CC BY

La valutazione dell'esposizione della popolazione italiana all'inquinamento atmosferico nel periodo 2016-2019 per lo studio della relazione tra inquinamento atmosferico e COVID-19

Exposure assessment of air pollution in Italy 2016-2019 for future studies on air pollution and COVID-19

Massimo Stafoggia,[1] Giorgio Cattani,[2] Carla Ancona,[1] Andrea Ranzi[3]

[1] Dipartimento di epidemiologia del Servizio sanitario regionale del Lazio, ASL Roma 1, Roma
[2] Dipartimento per la valutazione, i controlli e la sostenibilità ambientale, Istituto superiore per la protezione e la ricerca ambientale, Roma
[3] Centro tematico regionale ambiente prevenzione e salute, Agenzia regionale per la prevenzione, l'ambiente e l'energia dell'Emilia-Romagna, Modena

Corrispondenza: Massimo Stafoggia; m.stafoggia@deplazio.it

RIASSUNTO

L'inquinamento atmosferico è una delle principali cause di morte in tutto il mondo, con effetti avversi legati a esposizioni sia a breve sia a lungo termine. Di recente è stato anche accostato alla pandemia COVID-19. Per analizzare questa possibile associazione a livello italiano, è necessario indagare tutta l'area della penisola, sia le zone urbane sia quelle non urbane. Si rileva quindi la necessità di uno strumento per la valutazione dell'esposizione, omogeneo e applicabile su tutto il territorio nazionale.

Esistono già esperienze in Italia di modelli ad alta risoluzione spazio-temporale, per la stima del materiale particolato (PM), con l'uso di predittori spazio-temporali, dati satellitari, dati di monitoraggio della qualità dell'aria.

Il presente lavoro completa la disponibilità di queste stime per gli anni più recenti (2016-2019) e viene applicato anche per stimare gli ossidi di azoto e l'ozono. La risoluzione spaziale è di 1x1 km.

Il modello si conferma capace di catturare la variabilità del PM (R^2=0,78 e 0,74 per PM_{10} e $PM_{2,5}$, rispettivamente) e fornisce stime attendibili anche per l'ozono (R^2=0,76); per l'NO_2 le *performance* del modello sono inferiori (R^2=0,57). Le stime modellistiche sono state utilizzate per calcolare la PWE (*population-weighted exposure*) come media annua pesata sulla popolazione residente in ogni singola cella, che rappresenta la stima dell'esposizione cronica all'inquinamento atmosferico della popolazione italiana.

Queste stime sono pronte per essere utilizzate negli studi sull'associazione tra esposizione cronica all'inquinamento atmosferico e patologia COVID-19, così come per indagini sul ruolo dell'inquinamento dell'aria sulla salute della popolazione italiana.

Parole chiave: inquinamento atmosferico, modelli random-forest, *population-weighted exposure*, COVID-19

COSA SI SAPEVA GIÀ

- **Esiste una variabilità spazio-temporale delle concentrazioni dei principali inquinanti aerodispersi in ambito urbano e sub-urbano.**
- **L'utilizzo della modellistica permette di caratterizzare meglio l'esposizione della popolazione residente.**
- **Una mappatura delle concentrazioni di $PM_{2,5}$ e PM_{10} a scala 1x1 km è disponibile per gli anni 2013-2015 su tutto il territorio nazionale.**

COSA SI AGGIUNGE DI NUOVO

- **Sono state stimate le concentrazioni per 4 inquinanti (PM_{10}, $PM_{2,5}$, O_3, NO_2) a scala 1x1 km sul territorio nazionale, relativamente agli anni 2016-2019.**
- **La *population-weighted exposure* calcolata su media annua è in linea con le misurazioni delle stazioni di monitoraggio, seppure con differenze tra i diversi inquinanti.**
- **Il dato è pronto per un utilizzo in indagini epidemiologiche di associazione fra esposizione cronica a inquinamento atmosferico e patologia COVID-19.**

ABSTRACT

Air pollution is one of the leading causes of death worldwide, with adverse effects related both to short-term and long-term exposure. It has also recently been linked to COVID-19 pandemic. To analyze this possible association in Italy, studies on the entire area of the peninsula are necessary, both urban and non-urban areas. Therefore, there is a need for a homogeneous and applicable exposure assessment tool throughout the country.

Experiences of high spatio-temporal resolution models for Italian territory already exist for PM estimation, using space-time predictors, satellite data, air quality monitoring data.

This work completes the availability of these estimations for the most recent years (2016-2019) and is also applied to nitrogen oxides and ozone. The spatial resolution is 1x1 km.

The model confirms its capability of capturing most of PM variability (R^2=0.78 and 0.74 for PM_{10} e $PM_{2.5}$, respectively), and provides reliable estimates also for ozone (R^2=0.76); for NO_2 the model performance is lower (R^2=0.57). The model estimations were used to calculate the PWE (population-weighted exposure) as the annual mean, weighted on the resident population in each individual cell, which represents the estimation of the Italian population's chronic exposure to air pollution.

These estimates are ready to be used in studies on the association between chronic exposure to air pollution and COVID-19 pathology, as well as for investigations on the role of air pollution on the health of the Italian population.

Keywords: air pollution, random-forest models, population-weighted exposure, COVID-19

INTRODUZIONE

Le stime del *Global burden of disease* indicano che l'inquinamento atmosferico, subito dopo dieta, fumo, ipertensione e diabete causa ogni anno 2,9 milioni di morti premature in tutto il mondo.[1] Cardiopatia ischemica, ictus, malattia polmonare ostruttiva cronica, sono le principali condizioni associate alla mortalità correlata all'inquinamento atmosferico.

Nel nostro Paese, la principale fonte di inquinamento atmosferico nelle realtà urbane è rappresentata dagli scarichi del traffico veicolare, in particolare dei motori diesel; tuttavia, è stato anche riconosciuto negli ultimi anni il ruolo importante e crescente della combustione delle biomasse (quali legno e pellet usati per il riscaldamento), responsabile di un aumento della concentrazione di polveri soprattutto nelle aree del Nord e nelle aree rurali. A queste fonti inquinanti si aggiungono l'agricoltura e gli allevamenti di bestiame, per le emissioni di ammoniaca, che per reazione chimica nell'atmosfera si trasforma in particolato secondario, e le emissioni industriali.

In questi ultimi mesi, la possibile relazione tra i livelli di inquinamento atmosferico e l'epidemia di COVID-19 (malattia del coronavirus causata dalla SARS-CoV-2) è stata oggetto di dibattito. È noto che i principali inquinanti atmosferici antropogenici hanno proprietà fortemente ossidanti e che l'esposizione a questi inquinanti inneschi reazioni infiammatorie polmonari e sistemiche acute e croniche, anche in soggetti giovani e sani.[2] L'esposizione al particolato (PM_{10}, $PM_{2,5}$) o al biossido di azoto (NO_2) può, dunque, rendere il sistema respiratorio più suscettibile alla infezione da SARS-CoV-2 e alle complicanze della malattia da coronavirus. Le persone che vivono in un'area con livelli elevati di inquinanti atmosferici sono infatti a maggior rischio di infezioni respiratorie.[3]

L'ipotesi di un possibile collegamento tra la diffusione del COVID-19 e l'inquinamento atmosferico è stata indagata in diversi studi che hanno utilizzato indici di qualità dell'aria come *proxy* dell'esposizione della popolazione[4] oppure le concentrazioni degli inquinanti a livelli di aggregazione molto elevati.[5,6] Questi studi, di tipo ecologico, non consentono una interpretazione causale della relazione tra esposizione ed esito. Per valutare gli effetti dell'inquinamento atmosferico sulla diffusione di COVID-19 e sulla prognosi della malattia è prima di tutto necessaria una adeguata valutazione dell'esposizione della popolazione in studio, in quanto gli studi epidemiologici, possibilmente di tipo analitico, devono tenere conto della forza delle prove esistenti sulle relazioni ambiente-salute, delle caratteristiche peculiari dei casi COVID-19, di un razionale chiaro e condiviso sul potenziale legame dell'inquinamento atmosferico con COVID-19.[7]

La valutazione dell'esposizione è quindi il primo passo da affrontare; è necessaria una stima della variabilità spaziale dell'esposizione della popolazione italiana all'inquinamento atmosferico, sia a livello urbano sia in zone extraurbane, queste ultime solitamente meno indagate negli studi epidemiologici. Le informazioni devono essere omogenee a livello spaziale (cioè ottenute sulla base di dati o modelli standardizzati sull'intero territorio di indagine) e riferite al periodo di studio della pandemia (per la valutazione degli effetti acuti), ma ancora di più (vista la maggiore plausibilità di un possibile ruolo) su un arco temporale retrospettivo più ampio, per esempio gli anni precedenti (2016-2019), per valutare gli effetti cronici.

In Italia, le concentrazioni dei principali inquinanti aerodispersi sono regolarmente misurate dalla rete di monitoraggio di qualità dell'aria, gestita dal Sistema nazionale di protezione ambientale (SNPA) attraverso le singole Agenzie regionali e provinciali e la struttura centrale di ISPRA.[8] Questo dato rappresenta il patrimonio principale di misure sul territorio nazionale ed è regolato dal DLgs 155/2010, che definisce i criteri di zonizzazione e di progettazione della rete di monitoraggio della qualità dell'aria, in considerazione di diversi parametri tra cui la popolazione residente. Nonostante la rete di monitoraggio sia oggi costituita da un numero di stazioni e punti di misura per inquinante maggiore del minimo previsto dalla normativa e l'informazione puntuale sia caratterizzata da un alto livello qualitativo, l'informazione relativa alla variabilità spaziale dei livelli di esposizione della popolazione che può essere da essa desunta è ancora limitata.

Negli ultimi anni, sono state realizzate stime della variabilità spaziale della concentrazione di diversi inquinanti aerodispersi a scala locale e regionale basate su modelli empirici;[9,10] più di recente sono stati sviluppati modelli ad alta risoluzione spaziale e temporale per la stima delle concentrazioni degli inquinanti atmosferici su scala nazionale e continentale utilizzando, tra le variabili esplicative, dati di osservazione della terra dallo spazio, come, per esempio, lo spessore ottico dell'aerosol (*Aerosol Optical Depth* -AOD).[11,12]

Anche per il territorio italiano si sono sviluppate, negli ultimi anni, esperienze per la stima del particolato ad alta risoluzione spazio-temporale.[13,14] Di recente sono stati pubblicati i risultati delle stime ottenute con un modello di *machine learning*, il *random forest,* per il periodo 2013-2015 per PM_{10} e $PM_{2,5}$.[15] Il presente contributo riporta l'estensione dello stesso modello agli anni più recenti (2016-2019), includendo i principali inquinanti di interesse sanitario (PM_{10}, $PM_{2,5}$, NO_2, O_3).

Il lavoro presenta anche una valutazione della *population-weighted exposure* (PWE), ovvero la stima dell'esposizione di tutta la popolazione residente in Italia. Vengono riportate le stime di esposizione di lungo periodo (media annua) e il loro confronto con i dati di monitoraggio delle diverse tipologie di stazioni fisse.

Il presente contributo intende fornire uno strumento per la valutazione dell'esposizione media annua della popolazione residente, omogeneo su tutto il territorio nazionale, che possa essere utilizzato per studi di epidemiologia ambientale a scala nazionale o locale sulla relazione fra inquinamento e patologie avverse, per esempio la pandemia COVID-19.

METODI

DOMINIO DI STUDIO

Il dominio di riferimento di questo progetto è il territorio italiano. La superficie nazionale si estende tra le latitudini 35°N e 47°N e le longitudini 6°E e 19°E e ha una superficie totale di 301.340 km^2. In questo progetto è stata costruita una griglia di celle di ampiezza pari a 1 km^2 tale da contenere l'intera superficie italiana. Il totale delle celle considerate è 307.635.

PREDITTORI

Al fine di sviluppare un modello di predizione delle concentrazioni giornaliere di inquinanti atmosferici, si è resa necessaria la raccolta di un esteso *set* di parametri spaziali (variabili di cella in cella, ma fissi nel tempo) e spaziotemporali (variabili di cella in cella e di giorno in giorno), capaci di cogliere la variabilità spaziale (i primi) e temporale (i secondi) degli inquinanti in studio. Per un dettaglio dei singoli predittori si rimanda all'articolo di Stafoggia et al.[15] In sintesi:

- I predittori spaziali includono: aree amministrative (Regioni, Province, Comuni); popolazione residente per sezione di censimento al 2011 (Istat); aree geoclimatiche (Ispra); variabili di copertura e di uso del territorio (Corine Land Cover 2012, EEA-Ispra); emissioni industriali di CO, PM_{10} e NOx (2010, Ispra); densità di superfici impervie (Copernicus monitoring land service, CMLS), altitudine (Digital elevation model over europe, CMLS); luci notturne (da satellite Suomi - National polar-orbiting partnership), sensore Visible infrared imaging radiometer Suite (VIIRS, NOAA-NASA); densità stradale per cella e prossimità del centroide di cella dalla strada più vicina, per tipologia (autostrade, strade maggiori, minori, locali, TeleAtlas TomTom);
- I predittori spaziotemporali includono: stime modellistiche di PM_{10}, $PM_{2,5}$, NO_2 e O_3 da modelli di dispersione europei messi a disposizione dalla rete Copernicus (CAMS); stime di AOD a diverse lunghezze d'onda, (MAIAC, NASA e *Copernicus atmosphere monitoring service*, CAMS); stime giornaliere di variabili meteorologiche (temperatura dell'aria, *dew-point*, altezza dello strato limite planetario, umidità, pressione barometrica, velocità e direzione del vento, copertura nuvolosa, precipitazioni) ottenute dalla ri-analisi atmosferica globale prodotta dal Centro europeo per le previsioni meteorologiche a medio termine (ERA-5, ECMWF); indici mensili di vegetazione (*Normalized difference vegetation index* – NDVI, da MODIS-TERRA, NASA), e Leaf area index – LAI, *Copernicus global land service*, CGLS).

CONCENTRAZIONI MISURATE DI INQUINANTI AERODISPERSI

Sono state utilizzate le concentrazioni medie giornaliere e orarie di PM_{10}, $PM_{2,5}$, NO_2 ed O_3 misurate dal 01.01.2016 al 31.12.2019 presso le stazioni delle reti di monitoraggio regionali gestite dalle agenzie del SNPA (pubblicamente disponibili nell'ambito dello scambio di informazioni ex Decisione 2011/850/EU, https://www.eea.europa.eu/data-and-maps/data/aqereporting-8). Tutti i dati, originariamente con risoluzione oraria, sono stati aggregati come medie giornaliere. Occorre sottolineare che, mentre i dati fino al 31.12.2018 sono stati già validati dal SNPA e inviati ufficialmente alla *European environmental agency*, la raccolta e l'invio dei dati del 2019 sono ancora in corso, dunque sono stati utilizzati i dati già disponibili al 30 giugno, relativi a circa l'80% dell'intero *network* nazionale. Per questa ragione, si è preferito, in sede di presentazione di risultati, utilizzare l'anno 2018 come anno di riferimento.

MODELLO

Il modello di predizione consiste di due fasi successive, come descritto in Stafoggia et al. (2019). Per il solo $PM_{2,5}$, l'obiettivo dello *stage* 1 è stato quello di espandere il numero di stazioni e punti di misura del $PM_{2,5}$ sfruttando la correlazione tra $PM_{2,5}$ e PM_{10} negli stessi punti di misura. In particolare, è stato sviluppato un modello *random forest* in cui i dati osservati di $PM_{2,5}$ erano la variabile risposta, i dati colocati di PM_{10} erano il predittore principale, e le coordinate del monitor e le variabili di mese e giorno erano predittori addizionali. Quindi, tale modello è steso usato per predire le concentrazioni attese di $PM_{2,5}$ in tutti i siti che misuravano solo il PM_{10} (*stage* 1). In seguito, separatamente per ogni anno (2016-2019) e per ognuno dei 4 inquinanti ($PM_{2,5}$ come predetto nello *stage* 1, PM_{10}, NO_2 e O_3) si è addestrato un modello *random forest* avente come variabile risposta la concentrazione giornaliera dell'inquinante, e come variabili predittive tutti i predittori spaziali e temporali precedentemente elencati. A partire da un modello completo (con tutti i predittori) si è passati a un modello finale più parsimonioso, dove sono stati inclusi solo i predittori con potere predittivo non trascurabile (coefficiente di importanza >1%). Tutte le fasi di sviluppo del modello predittivo e di valutazione della sua *performance* sono state condotte applicando tecniche di cross-validazione (CV) *by monitor*, cioè suddividendo i punti di misura in 10 gruppi casuali, addestrando il modello su 9 gruppi (*training*) e predicendolo sul 10° (*testing*), per poi iterare la stessa procedura su tutti i 10 gruppi in sequenza. In questo modo, è stato possibile simulare la capacità del modello predittivo di stimare l'intera serie giornaliera di inquinante su un recettore esterno (cioè su un punto che non aveva contribuito al modello stesso). Infine, il modello è stato predetto su tutte le celle del dominio italiano, e tutti i giorni dell'anno considerato.

POPULATION-WEIGHTED EXPOSURE

A partire dalle concentrazioni giornaliere stimate dal modello *random forest*, l'esposizione media annuale della popolazione italiana è stata calcolata pesando le concentrazioni di ciascun inquinante a livello di singola cella per la popolazione residente all'interno della cella stessa, al fine di ottenere una stima della *population-weighted exposure* (PWE). In particolare, si sono seguiti i seguenti pas-

si. In primo luogo, la popolazione per cella è stata calcolata sovrapponendo lo *shapefile* delle sezioni di censimento 2011 allo *shapefile* del grigliato di definizione del modello *random forest*; in questo passaggio, si è assunta una omogeneità geografica della distribuzione della popolazione all'interno di ogni sezione censuaria. Quindi, le sottopopolazioni delle intersezioni delle sezioni con la cella target sono state sommate per avere un valore di popolazione per cella. In secondo luogo, l'esposizione delle singole celle ai diversi inquinanti atmosferici è stata pesata sulla base della popolazione residente, con pesi direttamente proporzionali a quest'ultima. Infine, le esposizioni pesate per cella sono state mediate a livello di aree amministrative (Comuni, Province), in modo tale da attribuire più importanza alle celle con maggiore popolazione e dare minore peso alle aree disabitate o scarsamente popolate.
In questo lavoro sono riportati i valori di media annua della PWE, calcolati a partire dalle stime giornaliere del modello *random-forest*. Tali stime di PWE sono state poi confrontate con le medie annue derivate dai dati delle stazioni di monitoraggio di qualità dell'aria. Sono state considerati i punti di misura che avessero dati giornalieri validi per almeno il 75% dei giorni all'interno di un anno solare. Le medie sono state calcolate per tutti i punti di misura, collocati nelle stazioni da traffico e quelle di *background* (urbano, suburbano e rurale).

RISULTATI

La tabella 1 riporta una sintesi della *performance* dei modelli *random forest* per anno, inquinante e tipologia di stazione. In particolare, la tabella riporta il numero di siti e il coefficiente R^2 (frazione di variabilità spiegata) sui monitor di *testing* a seguito di procedura di CV *by monitor*, come descritto in precedenza. Si osserva una elevata capacità dei modelli di predire adeguatamente le concentrazioni di $PM_{2,5}$, PM_{10} e O_3 nei punti di misura esterni, mentre la *performance* del modello dell'NO_2 è inferiore. Non si apprezzano significative differenze per tipologie di monitor, a dimostrazione del fatto che vengono adeguatamente descritte (e predette) le concentrazioni in situazioni di traffico o di *background*.
In generale, i predittori più importanti per la stima del PM sono risultati termini spaziotemporali quali la stima della concentrazione al suolo di PM da modello CAMS, la temperatura, il giorno dell'anno e l'altezza dello strato limite. Per l'NO_2, invece, i parametri più discriminanti sono stati alcuni predittori spaziali, quali la popolazione residente, la densità di strade maggiori e minori, e l'illuminazione notturna. Infine, per l'ozono i predittori più importanti sono risultati la temperatura dell'aria, il giorno ed il mese, l'altezza dello strato limite planetario e l'altitudine (dati non presentati).
La figura 1 presenta le predizioni medie annuali (PM e NO_2) e le predizioni medie della stagione calda (definita come maggio-ottobre, O_3) a livello di cella di 1 km² per l'anno 2018. Si notano concentrazioni di PM più elevate in pianura padana e nei maggiori centri abitati, concentrazioni di NO_2 più elevate nelle maggiori metropoli del nord e in corrispondenza delle principali arterie stradali. I livelli di ozono, come atteso in base ai complessi meccanismi di formazione di questo inquinante, sono più alti in quota, sui rilievi appenninici e alpini, e mostrano generalmente un gradiente positivo dall'interno verso l'esterno dei centri abitati.
La tabella 2 riporta i valori di PWE per l'intera popolazione italiana, suddivisi per macro-aree geografiche. Anche le stime modellistiche riportano il 2017 come anno a maggiore esposizione ad inquinamento atmosferico del quadriennio.[8] La PWE calcolata per le 10 città maggiormente popolate (con più di 300.000 abitanti), evidenzia come, mediamente, la popolazione di queste città sia esposta a livelli di biossido di azoto maggiori rispetto al resto della popolazione residente in aree diverse.
Si apprezzano comunque importanti differenze tra le diverse aree urbane. Per esempio, emerge una significativa differenza di esposizione della popolazione di Milano ri-

ANNO	TIPOLOGIA STAZIONE	$PM_{2,5}$		PM_{10}		NO_2		O_3	
		SITI (n.)	CV-R^2	SITI (n.)	CV-R^2	SITI (n.)	CV-R^2	SITI (n.)	CV-R^2
2016	totale	505	0,80	494	0,76	565	0,54	292	0,77
	traffico	158	0,79	157	0,75	174	0,43	9	0,54
	background	290	0,80	280	0,77	326	0,57	272	0,77
2017	totale	510	0,81	496	0,76	579	0,55	302	0,76
	traffico	158	0,81	156	0,76	171	0,46	7	0,60
	background	291	0,82	280	0,77	337	0,59	283	0,77
2018	totale	572	0,75	562	0,69	625	0,54	348	0,72
	traffico	164	0,75	163	0,70	175	0,41	11	0,55
	background	334	0,76	322	0,69	364	0,60	307	0,73
2019	totale	413	0,77	404	0,73	435	0,63	251	0,80
	traffico	115	0,80	114	0,80	119	0,51	12	0,50
	background	257	0,80	249	0,76	272	0,68	224	0,82

Tabella 1. Statistiche di adattamento dei modelli random forest, per inquinante, anno e tipologia di stazione.
Table 1. Statistics of adaptation of random forest models, by pollutant, year, and station type.

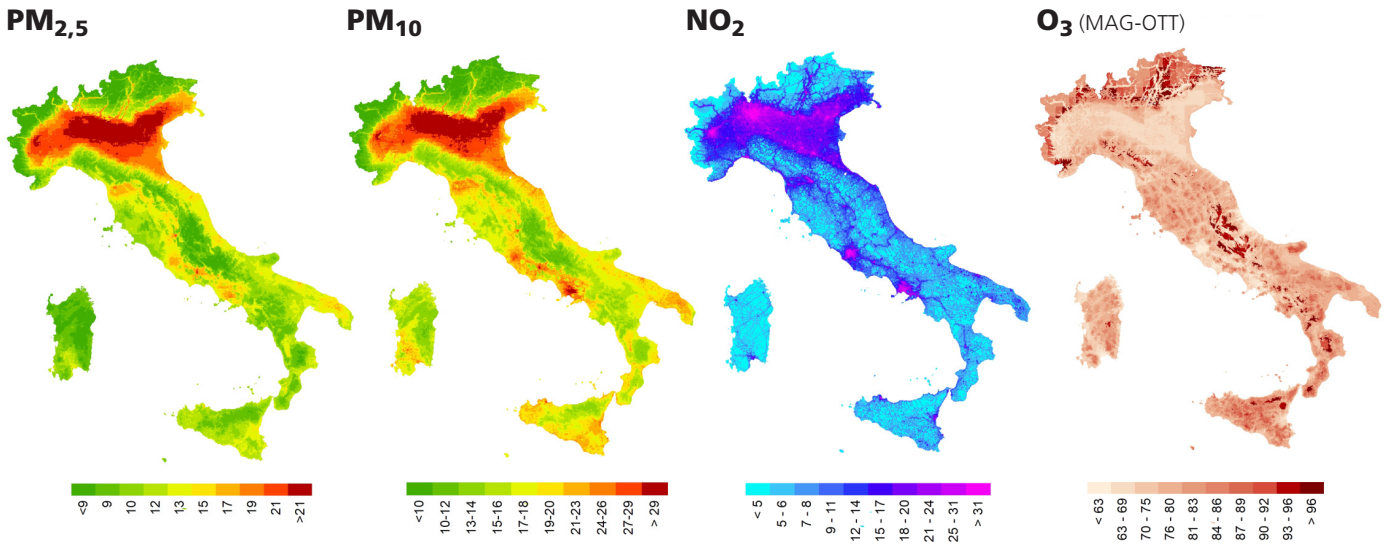

Figura 1. Concentrazioni medie annue (PM_{10}, $PM_{2,5}$ e NO_2, µg/m³) e dei mesi maggio-ottobre (O_3, µg/m³), anno 2018.
Figure 1. Annual mean concentrations (PM_{10}, $PM_{2,5}$, and NO_2, µg/m³) and mean of the months May-October (O_3, µg/m³), year 2018.

	POPOLAZIONE (CENS. 2011)	2016				2017				2018				2019			
		PM_{10}	$PM_{2,5}$	NO_2	O_3#	PM_{10}	$PM_{2,5}$	NO_2	O_3#	PM_{10}	$PM_{2,5}$	NO_2	O_3#	PM_{10}	$PM_{2,5}$	NO_2	O_3#
ITALIA																	
Totale	59.433.744	25,2	16,7	22,9	67,0	26,3	17,6	22,8	72,2	24,3	16,0	20,8	69,4	23,4	16,0	20,3	67,4
Nord	27.214.372	27,8	20,4	26,9	66,9	31,3	22,7	27,5	71,6	26,9	19,4	24,9	70,0	25,9	18,8	24,1	67,1
Centro	11.600.675	23,2	14,9	22,1	64,1	22,3	14,3	21,9	69,6	22,2	14,1	19,9	67,2	21,6	14,8	19,2	65,5
Sud e Isole	20.619.697	22,7	12,9	18,0	68,6	22,0	12,8	17,2	74,3	22,1	12,5	15,8	69,9	21,0	13,0	15,9	68,9
ZONE CLIMATICHE*																	
Area alpina	3.018.258	19,0	13,5	17,6	66,1	20,8	15,8	18,2	73,3	18,1	14,1	16,3	68,8	16,7	14,5	15,5	69,7
Valle Padana	17.249.378	30,2	21,6	30,0	66,7	34,7	25,1	30,7	71,6	29,4	21,1	27,9	70,2	27,9	20,1	27,1	68,0
Alto Adriatico	5.521.578	28,0	19,2	27,8	64,8	29,8	21,4	23,6	69,2	26,5	19,0	21,4	69,2	27,1	18,5	20,7	63,9
Appennino	4.690.575	18,4	12,8	19,7	69,1	17,8	12,0	12,6	78,4	18,3	12,1	11,1	73,2	17,4	13,0	10,4	71,7
Alto Tirreno	4.925.079	21,4	14,2	21,8	70,5	21,4	14,1	22,2	72,3	20,9	13,5	19,7	71,9	20,4	14,5	18,8	65,2
Basso Tirreno	10.394.371	26,3	16,8	26,7	60,3	25,4	14,5	25,5	66,5	25,0	14,2	24,2	63,6	23,2	14,5	24,1	63,3
Basso Adriatico	6.270.880	22,2	13,1	22,2	70,5	21,5	13,4	15,7	76,3	21,2	13,1	14,1	71,2	20,8	13,5	14,6	72,2
Sud e Sicilia	5.724.263	21,7	11,5	22,5	72,1	20,5	11,9	15,6	77,4	21,7	11,9	14,0	72,2	21,3	12,7	13,6	69,3
Sardegna	1.639.362	21,0	9,3	22,0	71,8	20,7	11,5	11,8	68,8	19,6	10,0	10,5	65,1	19,7	10,6	9,8	65,0
CITTÀ >300.000 ABITANTI																	
Totale	8.111.244	28,5	18,3	37,1	63,3	29,8	18,8	37,1	67,5	27,3	16,8	33,3	64,2	25,9	16,7	32,2	61,2
Roma	2.617.175	27,7	16,8	35,1	56,2	26,5	16,0	35,3	60,7	25,7	15,4	32,2	57,4	25,2	15,6	31,2	57,3
Milano	1.242.123	35,7	25,8	50,3	63,6	41,0	28,4	49,3	69,9	34,3	23,6	43,7	69,3	32,3	21,2	43,0	66,1
Napoli	962.003	29,3	16,4	35,6	61,2	28,9	14,9	33,7	70,8	27,8	13,9	32,1	66,7	24,7	13,7	33,5	61,2

media per il periodo maggio-ottobre / *mean value for the period May-October* * Zone climatiche secondo la classificazione ISPRA / *Climatic areas according to ISPRA classification*

Tabella 2. Population-weighted exposure (PWE) per aree geografiche per PM_{10}, $PM_{2,5}$, NO_2 e O_3 (µg/m³). Anni 2016-2019.
Table 2. Population-weighted exposure (PWE) by geographic areas for PM_{10}, $PM_{2,5}$, NO_2, and O_3 (µg/m³). Years 2016-2019.

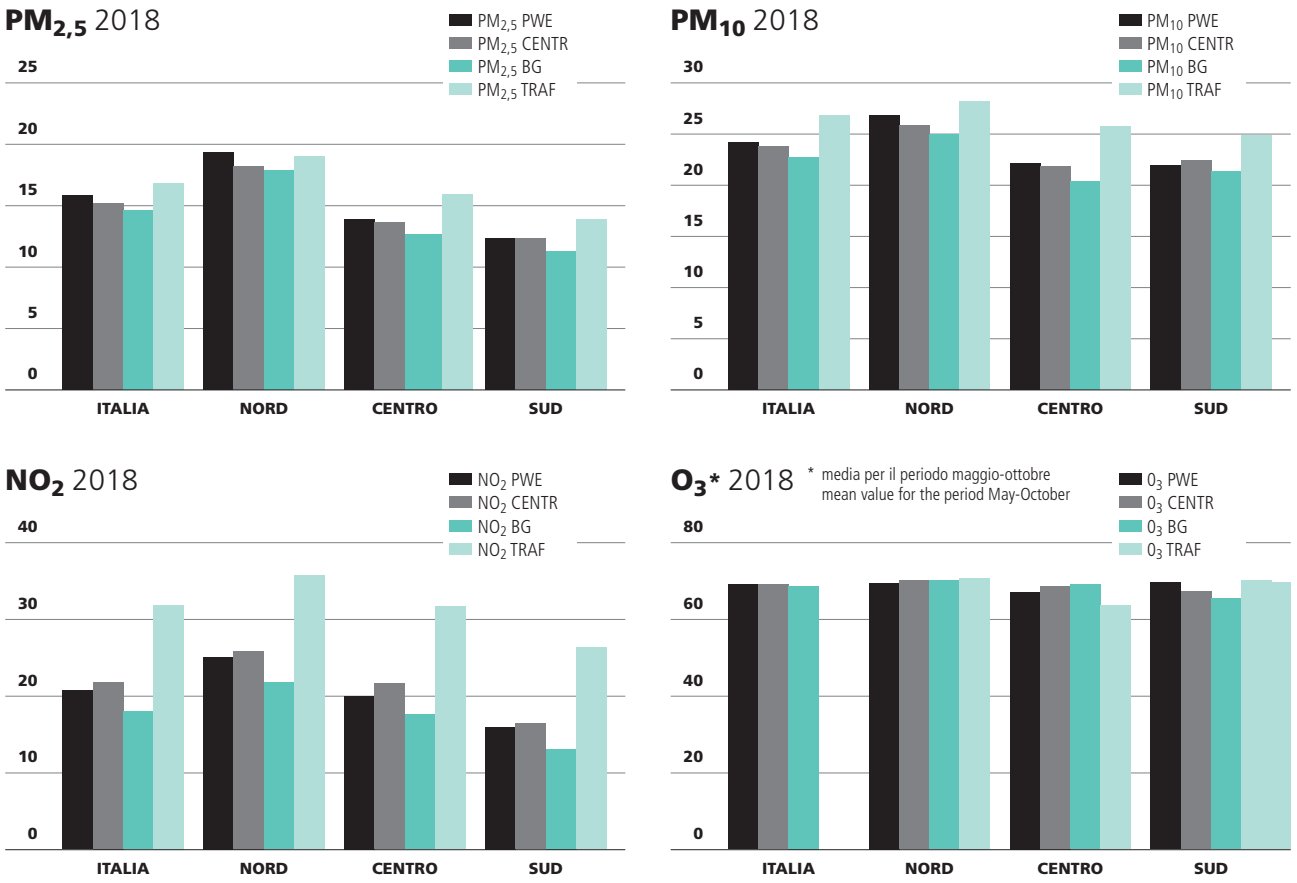

Figura 2. Confronto valori medi annui di *population-weighted exposure* (PWE) (µg/m³) con i dati delle stazioni di monitoraggio qualità dell'aria (tutte, background, traffico). Anno 2018.
Figure 2. Comparison of the average annual mean values of population-weighted exposure (PWE) (µg/m³) and air quality monitoring stations data (all monitors, background, traffic). Year 2018.

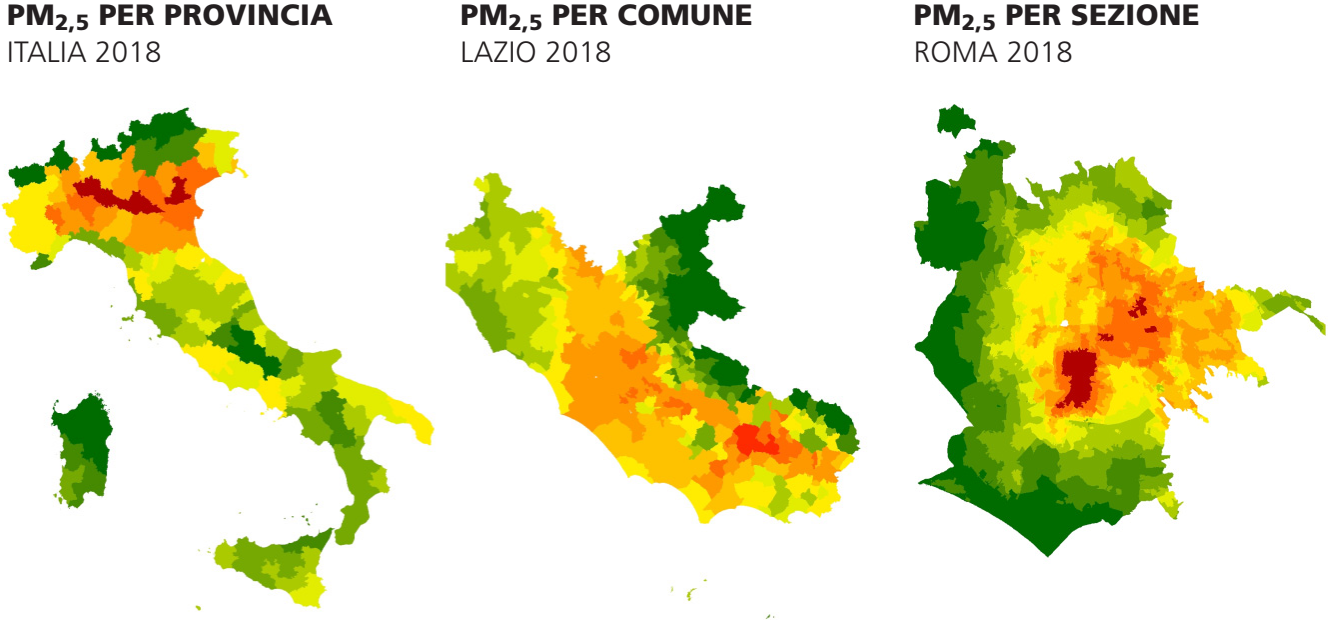

Figura 3. Concentrazioni medie annue di $PM_{2,5}$ (µg/m³) per aree amministrative. Anno 2018
Figure 3. $PM_{2.5}$ (µg/m³) annual mean concentrations, by administrative areas. Year 2018.

spetto a quella di Roma o Napoli. Tali differenze sono dovute non tanto al carico emissivo complessivo, quanto al contesto orografico e climatico completamente diverso e sfavorevole per Milano e tutta l'area del bacino padano.
Per l'anno 2018, l'ultimo per il quale siano già disponibili tutti i dati delle stazioni di monitoraggio della qualità dell'aria, licenziati dopo le diverse fasi di validazione, è stato effettuato un confronto fra le esposizioni medie annue calcolate attraverso la PWE e le medie annue delle serie storiche dei dati dei monitoraggi.
La figura 2 riporta graficamente il confronto dei livelli per l'intera penisola e per Nord, Centro e Sud Italia. I valori medi di PWE sembrano in linea con le medie di tutti i punti di misura disponibili, e si collocano in una zona intermedia tra i valori di stazione di *background* e quelli di traffico. Questo è particolarmente vero per il particolato, mentre per il biossido di azoto si notano andamenti più irregolari, per la nota maggiore variabilità di questo inquinante in ambito intraurbano. I valori di PWE rimangono comunque intermedi tra le medie dei punti di misura collocati nelle due tipologie di stazioni.
La figura 3 riporta l'aggregazione per Provincia, Comune e sezione di censimento dei valori di concentrazione stimata dal modello *random-forest*. Tali mappe rappresentano esempi di applicazioni di tali stime per la realizzazione di indagini epidemiologiche su scala nazionale, regionale o comunale, con dettaglio del contrasto di esposizione via via più elevato.

DISCUSSIONE

L'approccio epidemiologico agli studi degli effetti dell'inquinamento atmosferico outdoor distingue le esposizioni acute (a breve termine) da quelle croniche (a lungo termine). Nello studio sulle possibili associazioni fra inquinamento e COVID-19 si è maggiormente interessati agli effetti a lungo termine perché aumentano la dimensione della popolazione suscettibile, senza peraltro escludere un effetto a breve termine correlabile a un aumento della velocità di propagazione della epidemia oppure a un aggravamento delle condizioni cliniche nei soggetti già malati. Il modello sviluppato e presentato in questo lavoro può essere usato per caratterizzare entrambe le esposizioni per la popolazione residente su tutto il territorio nazionale.
Le stime dell'esposizione all'inquinamento atmosferico presentate rappresentano un importante sviluppo rispetto a quanto fatto in passato a livello italiano. Innanzitutto, sono stati aggiornati i modelli di predizione al 2019, con ottimi risultati di *fitting*, come desumibile dal confronto tra osservazioni e predizioni nei siti di "*testing*". In secondo luogo, sono stati applicati per la prima volta i modelli *random forest* all'NO_2 e all'O_3. Nel primo caso, i modelli non sono risultati ottimali, e ulteriori sforzi andranno fatti in futuro per migliorarne la *performance*, o attraverso la raccolta di ulteriori predittori spaziali o temporali, ovvero usando tecniche di *machine learning* alternative (quali l'*extreme gradient boosting,* le reti neurali, o una media pesata di diversi modelli). Al contrario, nel caso dell'ozono i modelli hanno mostrato una elevata affidabilità, descrivendo bene le concentrazioni osservate, soprattutto nei punti di misura di *background*.
Uno dei passaggi fondamentali nella validazione del modello *random-forest* è stato la cross-validazione delle stime modellistiche con le misure delle stazioni di monitoraggio di qualità dell'aria. In particolare, la procedura di suddividere le stazioni in un gruppo di *training* ed uno di *testing* ha lo scopo di valutare la *performance* del modello in punti esterni, simulandone il funzionamento in aree interamente prive di misura. È presumibile che le aree dove la densità di punti di misura è inferiore siano caratterizzate da una incertezza di stima maggiore rispetto alle altre; dunque, è ragionevole attendersi che le statistiche di *fitting* riportate siano una approssimazione per eccesso della reale incertezza dei modelli di predizione su scala nazionale.
Occorre, tuttavia, sottolineare che i criteri di ubicazione su macro-scala dei punti di misura determinano una maggiore densità di punti di misura nelle zone urbane e suburbane rispetto alle zone rurali, meno densamente popolate. Di conseguenza le aree con presumibile maggiore incertezza di stima del modello coincidono con le zone meno abitate. Di fatto, quindi, quando questi valori di concentrazione stimata vengono utilizzati per calcolare l'esposizione della popolazione residente, ai fini di indagare gli effetti dell'inquinamento sulla salute della popolazione, è presumibile che i dati maggiormente utilizzati siano quelli con maggiore grado di affidabilità. È il caso del calcolo della PWE, che rappresenta l'esposizione media di tutta la popolazione residente all'interno della zona considerata.
È stato effettuato il confronto tra il dato di esposizione media annua calcolato col metodo della PWE con le medie annue di concentrazione fornite dall'insieme dei punti di misura delle stazioni collocate nelle aree considerate. Sono state considerate dapprima tutte le stazioni, poi il confronto è stato fatto anche per tipo (di traffico e di fondo).
A livello di macro-aggregazioni, la PWE per ciascun inquinante sembra ben rappresentata dalla media di tutti i punti di misura, a dimostrazione di un criterio di collocazione e scelta delle stazioni, definito in base alla zonizzazione normativa vigente, che coglie in maniera soddisfacente l'esposizione media della popolazione. È probabile che a livello di maggior dettaglio ci possano essere delle differenze tra le due metodiche, derivanti anche dalla minore numerosità delle stazioni di confronto.
Il test effettuato rappresenta un risultato incoraggiante, da un lato per l'utilizzo delle medie delle stazioni nelle analisi su larga scala (come, per esempio, le valutazioni di impatto fatte su scala nazionale), dall'altro per l'utilizzo della PWE per indagini su aree sub-regionali.
Queste considerazioni sono applicabili al particolato, mentre le caratteristiche dell'NO_2, a partire dalla sua forte dipendenza dal traffico veicolare, rendono il confronto tra

PWE e punti di misura più variabile, specialmente quando si ragiona su aree non estese, dove la presenza e numerosità delle stazioni da traffico piuttosto che di fondo può spostare significativamente la media delle concentrazioni misurate. Una verifica sulle tre grandi città maggiormente popolate (Roma, Milano, Napoli) ha confermato queste considerazioni, mostrando una grande variabilità per il biossido di azoto (dati non riportati); la numerosità e posizione delle stazioni da traffico e di *background*, che possono raggiungere valori rispettivamente molto e poco elevati, risultano in entrambi i casi poco rappresentative della esposizione media della popolazione, se non considerati nel loro insieme. Ulteriori analisi comparative che tengano conto di una media pesata sulla tipologia di stazione potrebbero fornire indicazioni più precise.

Questo lavoro fornisce uno strumento operativo e immediatamente fruibile per la stima dell'esposizione media annua della popolazione residente sul territorio italiano, a partire da un modello *random-forest* costruito su un insieme di dati liberamente disponibili (dati di uso del territorio, modelli di dispersione e dati satellitari), e calibrato sui dati di monitoraggio della qualità dell'aria forniti dal sistema agenziale SNPA. È utile sottolineare il grande vantaggio di raccogliere dati uniformi a livello nazionale, che si traduce nella possibilità di sviluppare studi su tutto il territorio italiano, così come indagini su aree più ristrette (regione, provincia, comune), ma che consentano la comparabilità dei risultati tra aree diverse per quanto riguarda l'esposizione della popolazione all'inquinamento atmosferico, in quanto basati su protocolli di raccolta e analisi dei dati comuni e standardizzati. Questo strumento è disponibile e gestito dalla Rete italiana ambiente e salute (RIAS; https://rias.epiprev.it/), che coinvolge le strutture sanitarie e ambientale del Paese. Potrà, quindi, essere utilizzato per la valutazione dell'esposizione della popolazione in studi sulla relazione fra inquinamento atmosferico e salute umana. Come prima applicazione, sarà utilizzato dal progetto nazionale EpiCovAir, promosso dall'ISS e dal SNPA, in collaborazione con la Rete italiana ambiente e salute, che prevede un programma di studi epidemiologici per indagare le possibili relazioni fra inquinamento atmosferico e COVID-19. In modo analogo, potrà essere utilizzato per analoghi studi regionali e locali su inquinamento atmosferico e COVID-19, così come sullo studio delle relazioni fra inquinamento e altre patologie.

Conflitti di interesse dichiarati: nessuno.

Data di sottomissione: 22.07.2020
Data di accettazione: 26.08.2020

BIBLIOGRAFIA

1. Health effects institute. State of global air 2019. Boston, USA, 2019. Disponibile all'indirizzo: www.stateofglobalair.org
2. Glencross DA, Tzer-Ren H, Camiña N, Hawrylowicz CM, Pfeffer PE. Air pollution and its effects on the immune system. Free Radic Biol Med 2020;151:56-68.
3. Grigg J. Air pollution and respiratory infection: an emerging and troubling association. Am J Respir Crit Care Med 2018;198(6):700-01.
4. Conticini E, Frediani B, Caro D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ Pollut 2020;261:114465.
5. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and Covid-19 mortality in the United States: A nationwide cross-sectional study. medRxiv. 2020;2020.04.05.20054502.
6. Liang D, Shi L, Zhao J, et al. Urban air pollution may enhance Covid-19 case-fatality and mortality rates in the United States. medRxiv. 2020;2020.05.04.20090746.
7. Ancona C, Angelini P, Bauleo L et al. Inquinamento atmosferico e epidemia Covid-19: la posizione della Rete Italiana Ambiente e Salute. Disponibile all'indirizzo: https://repo.epiprev.it/1145
8. Istituto superiore per la protezione e la ricerca ambientale. Cattani G, Di Menno A, Fioravanti G et al. Analisi dei trend dei principali inquinanti atmosferici in Italia (2008-2017). Roma, Ispra, 2018.
9. Hoek G, Beelen R, de Hoogh K et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 2008;42(33):7561-78.
10. Ranzi A. Stima della variabilità spaziale dell'esposizione a inquinanti aerodispersi in ambito urbano. Epidemiol Prev 2012;36(5):287-95.
11. Kloog I, Nordio F, Coull BA, Schwartz J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the Mid-Atlantic states. Environ Sci Technol 2012;46,11913-21.
12. De Hoogh K, Héritier H, Stafoggia M, Künzli N, Kloog I. Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. Environ Pollut 2018;233:1147-54.
13. Badaloni C, Cattani G, De' Donato F et al. Big data in epidemiologia ambientale. Dati satellitari e uso del territorio per la stima delle esposizioni a livello nazionale. Epidemiol Prev 2018;42:46-59.
14. Stafoggia M, Schwartz J, Badaloni C, et al. Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. Environ Int 2016;99:234-44.
15. Stafoggia M, Bellander T, Bucci S et al. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model. Environ Int 2019;124:170-79.