla dispersione. Tale calo, chiaramente, risulta invece eliminato quando si considerano le serie normalizzate (grafico 5.2). Guardando sia i grafici delle concentrazioni reali che quello delle serie normalizzate si osserva come nei mesi di *lockdown* le concentrazioni abbiano subito un discreto ribasso, di circa il 20%, rispetto agli anni precedenti, sicuramente dovuto al forte blocco del traffico e di alcune attività produttive.

5.1.2 NO_2

Abbiamo verificato anche i dati relativi al biossido di azoto, aspettandoci comunque di ottenere risultati abbastanza simili a quelli visti sugli ossidi trattati nel loro complesso.

Anche in questo caso, infatti, oltre al classico calo dei mesi primaverili, si vede una differenza tra le concentrazioni del 2020 e quelle degli anni precedenti, sia considerando le concentrazioni misurate che guardando le serie normalizzate. Anche in questo caso le cause sono sicuramente da attribuire alle stesse ragioni che hanno portato al calo degli ossidi.

Chiaramente una misura così restrittiva non è sicuramente applicabile in modo continuo, ma ci dimostra come interventi su larga scala possano effettivamente avere un impatto sulle concentrazioni a differenza di quello che si era visto analizzando dei

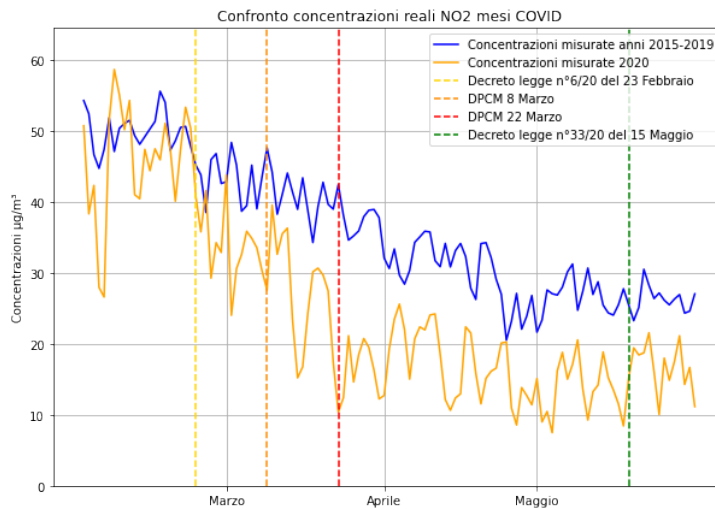


Figura 5.3: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

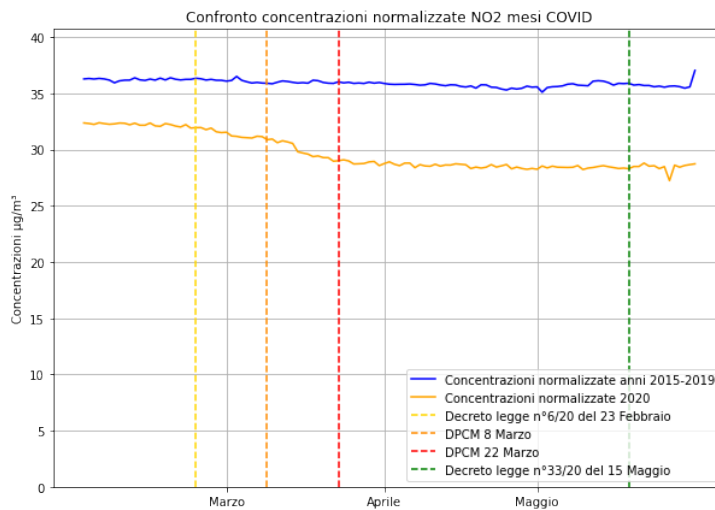


Figura 5.4: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

provvedimenti presi in particolari località, i cui effetti si sono rivelati minimi se non addirittura nulli. Interventi su larga scala che possano avere un'efficacia richiederanno investimenti sull'innovazione tecnologica, per poter arrivare ad avere una flotta di veicoli circolanti a sempre più basse emissioni, così come potrebbe essere sicuramente utile investire sullo sviluppo di motorizzazioni "pulite" come quella elettrica, anche se prima rimangono da risolvere importanti questioni di sostenibilità che al momento non gli consentono di competere come sostituti dei tradizionali motori a scoppio [34, 18]. Chiaramente questo tipo di interventi richiedono tempo e ci dimostra come non sia possibile intervenire sugli inquinanti pensando di abbatterne le concentrazioni rapidamente e senza piani concreti sul medio e lungo termine.

5.2 PM10 e PM2.5

Per quanto riguarda le polveri sottili e l'andamento delle concentrazioni nel periodo di lockdown la situazione risulta essere più complessa rispetto agli ossidi di azoto. Se da un lato, infatti, i cali di traffico e produzione industriale hanno sicuramente portato ad un calo delle emissioni, dall'altro si è visto un aumento di quelle derivanti da riscaldamento domestici, causato da un maggior fabbisogno giornaliero delle famiglie costrette a restare più tempo nell'abitazione a causa delle misure imposte.

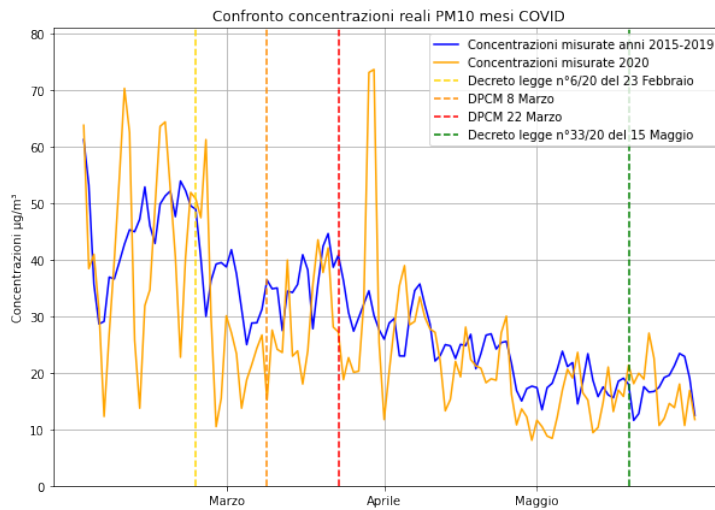


Figura 5.5: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

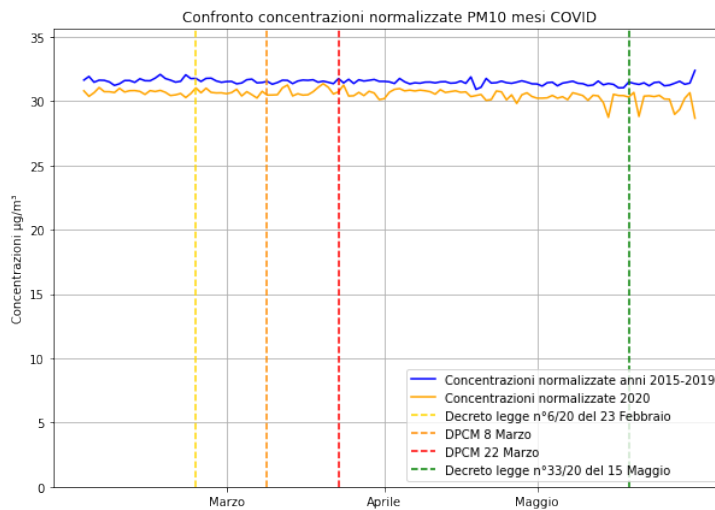


Figura 5.6: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

Per quanto riguarda la frazione più spessa del particolato (PM10) vediamo come le concentrazioni misurate si mantengano piuttosto in linea con quanto visto negli anni precedenti, con episodi come quello di fine Marzo in cui le concentrazioni hanno fatto registrare valori ben al di sopra del limite di legge, nonostante fossero già in vigore tutte le limitazioni imposte coi vari DPCM per il contenimento dei contagi. La serie normalizzata^{5.6}, invece, evidenzia un calo trascurabile rispetto agli anni precedenti, che però potrebbe essere collegato al leggero trend decrescente che era già stato visto sulle serie normalizzate di questo inquinante.

Esattamente come visto in precedenza, anche per la frazione più fine di particolato le concentrazioni si mantengono abbastanza invariate rispetto agli anni precedenti, con una piccola discrepanza sulle serie normalizzate nuovamente dovuta alla presenza di un trend leggermente calante individuato per le serie normalizzate di questo inquinante^{5.8}.

Per quanto riguarda il particolato, quindi, nonostante la riduzione del traffico e delle attività produttive, l'ipotizzabile maggior fabbisogno energetico delle famiglie potrebbe aver fatto sì che le concentrazioni rimanessero comunque su livelli più o meno simili a quelli degli anni precedenti. Questo ci mostra come effettivamente per trattare questi inquinanti, la cui origine è sicuramente varia e complessa, non si possano prendere provvedimenti riguardanti solamente un settore in particolare, come potrebbe essere il traffico, ma vanno pensati dei piani che riguardino tutte le

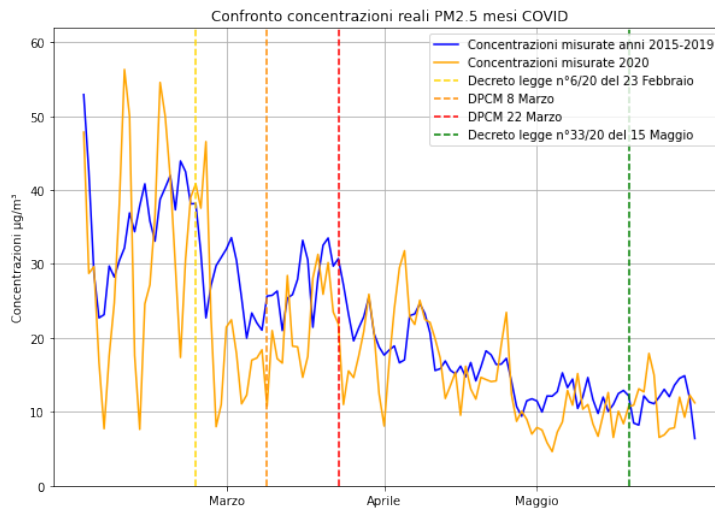


Figura 5.7: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

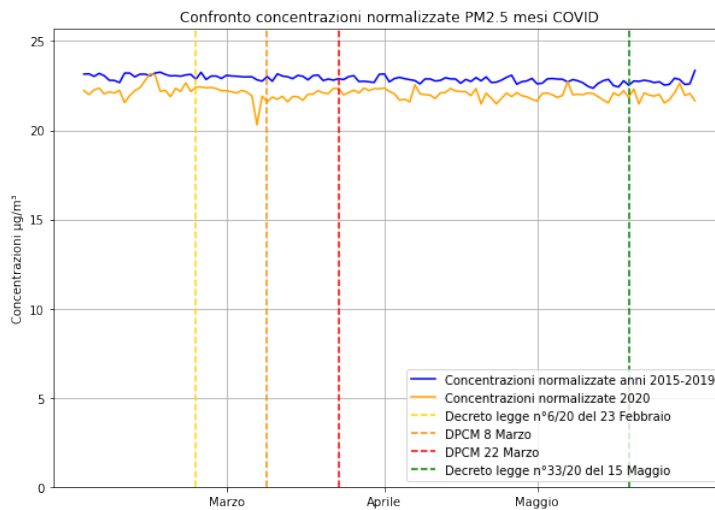


Figura 5.8: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

fonti emissive. La presenza di un trend negativo per tutte e due le frazioni di particolato è sicuramente un'indicazione di come la strada dell'innovazione tecnologica, che negli ultimi anni sta coinvolgendo tutti i settori, come è necessario che succeda per poter pensare di agire efficacemente su un inquinante così complesso e di origini varie, sia quella giusta da percorrere nonché l'unica che può portare ad avere risultati concreti. Si è visto infatti come, seppur il lockdown abbia drasticamente ridotto il numero di veicoli circolanti, questo non abbia portato a miglioramenti della situazione, ad ulteriore conferma di come non si possa accreditare il traffico come il principale responsabile delle emissioni di particolato.

5.3 CO

Anche per quanto riguarda il monossido di carbonio il quadro emissivo è cambiato durante i mesi dell'epidemia, con l'importante riduzione del traffico che però anche in questo caso è stata bilanciata dal maggior uso dei riscaldamenti domestici. Traffico e combustioni non industriali, infatti, sono le due categorie maggiormente responsabili delle emissioni di questo inquinante in atmosfera.

Successivamente all'applicazione delle limitazioni tramite l'emanazione dei DPCM del mese di Marzo si nota come effettivamente le concentrazioni misurate di mo-

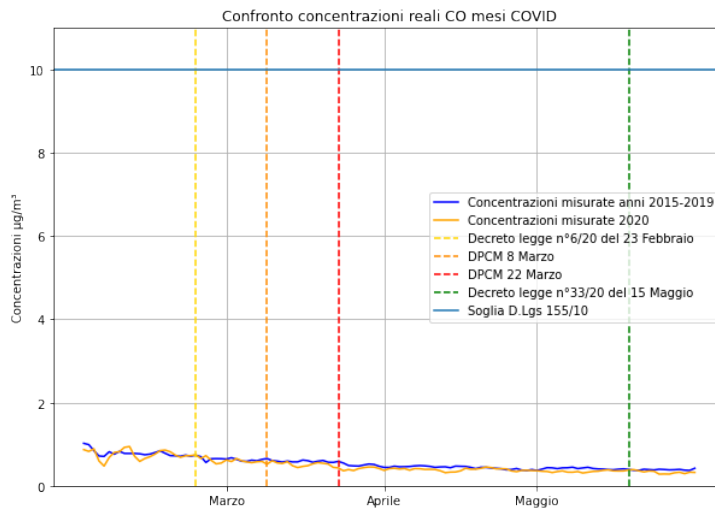


Figura 5.9: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

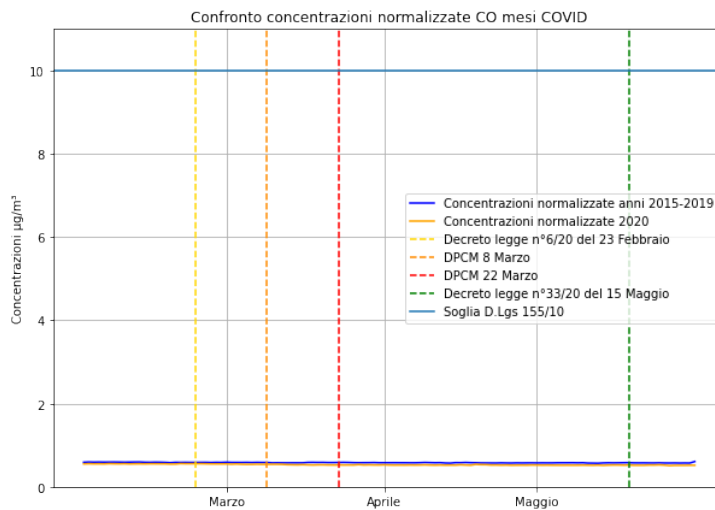


Figura 5.10: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

nossido di carbonio risultino essere leggermente minori rispetto a quelle degli anni precedenti^{5.9}. Tale differenza viene confermata anche dalle serie normalizzate, anche se bisogna ricordare che si tratta di differenze minime su valori già molto bassi.

Le limitazioni imposte, quindi, potrebbero aver avuto qualche effetto sulle concentrazioni di monossido di carbonio, anche se ormai la situazione per questo inquinante non desta più particolari preoccupazioni e quindi c'è anche meno interesse nella ricerca di nuove misure di contenimento, visto che quelle già presenti risultano essere più che sufficientemente efficaci.

5.4 Benzene

Il benzene è sicuramente un inquinante importante da analizzare in questo periodo, visto che spesso viene indicato come tracciante del traffico veicolare (specialmente quello motorizzato a benzina) che proprio in questi mesi ha subito importanti limitazioni e potrebbe quindi aver portato ad un calo delle concentrazioni misurate.

Anche per questo inquinante, così collegato al traffico, notiamo uno scostamento rispetto all'andamento degli anni precedenti del 20% ($0.2\mu\text{g}/\text{m}^3$), compatibile con quello visto sugli ossidi di azoto. Questo scostamento, confermato anche dalle serie normalizzate, conferma come effettivamente il traffico influenzi le concentrazioni di

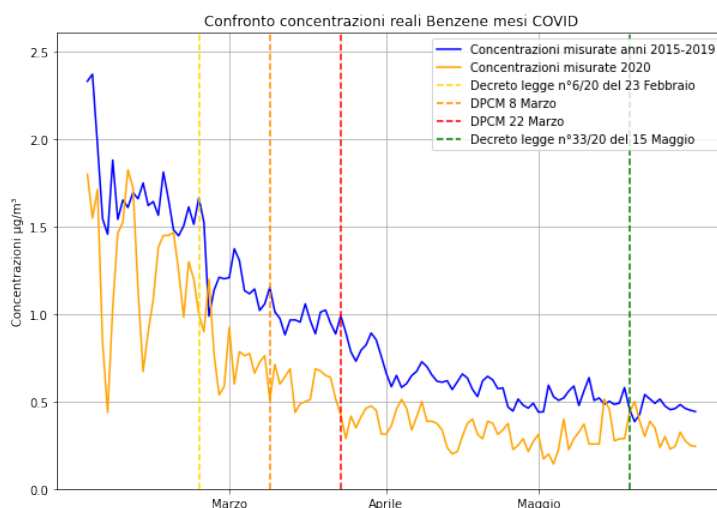


Figura 5.11: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

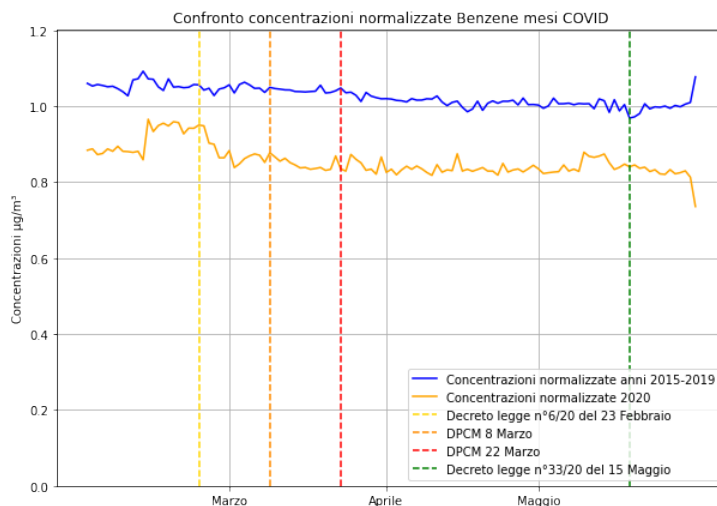


Figura 5.12: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

questo inquinante e quindi come una sua riduzione abbia conseguentemente portato ad un abbassamento delle concentrazioni, anche se è importante ricordare che si sta parlando di valori ben al di sotto della soglia di legge ($5\mu g/m^3$ per la media oraria) e che pertanto non destano più preoccupazioni per i possibili effetti sulla salute umana.

Sebbene la situazione per quanto riguarda il benzene non desti preoccupazioni, visto che i livelli ormai sono ampiamente sotto alle soglie di legge, abbiamo visto come una riduzione del traffico abbia portato ad un ulteriore miglioramento della situazione per quanto riguarda questo inquinante, che è auspicabile venga portato avanti dall'innovazione tecnologica nel corso dei prossimi anni.

5.5 SO₂

Analizziamo l'andamento delle concentrazioni di biossido di zolfo durante i mesi di lockdown, per vedere se notiamo differenze rispetto agli anni precedenti, pur ricordando che si sta trattando un inquinante le cui concentrazioni sono ormai prossime al fondo naturale.

Si vede infatti come le concentrazioni si siano mantenute su un livello praticamente in linea con quello degli anni precedenti e come anche le serie normalizzate siano piuttosto simili, anche se per il 2020 risulta effettivamente esserci stato un

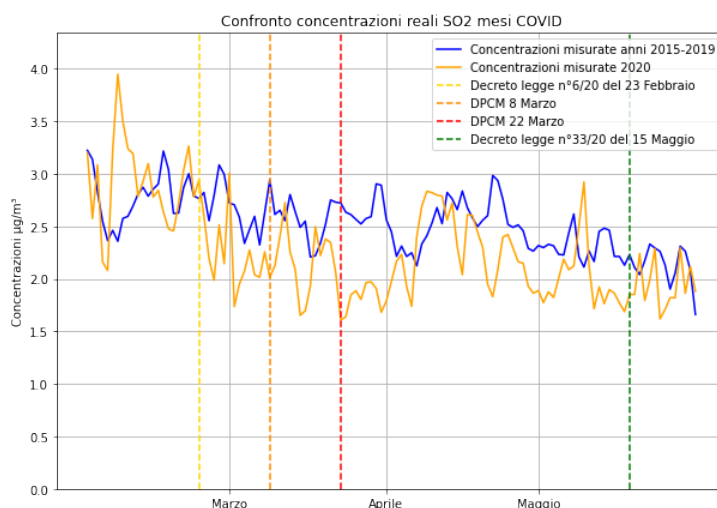


Figura 5.13: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

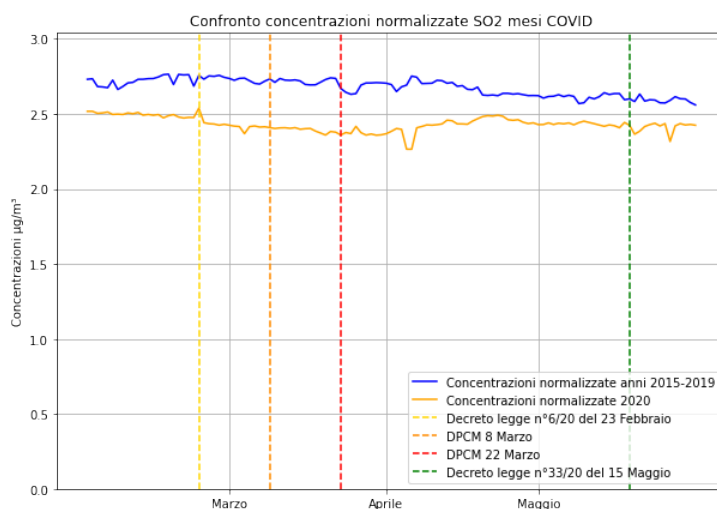


Figura 5.14: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

leggero calo, di cui però, vista la modesta quantità, è difficile stabilire se sia dovuto alle limitazioni imposte durante il lockdown o se la sua origine sia anche dovuta ad altri fattori (come ad esempio fluttuazioni nel fondo naturale).

5.6 Ozono

L'ozono, essendo un inquinante secondario, ha origine molto varie e quindi può essere interessante valutare quali siano state le concentrazioni registrate a seguito dei provvedimenti di lockdown.

Vediamo come le concentrazioni per quest'anno si siano mantenute abbastanza in linea con quelle degli anni precedenti mostrando il classico trend crescente dei mesi primaverili che caratterizza questo inquinante, con la serie normalizzata che segnala un aumento, dovuto probabilmente anche alla presenza del trend positivo già visto in precedenza.

Nonostante i grossi cambiamenti visti nei mesi di lockdown, quindi, l'ozono non risulta essere stato colpito da tali misure. Questo inquinante, considerate le sue complesse e varie origini, risulta quindi difficile da trattare e trovare misure per il suo contenimento è molto complesso. Si è comunque visto come un blocco generale delle attività non sia servito a migliorarne la situazione.

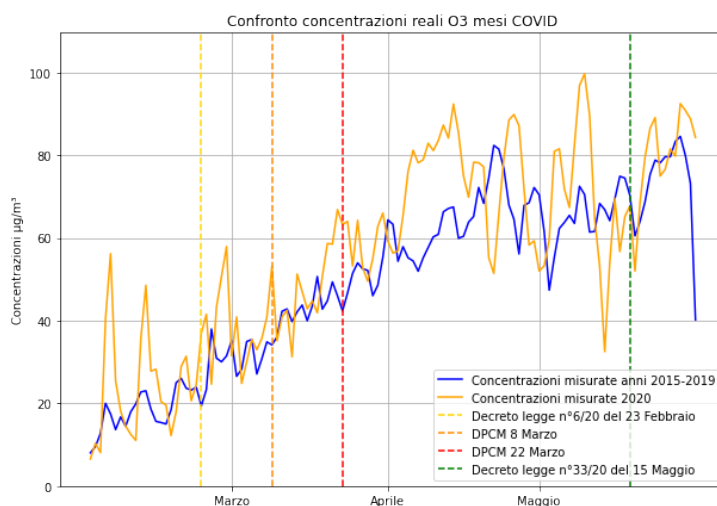


Figura 5.15: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

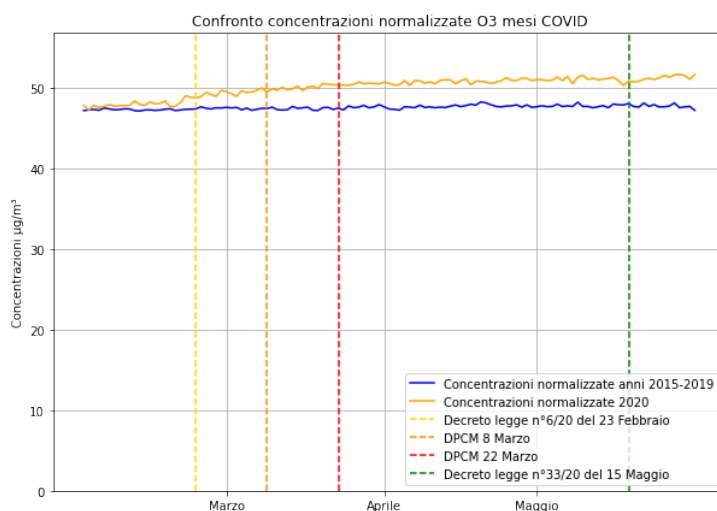


Figura 5.16: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

5.7 Ammoniaca

In ultimo andiamo a controllare l'andamento dell'ammoniaca, per la quale le serie normalizzate dei nostri modelli avevano individuato un andamento piuttosto fluttuante.

Effettivamente controllando le concentrazioni misurate si nota come tra Marzo e Maggio si sia visto un effettivo calo dei valori misurati, individuato anche dalle serie normalizzate. Questo calo probabilmente deriva da una minor attività anche per quanto riguarda il settore agricolo, che è praticamente l'unico responsabile delle concentrazioni misurate di questo inquinante, che è stato costretto a rallentare durante i mesi di lockdown.

Visto il forte collegamento tra questo inquinante e l'agricoltura qualsiasi misura di contenimento delle concentrazioni dovrà necessariamente agire su tale settore, poiché effettivamente con un calo dell'attività si è notato anche un calo nelle concentrazioni rilevate.

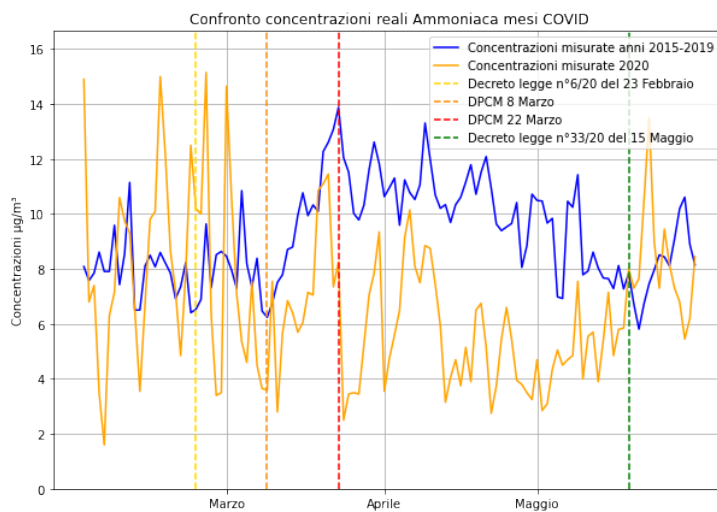


Figura 5.17: Confronto tra l'andamento della serie misurata del 2020 con la media di quelle degli anni 2015-2019

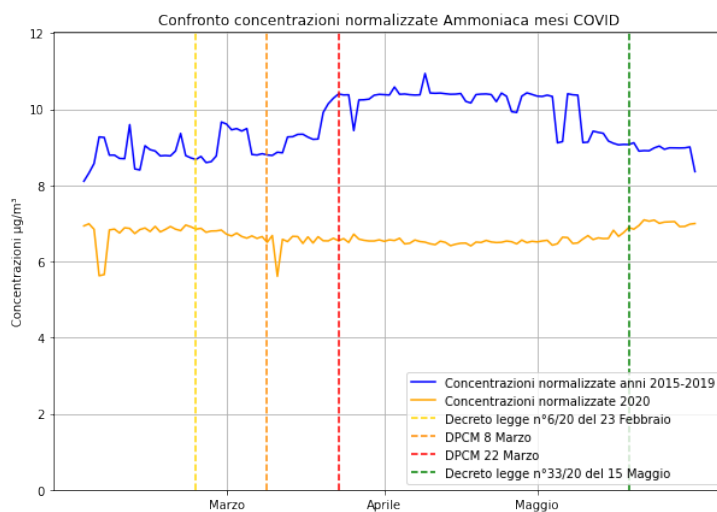


Figura 5.18: Confronto tra l'andamento della serie normalizzata del 2020 con la media di quelle degli anni 2015-2019

Capitolo 6

Appendice sulla validità dei modelli nel corso degli anni

A seguito della lettura di un interessante articolo [3] che, tramite l'uso di diverse tecniche di ML, tra cui anche Random Forest, ha provato a realizzare dei modelli per fare previsioni sulla qualità dell'aria della città di Milano utilizzando gli stessi dataset che sono stati impiegati anche nei nostri studi, ma limitatamente al periodo 2013-2016, e che suggeriva la possibilità di adottare dei modelli per la previsione della qualità dell'aria senza la necessità di avere i sensori sul campo, ho voluto provare a verificare come la nostra tecnica si potesse comportare in merito. Per verificare la possibilità di continuare ad utilizzare gli stessi modelli nel corso degli anni, controllandone quindi le performance ed i risultati ottenuti, l'idea è stata quella di creare dei modelli utilizzando i dati antecedenti ad una data per la fase di training, per poi andare invece a testarli sul periodo seguente, in modo da simulare la creazione di un modello in uno specifico momento per poi andare a verificare come si sarebbe comportato quel modello se fosse stato utilizzato nel futuro. Se i modelli avessero mostrato comunque buone performance e risultati vicini a quelli ottenuti utilizzando tutto il range di dati a disposizione si sarebbe potuto concludere che, effettivamente, sia possibile realizzare dei modelli per la previsione della qualità dell'aria senza aver bisogno di continuare a tenere operativi tutti i sensori sul campo.

La città che abbiamo scelto di prendere in considerazione è stata quella di Mantova, poiché risultava il capoluogo per il quale le serie delle variabili utilizzate iniziavano ad essere disponibili da prima di tutte le altre città. Come detto, per la costruzione dei modelli è stata quindi fissata una data perno: le rilevazioni antecedenti venivano usate per la costruzione del modello (training), mentre quelle successive per testarne le prestazioni. Inoltre, a tali modelli, sono state fatte costruire le serie normalizzate fino al 2020, in modo da poterle confrontare con quelle ottenute nelle analisi precedenti. Si è quindi deciso di fare questa verifica su tutti gli inquinanti a disposizione, facendo diverse prove per ciascuno, andando a variare la data perno di un anno alla volta, partendo dal 2010 fino al 2019. In tutti i modelli realizzati, indipendentemente dall'inquinante o dalla data perno scelta, si è però incorsi nel problema del *covariate shift* [28], ovvero la degradazione delle performance dei modelli di ML quando le fasi di training e di testing avvengono utilizzando due campioni con distribuzioni diverse. Nel nostro caso questo problema, che con tecniche come Random Forest risulta parecchio limitante, non permette quindi di ottenere modelli sufficientemente precisi quando sono costretti a trattare un periodo i cui dati non sono stati usati nella fase di costruzione.

A titolo esemplificativo vengono riportati i risultati ottenuti sul modello costruito per trattare il biossido di azoto (NO_2), prendendo come data perno il 1-1-2014, ma, come già detto, la stessa situazione è stata riscontrata in tutti i casi considerati. Il modello che avevamo creato in precedenza, utilizzando i dati di tutti gli anni a disposizione, in fase di testing era risultato avere un RMSE di $7.43 \mu g/m^3$ ed un coefficiente R^2 di 0.74. Il modello creato utilizzando i dati degli anni antecedenti al 2014, e poi testato sulle rilevazioni del periodo successivo, ha invece ottenuto un indice RMSE di $10.81 \mu g/m^3$ (+25% rispetto al modello "classico") ed un coefficiente R^2 di 0.23 (-66%), che ci mostrano, soprattutto il secondo, proprio come il nostro

modello abbia sofferto del cosiddetto covariate shift, che ne ha quindi degradato completamente le performance, rendendo il modello pressoché inutilizzabile. Questo peggioramento delle performance ha chiaramente poi inciso anche sulle serie normalizzate ottenute, che successivamente alla data perno tendono a stabilizzarsi intorno ad un valore, mostrando proprio come in quei casi il modello non sia più capace di capire le dinamiche che portano ad avere variazioni sulle concentrazioni. 6.1

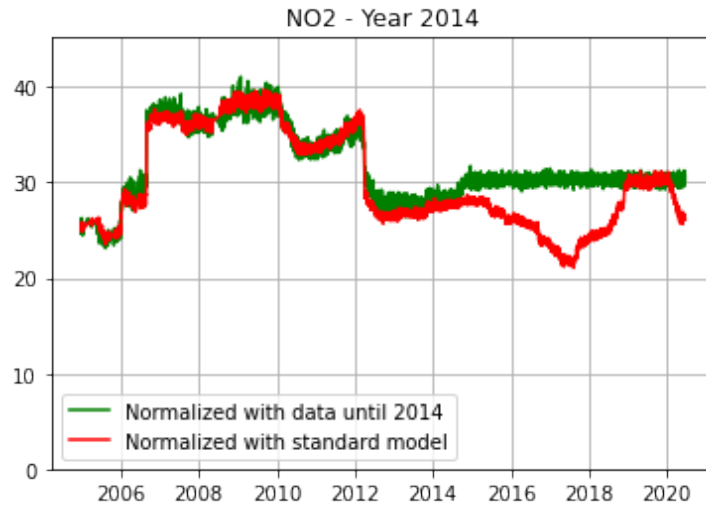


Figura 6.1: Confronto tra le serie normalizzate ottenute col modello classico (in rosso) e quella ottenuta dal modello costruito con i dati antecedenti al 2014 (in verde)

Dai risultati ottenuti, quindi, è emersa l'impossibilità di utilizzare la nostra tecnica, ma in generale anche altri algoritmi di ML, per creare dei modelli in grado di fare previsioni efficaci sulla qualità dell'aria che continuino a restare validi nel tempo, permettendo quindi di non utilizzare più sensori sul campo, poiché periodicamente è necessario farne un re-training utilizzando anche i nuovi dati a disposizione, così che gli stessi possano continuare a comprendere l'evoluzione della situazione nel corso del tempo.

Capitolo 7

Conclusioni

Nelle nostre analisi abbiamo dimostrato come sia possibile applicare l'algoritmo Random Forest per eliminare l'influenza della meteorologia dalle serie dei principali inquinanti atmosferici, in modo da poterne analizzare l'andamento e le variazioni in modo più preciso. Controllare l'influenza della meteorologia in analisi di questo tipo è fondamentale, poiché essa risulta per la maggior parte delle volte maggiore di quella che possono avere limitazioni o particolari eventi. Random Forest, così come molte altre tecniche del mondo del machine learning, si adattano perfettamente a questo tipo di modellazione e permettono di ottenere dei risultati validi ed affidabili, facendo affidamento sulla capacità di calcolo dei calcolatori moderni. Inoltre, includendo nei nostri modelli anche una variabile relativa al traffico, che nel nostro caso è stata il numero di ingressi in Area C registrati nella città di Milano, è stato possibile utilizzare la nostra tecnica per stimare le variazioni che si hanno sulle concentrazioni degli inquinanti a fronte di variazioni nei volumi di traffico. Questo approccio è chiaramente generalizzabile e mostra come sia possibile utilizzare la nostra tecnica per stimare l'influenza sulle concentrazioni di un determinato fenomeno o agente semplicemente includendo nei modelli una variabile che lo tracci e poi andando a confrontare le discrepanze tra le serie normalizzate con e senza l'eliminazione della sua influenza. Applicando la nostra tecnica per ottenere la normalizzazione meteorologica ai dati raccolti sul territorio della Regione Lombardia si è visto come ormai l'inquinamento atmosferico non sia più un problema grave, con la maggior parte degli inquinanti che sono ormai stabilmente sotto ai livelli classificati come pericolosi per la salute, fatta eccezione per ossidi di azoto e polveri sottili che rimangono gli ultimi elementi a rappresentare una criticità. Il trend decrescente presente per quasi tutti gli inquinanti ci mostra come effettivamente il progresso tecnologico degli ultimi anni sia fondamentale per la riduzione dell'inquinamento e come questa sia probabilmente l'unica soluzione efficace per quelli che ancora oggi si mantengono su livelli più preoccupanti. Riguardo all'efficacia dei provvedimenti presi per contrastare l'inquinamento si è visto come quelli attivati solo localmente, come possono essere state Area C o Area B per il comune di Milano, siano praticamente ininfluenti sia a causa del fatto che gli inquinanti in atmosfera possono viaggiare anche diversi chilometri che perché ormai l'influenza del traffico, grazie alle numerose innovazioni introdotte nel corso degli anni, è calata in modo consistente. Includendo come variabile dei nostri modelli anche il numero di ingressi in Area C abbiamo potuto verificare l'impatto sulle concentrazioni causato da variazioni nei volumi di traffico ed abbiamo visto come, a seguito di un'importante riduzione avuta nei mesi di Marzo e Aprile 2020, le concentrazioni siano cambiate: quelle degli inquinanti più caratteristici di tale settore, ovvero gli ossidi di azoto ed il benzene, hanno riportato un calo intorno al 20%. Per tutti gli altri, invece, non si sono viste variazioni nonostante la forte riduzione di tutte le attività che si è avuta in quei mesi. L'inquinamento è un fenomeno molto complesso, in cui entrano in gioco molti attori che ne complicano la trattazione, e spesso non è facile trovare provvedimenti realmente efficaci per intervenire in modo concreto. La nostra tecnica, così come potenzialmente l'uso di altri algoritmi di Machine Learning, può quindi essere un utile strumento per valutare l'efficacia di provvedimenti presi o l'influenza di determinati settori, permettendo quindi di studiare misure sempre più efficaci per affrontare tale fenomeno. Si è visto come non si possa pensare che un solo settore, come può essere il traffico, possa essere considerato l'unico responsabile del proble-

ma e che quindi sia necessario intervenire con misure che coinvolgano tutte le fonti emissive. In tal senso nei prossimi anni verrà sicuramente in soccorso l'innovazione tecnologica che continuerà a coinvolgere tutti gli ambiti e che continuerà a favorire il miglioramento della qualità dell'aria grazie alla riduzione delle emissioni. Il nostro lavoro ha inoltre dimostrato come sia importante avere a libera disposizione tutti i dati relativi all'inquinamento, poiché questo permette ai più curiosi di informarsi sul problema e di verificare personalmente qualsiasi informazione ad esso relativa. Considerata la scarsa informazione della popolazione italiana sul tema e che, purtroppo, spesso anche le notizie che vengono fornite risultano imprecise, questi dati rappresentano uno strumento molto importante per la formazione dei cittadini, oltre che ad un importante mezzo per effettuare diverse analisi sfruttando le potenzialità degli algoritmi di machine learning.

Bibliografia

- [1] PBL Netherlands Environmental Assessment Agency, R Hoogerbrugge, HAC Denier van der Gon, MC van Zanten, and J Matthijsen. *Trends in Particulate Matter*. PBL Netherlands Environmental Assessment Agency, 2010.
- [2] Magne Aldrin and Ingrid Hobæk Haff. Generalised additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment*, 39(11):2145–2155, 2005.
- [3] Edoardo Arnaudo, Alessandro Farasin, and Claudio Rossi. A comparative analysis for air quality estimation from traffic and meteorological data. *Applied Sciences*, 10(13), 2020.
- [4] I Barmpadimos, C Hueglin, J Keller, S Henne, and ASH Prévôt. Influence of meteorology on pm10 trends and variability in switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics*, 11(4):1813, 2011.
- [5] Max Bramer. Avoiding overfitting of decision trees. *Principles of data mining*, pages 119–134, 2007.
- [6] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [7] Leo Breiman. Statistical modeling: The two cultures. *Quality control and applied statistics*, 48(1):81–82, 2003.
- [8] G Cattani, A Bernetti, AM Caricchia, R De Lauretis, S De Marco, A Di Menno di Bucchianico, A Gaeta, G Gandolfo, and E Taurino. Analisi dei trend dei principali inquinanti atmosferici in italia 2003-2012. *ISPRA, Rapporti*, 203:2014, 2014.
- [9] D Dina, M Inglessis, C Notaro, G Settimo, and G Viviano. Concentrazione di materiale particolato pm10 e pm2,5 in area urbana: l’influenza di alcuni parametri meteorologici. *Environment including global change*, 2009.
- [10] Joe R Eagleman. Air pollution meteorology. 1991.
- [11] European Commission: DG Environment and Kantar. Attitudes of europeans towards air quality, 2019.
- [12] Stuart K Grange and David C Carslaw. Using meteorological normalisation to detect interventions in air quality time series. *Science of The Total Environment*, 653:578–588, 2019.
- [13] Stuart K Grange, David C Carslaw, Alastair C Lewis, Eirini Boleti, and Christoph Hueglin. Random forest meteorological normalisation models for swiss pm 10 trend analysis. *Atmospheric Chemistry and Physics*, 18(9):6223–6239, 2018.
- [14] RF Gunst and JT Webster. Regression analysis and problems of multicollinearity. *Communications in Statistics-Theory and Methods*, 4(3):277–292, 1975.
- [15] INEMAR - ARPA Lombardia. Inventario emissioni in atmosfera: emissioni in regione lombardia nell’anno 2017 – revisione pubblica. *INEMAR*, 2020.
- [16] ISPRA. Italian emission inventory 2020, 2020.

- [17] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [18] Sarah J Jones. If electric cars are the answer, what was the question? *British medical bulletin*, 129(1):13–23, 2019.
- [19] Louise W Kao and Kristine A Nañagas. Carbon monoxide poisoning. *Medical Clinics*, 89(6):1161–1194, 2005.
- [20] Claudia Libiseller, Anders Grimvall, Jari Waldén, and Helena Saari. Meteorological normalisation and non-parametric smoothing for quality assessment and trend analysis of tropospheric ozone data. *Environmental monitoring and assessment*, 100(1-3):33–52, 2005.
- [21] ARPA Lombardia. Dati sensori aria, 2012-2020. <https://www.dati.lombardia.it/Ambiente/Dati-sensori-aria/nicp-bhqi>.
- [22] ARPA Lombardia. Dati sensori meteo, 2012-2020. <https://www.dati.lombardia.it/Ambiente/Dati-sensori-meteo/647i-nhvk>.
- [23] ARPA Lombardia. Rapporto sulla qualità dell’aria della città metropolitana di milano, 2018.
- [24] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans et al. Benzene. 2018.
- [25] World Health Organization et al. Who air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Technical report, World Health Organization, 2006.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [27] P Steven Porter, S Trivikrama Rao, Igor G Zurbenko, Alan M Dunker, and George T Wolff. Ozone air quality over north america: part ii—an analysis of trend detection and attribution techniques. *Journal of the Air & Waste Management Association*, 51(2):283–306, 2001.
- [28] Joaquin Quiñonero-Candela, Masashi Sugiyama, Neil D Lawrence, and Anton Schwaighofer. *Dataset shift in machine learning*. Mit Press, 2009.
- [29] S Trivikrama Rao and Igor G Zurbenko. Detecting and tracking changes in ozone air quality. *Air & waste*, 44(9):1089–1092, 1994.
- [30] Mohsin Raza, Longfei Chen, Felix Leach, and Shiting Ding. A review of particulate number (pn) emissions from gasoline direct injection (gdi) engines and their control techniques. *Energies*, 11(6):1417, Jun 2018.
- [31] Stefano Scolari. Evoluzione della qualità dell’aria a milano dagli anni cinquanta ad oggi. 2017.
- [32] Mordecai Segal, Richard T McNider, Roger A Pielke, and DS McDougal. A numerical model simulation of the regional air pollution meteorology of the greater chesapeake bay area—summer day case study. *Atmospheric Environment (1967)*, 16(6):1381–1397, 1982.
- [33] John H Seinfeld. Ozone air quality models. *JAPCA*, 38(5):616–645, 1988.
- [34] Anders Skonhoft, Bjart Holtmark, et al. The norwegian support and subsidy of electric cars. should it be adopted by other countries? Technical report, 2014.
- [35] Roland B Stull. *An introduction to boundary layer meteorology*, volume 13. Springer Science & Business Media, 2012.

- [36] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2012. https://dati.comune.milano.it/dataset/ds779_area_c_accessi_giornalieri_2012.
- [37] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2013. https://dati.comune.milano.it/dataset/ds780_area_c_accessi_giornalieri_2013.
- [38] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2014. https://dati.comune.milano.it/dataset/ds781_area_c_accessi_giornalieri_2014.
- [39] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2015. https://dati.comune.milano.it/dataset/ds782_area_c_accessi_giornalieri_2015.
- [40] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2016. https://dati.comune.milano.it/dataset/ds783_area_c_accessi_giornalieri_2016.
- [41] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2017. https://dati.comune.milano.it/dataset/ds873_area_c_accessi_mensili_gennaio_2017.
- [42] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2018. https://dati.comune.milano.it/dataset/ds885_area_c_accessi_mensili_gennaio_2018.
- [43] AMAT Agenzia Mobilità Ambiente Territorio. Accessi giornalieri area c, 2019-2020. https://dati.comune.milano.it/dataset/ds1072_area-c-accessi-giornalieri-20192020.
- [44] A. Trentini and M. Belotti. Normalizzazione meteorologica dei dati dell'inquinamento atmosferico. <https://github.com/MarcoBelotti96/Tesi>, 2020.
- [45] Andrea Trentini. Lombardy epa obtorto collo data and anti-pollution policies fallacies. *Journal of e-Learning and Knowledge Society*, 10(2), 2014.

Ringraziamenti

Vorrei ringraziare tutte le persone che mi hanno aiutato nella realizzazione di questo elaborato. Il Prof. Andrea Trentini, relatore di questa tesi, per avermi permesso di studiare questo tema e per il supporto continuamente offerto in tutti questi mesi di lavoro. Ringrazio inoltre ARPA Lombardia per aver messo a libera disposizione tutti i dati utilizzati nei nostri studi, senza i quali nulla di ciò che si è realizzato sarebbe stato possibile.

Un ringraziamento speciale va alla mia famiglia, a Paolo, Fiorella ed Ivan, che in tutti questi anni mi hanno sempre supportato senza nessuna pretesa, anche quando avrebbero avuto tutte le ragioni per averne. A Vanessa, per avermi ridato speranza e felicità in un momento in cui ne ero rimasto senza. Ai ciucù, compagni di mille avventure da una vita. Ed a tutti gli altri amici e colleghi con cui, in tutti questi anni, ho condiviso esperienze ed emozioni.

Infine un grazie ai Laghi Gemelli ed al rifugio, la scuola e l'università mi hanno fatto crescere tanto, lì sono diventato un uomo ed ho imparato che per essere veramente felici basta poco.

Grazie infinite

Marco