



STIMATORE STATISTICO LINEARE PER LA STIMA DELLA CONCENTRAZIONE MEDIA GIORNALIERA DI PM_{10}

Andrea Bolignano, Roberto Sozzi(), Matteo Morelli*

Alessandro Domenico Di Giosa

ARPA Lazio – Divisione Atmosfera e impianti (Servizio Tecnico)

Sergio Ceradini, Fabrizio Sacco, Aldo Di Giulio

ARPA Lazio – Servizio Aria (Sezione Provinciale di Roma)

(*) roberto.sozzi@arpalazio.it

Riassunto – Secondo la Direttiva Comunitaria 2008/50/CE la valutazione della qualità dell'aria è, in primo luogo, la ricostruzione dei campi spaziali e temporali della concentrazione di alcune specie chimiche potenzialmente dannose per la salute umana ed, in secondo luogo, la verifica punto a punto tra alcuni indicatori statistici caratteristici di ciascun inquinante (concentrazione media, ecc.) e gli standard di legge. Per fare ciò, normalmente è necessario e conveniente assimilare le concentrazioni misurate all'interno di sistemi modellistici e fondamentale a questo scopo è un'ampia disponibilità di dati misurati di concentrazione. In pratica, la fonte informativa principale è costituita dalle misure prodotte da una rete di monitoraggio della qualità dell'aria, ma la loro effettiva disponibilità è limitata da alcuni problemi pratici. In primo luogo, la validazione delle misure di concentrazione prodotte dai vari analizzatori, cioè la decisione se ciò che è prodotto da un analizzatore è effettivamente una misura o meno, è un'impresa difficile da mettere in atto e sarebbe veramente utile a tale scopo una metodologia automatizzabile e programmabile a supporto degli operatori della rete. In secondo luogo, nonostante la cura che si pone alla manutenzione degli analizzatori presenti nella rete di monitoraggio, praticamente sempre si hanno dati mancanti nelle serie storiche ottenute e ciò può produrre seri problemi nella valutazione della qualità dell'aria. Sarebbe utile poter disporre anche in questo caso di una metodologia obiettiva di *gap-filling* per un'affidabile stima delle misure non disponibili. Infine, è possibile e conveniente realizzare misure di concentrazione dei vari inquinanti usando sistemi mobili di monitoraggio posti in differenti siti del territorio, ma necessariamente queste misure sono limitate nella loro durata (normalmente alcune settimane). Usare questa metodologia operativa implica inevitabilmente che per ciascun sito, sede della campagna sperimentale, le serie storiche di concentrazione risulteranno incomplete, molto utili a fini esplorativi, ma ben poco utili per la valutazione della qualità dell'aria o per il Data Assimilation. Queste sono le ragioni per cui in questo lavoro si è sviluppato, limitatamente alla concentrazione media giornaliera di PM_{10} , uno stimatore statistico di tipo BLUE (Best Linear Unbiased Estimator) basato sulle serie storiche operativamente disponibili al Centro Regionale della Qualità dell'Aria. Dopo aver presentato le basi matematiche e statistiche necessarie allo sviluppo, vengono presentate tre metodologie distinte, tutte basate sullo stimatore BLUE: una a supporto della validazione giornaliera delle misure, una per il *gap-filling* dei dati mancanti ed una per ricostruire serie storiche complete a partire dalle misure realizzate durante campagne sperimentali con mezzi mobili di durata limitata, ma ripetute periodicamente nel medesimo sito. In questo articolo vengono presentati tutti i dettagli delle metodologie sviluppate ed anche esempi di loro applicazioni pratiche in ambito operativo. Infine, l'articolo si conclude indicando le direzioni della ricerca futura.

Summary – According to European Directive 2008/50/CE, the Air Quality Assessment is represented by the reconstruction, in space and time, of the concentration fields of some chemical species dangerous for human health and the evaluation, point by point, of statistical quantities (mean concentration, number of exceedances, etc.) defined by the regulatory standards. To achieve these objectives, in general, the use of data-assimilation techniques (between concentration measurements and modeling system) is necessary and, for this scope, profitable is a wide availability of the concentration data. The measurements provided by the air quality monitoring network are the main information source but, in general, the real availability of this data is limited by some operative problems. Firstly, the validation of pollutant concentrations provided by an analyzer, that is the decision if this measure is reliable or not, is a delicate aspect because of its intrinsic subjectivity and an objective methodology, possibly automatic and programmable, could be very helpful for network's operators. Secondly, in spite of the most care posed in the maintenance of the analyzers installed on monitoring network, the presence of missing data on time series of pollutant concentrations can prevent a correct air quality assessment. In this case a gap-filling methodology finalized to a reliable estimation of the missing data should be helpful. Finally it is possible and profitable to carry out measures of pollutant concentration with the use of mobile systems located in different points of the territory but, necessarily, these measurements are limited in time (normally few weeks).

The use of this operative methodology imply that for each point, where the experimental campaign is performed, the availability of time series is necessarily incomplete. This is very helpful for exploratory purpose, but no very suitable for air quality assessment nor for data assimilation. This is the reason why a BLUE estimator (Best Linear Unbiased Estimator), based on time series of PM10 daily mean concentrations provided by the analyzers installed on the air quality monitoring network of Lazio Region and availables in the Air Quality Regional Center of Arpalazio, has been developed. After the description of the mathematical and statistical aspects necessary for the development of BLUE estimator, three different operational approaches of methodology are showed: the first describes how the methodology can be used as the support to the validation procedure of PM10 daily concentrations, the second describes how it can be used for carry out the gap-filling of PM10 daily levels and the third application is aimed to rebuild a time series of PM10 starting from the measurements provided by an experimental campaign time-limited and periodic. In this paper are showed the details of each operative methodology with the results of his application in a operational framework. The paper ends showing developments we want carry out in the future.

Keyword – *air quality monitoring network, mobile laboratory, BLUE estimator, kriging, geostatistics*

1. Premessa

Anche se le emissioni in aria della maggior parte degli inquinanti sono diminuite in modo sostanziale negli ultimi decenni, l'inquinamento atmosferico continua ad essere un'emergenza ed uno dei maggiori fattori di rischio per la salute della popolazione in tutto il mondo. Mentre nei decenni passati i maggiori problemi noti erano costituiti dalle elevate concentrazioni al suolo di biossido di zolfo e di monossido di carbonio, oggi le preoccupazioni maggiori sono costituite dalle alte concentrazioni di biossido di azoto e, soprattutto, di particolato sottile (PM_{10} e $PM_{2.5}$) [1] [2].

Il controllo della qualità dell'aria, cioè dei livelli di concentrazione delle diverse sostanze inquinanti, viene realizzato tradizionalmente collocando su un territorio una rete di monitoraggio, cioè un insieme di postazioni fisse in cui operano vari analizzatori dedicati alla misura in continuo della concentrazione di una serie di sostanze che le attuali conoscenze sanitarie ed epidemiologiche reputano dannosi alla salute umana e a quella degli ecosistemi, sulla base di dati ormai inconfutabili. Ciò che emerge dalle misure delle reti di monitoraggio sia nel mondo industrializzato che in quello in via di sviluppo è una situazione piuttosto critica che non tende a significativi miglioramenti. Il problema riveste quindi una notevole gravità visto che continuano ad accumularsi conoscenze sempre più approfondite sulle conseguenze sanitarie a breve e a lungo termine che l'inquinamento dell'aria ha sulla salute umana. Ciò che emerge poi, da questi studi, è la necessità di possedere una conoscenza sempre più dettagliata dell'inquinamento atmosferico proprio per poter meglio definire le leggi che quantificano l'influenza della concentrazione al suolo degli inquinanti sulla salute umana, visto che i vari inquinanti si distribuiscono nello spazio e nel tempo in maniera per nulla uniforme, soprattutto nelle aree densamente urbane come le grandi città e le zone ad orografia complessa dove spesso si vanno ad insediare rilevanti installazioni industriali.

Le reti di monitoraggio, che garantiscono la continuità della misura della concentrazione degli inquinanti, presentano evidenti limiti operativi. In primo luogo le postazioni di misura che le compongono non possono essere numerose come si vorrebbe per i problemi economici legati sia agli investimenti necessari al loro acquisto che per i costi richiesti dalla loro manutenzione, costi che risultano più che proporzionali al numero di postazioni di una rete. Inoltre, la tecnologia attualmente utilizzata (e prescritta) è piuttosto invasiva e non consente, se non a prezzo di complicazioni notevoli, di monitorare all'interno delle città, soprattutto di quelle città storiche europee la cui geometria è estremamente complessa. Per questa ragione la Direttiva 2008/50/CE [3], valida in tutta l'Unione Europea, consiglia per il controllo della qualità dell'aria l'uso delle reti di monitoraggio tradizionali affiancate da altre metodologie meno onerose ed invasive come l'utilizzo di modelli matematici, di metodi statistici obiettivi e di misurazioni mobili, il tutto finalizzato a garantire una informazione continua e dettagliata della concentrazione al suolo delle principali sostanze inquinanti.

Tali sostanze inquinanti vengono emesse da una miriade di sorgenti quali: le varie ciminiere industriali, gli autoveicoli che percorrono le arterie stradali, la combustione necessaria per il riscaldamento invernale degli edifici, le navi che stazionano nei porti e gli aerei. Tutte queste sostanze, una volta emesse dalle diverse sorgenti emittenti, vengono trasportate dalle masse d'aria in movimento entro la parte bassa della troposfera (il PBL, Planetary Boundary Layer), vengono disperse dalla turbolenza che caratterizza permanentemente questo strato di atmosfera e reagiscono chimicamente tra loro generando sostanze inquinanti secondarie come l'ozono ed una parte del particolato sottile che, pur non essendo emesse direttamente, compaiono spesso al suolo con concentrazioni rilevanti. E' quindi immediato notare come la qualità dell'aria (cioè la concentrazione delle diverse sostanze inquinanti al livello del suolo) non dipenda solo dal loro tasso di emissione, ma anche e soprattutto dalle caratteristiche disperdenti del PBL. Come ciò accada è ormai ben noto: Dutton [4] descrive dettagliatamente la fisica che governa i moti atmosferici, Stull [5], Garratt [6] e Sorbjan [7] introducono le peculiarità della turbolenza meccanica e convettiva che governa le capacità disperdenti del PBL, Kaimal e Finnigan [8] e Sozzi e al. [9] trattano le metodologie di misura e di stima della stessa, Venkatram e Wyngaard [10] presentano la complessità della dispersione degli inquinanti in aria e Seinfeld e Pandis [11] nel loro monumentale trattato raccolgono e dettagliano l'intera materia assieme alla chimica dell'atmosfera. Il quadro che emerge è di una complessità straordinaria (ben lontano dal quadro semplicistico ed erroneo fornito da alcune fuorvianti presentazioni ingegneristiche), ma ad oggi la modellizzazione di tutto ciò è una realtà concreta come mostrato da Pielke [12] e da Jacobson [13]. Tanto concreta da consentire la realizzazione di sistemi modellistici altamente realistici ed utilizzabili per la previsione e la ricostruzione della qualità dell'aria, estremamente preziosi anche per stimare l'impatto dell'inquinamento atmosferico sulla salute umana [14].

La visione diretta della qualità dell'aria mediante sistemi strumentali di monitoraggio (come le reti fis-

se) e la sua ricostruzione (e previsione) modellistica affrontano il problema della conoscenza della qualità dell'aria da due punti di vista totalmente opposti.

In effetti, la visione diretta della qualità dell'aria ottenuta dall'osservazione strumentale della concentrazione al suolo delle diverse specie inquinanti ignora completamente le cause che hanno condotto alla situazione osservata (tassi di emissione delle sostanze inquinanti e capacità disperdente dell'atmosfera) e si concentra esclusivamente nel fornire un quadro il più realistico possibile della situazione in atto, il tutto con una serie di difficoltà che di fatto risultano insuperabili. Infatti, i campi di concentrazione delle diverse sostanze non possono essere osservati strumentalmente nella loro interezza (almeno con la tecnologia attualmente disponibile), ma possono essere solo *campionati* in un numero ridottissimo di punti del territorio cui si è interessati. Dato che tali campi ben raramente sono regolari e lentamente variabili nello spazio e nel tempo (soprattutto nelle città e nelle zone complesse come le zone montane, vallive e costiere), ciò che in effetti si riesce a ricostruire con le misure fornite da una rete di monitoraggio risulta privo di dettagli importanti, lasciando estese porzioni del territorio completamente prive di informazione. Inoltre, non va trascurata e minimizzata l'incertezza intrinseca alle misure cui spesso non si fa cenno alcuno ma che è ben lungi dall'essere trascurabile. In sostanza, il quadro che si ottiene dall'impiego della sola rete di monitoraggio fissa della qualità dell'aria è normalmente altamente incompleto e decisamente incerto, da qui la raccomandazione della Direttiva Europea 2008/50/CE di completare il quadro che emerge dalle misure di una rete fissa con altre metodologie come, per esempio i modelli matematici.

L'impiego dei modelli matematici teoricamente supererebbe tutte queste limitazioni e, sempre teoricamente, dovrebbe fornire un quadro completo ad alta risoluzione della qualità dell'aria di un territorio, oltre che realistiche previsioni della sua evoluzione nel tempo. Infatti, dato che il modello fisico-chimico che descrive la dispersione degli inquinanti in atmosfera è costituito da un insieme di equazioni differenziali (alcune delle quali alle derivate parziali), se si conosce con esattezza l'insieme delle condizioni iniziali ed al contorno ed i termini di sorgente (per esempio il tasso di emissione delle varie sostanze chimiche), la soluzione del problema discende immediatamente almeno in termini numerici, visto che la complessità delle equazioni del modello impedisce una loro soluzione analitica, e questa soluzione è costituita da un insieme di campi quadridimensionali (spaziali e temporali) il cui numero è pari ai campi meteorologici e micrometeorologici considerati ed alle sostanze inquinanti cui si è interessati. Ma tutto ciò è vero solo in teoria; infatti le condizioni iniziali ed al contorno di tipo meteorologico e chimico sono note con un certo grado di incertezza ma, soprattutto, elevata incertezza caratterizza sia la localizzazione che l'entità delle varie emissioni di inquinanti. Questi errori si propagano nel modello ed i campi ricostruiti e/o predetti dallo stesso sono affetti da un errore spesso notevole. E questa è la ragione per cui la Direttiva 2008/50/CE non consente, se non in zone con particolari caratteristiche, l'uso esclusivo dei modelli di dispersione per la valutazione della qualità dell'aria.

Ovviamente l'ideale è una forma di *fusione* dei due punti di vista: l'osservazione strumentale e la ricostruzione modellistica. Questo problema è ben noto da tempo in campo meteorologico ed è stato affrontato con successo con svariate tecniche matematiche, tutte volte a costruire procedure di calcolo che, impiegando tutte le informazioni possibili (leggi modellistiche ed osservazioni sperimentali), conducessero ad una stima la più realistica possibile dello stato dell'atmosfera. Tutte queste procedure vengono normalmente indicate collettivamente come tecniche di Data Assimilation [15] [16]. Nell'ultimo decennio, esse sono state estese anche ai campi delle specie chimiche inquinanti con notevole successo ed i vari sistemi modellistici impiegati per la valutazione della qualità dell'aria impiegano questa tecnica efficacemente.

Le tecniche di Data Assimilation estraggono informazioni pertinenti dalle osservazioni disponibili e le trasmettono alla struttura matematica del modello con l'intento di minimizzarne gli errori dovuti all'incertezza associata alle condizioni iniziali ed al contorno ed ai termini di sorgente considerati. Pertanto, maggiore è il numero di misure disponibili, più efficace sarà, almeno in teoria, l'azione di miglioramento delle ricostruzioni modellistiche. Come si è visto, aumentare il numero delle postazioni fisse di misura di una rete di monitoraggio non è agevole e spesso risulta impossibile, mentre è normalmente possibile realizzare misure occasionali (per esempio le misure realizzate con i mezzi mobili) che, però, proprio a causa della loro occasionalità forniscono direttamente solo informazioni parziali e di difficile utilizzo.

Nella regione Lazio, in ottemperanza ai dettami della Direttiva 2008/50/CE (e del D.Lgs. 155/2010 [17] che la recepisce), opera una moderna rete di monitoraggio della qualità dell'aria di tipo tradizionale completamente integrata con un sistema modellistico (ricostruttivo e previsionale) costituita da un modello meteorologico ad elevata risoluzione accoppiato ad un modello fotochimico per la dispersione in aria degli inquinanti. Il *sistema di monitoraggio* così ottenuto (costituito dall'insieme della rete strumentale e della catena modellistica) è in grado di *valutare* la qualità dell'aria regionale, cioè di ricostruirla a livello regionale nello spazio e nel tempo e di individuare il rispetto degli indicatori statistici prescritti dalla norma, ma la sua reali-

sticità è tanto maggiore quanto maggiore è l'informazione strumentale disponibile. Per questa ragione, nel presente lavoro viene messa a punto una metodologia che consente di aumentare le informazioni strumentali disponibili, dirette ed indirette, in modo da fornire alle tecniche di assimilazione la maggior informazione possibile sui fenomeni in atto. Per semplicità, ci si è limitati a considerare solo le concentrazioni medie giornaliere di PM_{10} che comunque costituiscono una delle maggiori criticità ambientali. I problemi che limitano la disponibilità effettiva di misure di particolato sottile sono sostanzialmente tre:

- le misure prodotte dai vari analizzatori presenti nelle postazioni fisse della rete di monitoraggio, pur presentando valori all'apparenza del tutto realistici, possono talvolta essere il risultato di anomalie strumentali non immediatamente percepibili da parte degli operatori. In questo caso, se queste misure fossero impiegate in un processo di assimilazione, introdurrebbero ulteriori incertezze nel sistema modellistico impedendogli di ricostruire realisticamente la distribuzione nello spazio e nel tempo della concentrazione media giornaliera di particolato sottile. Sarebbe quindi importante poter disporre in ogni punto di misura di un *metodo di stima obiettiva* in grado di fornire una stima realistica della concentrazione media giornaliera di PM_{10} da confrontare con la misura ottenuta dall'analizzatore, in modo da costruire un efficiente sistema di *Quality Control* delle misure contribuendo al processo di validazione delle stesse;
- spesso le misure non sono disponibili a causa di avarie dell'analizzatore ed in questo caso si perde una preziosa informazione nel sistema di ricostruzione della qualità dell'aria. Se per ogni punto di misura fosse disponibile un *metodo di stima obiettiva* di nota incertezza, sarebbe possibile compensare l'indisponibilità della misura strumentale con una sua stima attendibile (*gap filling*).
- se poi fosse possibile costruire questo *metodo di stima obiettiva* anche sulla base di misure occasionali e discontinue nel tempo realizzate con l'ausilio di mezzi mobili posti in alcuni punti di interesse del territorio, l'informazione disponibile per la valutazione della qualità dell'aria aumenterebbe notevolmente e di conseguenza aumenterebbe pure la realistica della valutazione stessa. In pratica, si tratterebbe di realizzare in alcuni siti selezionati del territorio regionale campagne sperimentali con mezzi mobili di durata limitata (per esempio due settimane ciascuna) ma ripetute periodicamente nelle diverse stagioni. In questo modo si otterrebbero in questi siti delle serie parziali della concentrazione media giornaliera di PM_{10} . Sarebbe auspicabile mettere a punto sulla base di queste informazioni strumentali parziali un modello di stima obiettiva che faccia tesoro di queste informazioni e che produca costantemente, per ciascuno di questi siti, una stima attendibile della concentrazione di PM_{10} anche quando il mezzo mobile è assente.

Nei punti che seguono verrà per prima cosa descritta la teoria su cui si fonda il metodo di stima obiettiva proposto e se ne verificheranno le *performances* teoriche impiegando le serie storiche di PM_{10} rilevate dalla rete regionale della qualità dell'aria presente nel Lazio. Successivamente si verificherà come e con che efficacia l'impiego del metodo di stima obiettiva messo a punto è in grado di effettuare un realistico *gap filling* ed infine si verificherà la possibilità di impiegare tale metodo in combinazione con le misure realizzate con i mezzi mobili creando quindi una sorta di *postazioni virtuali* che si aggiungono alle postazioni di una rete fissa aumentando complessivamente l'informazione disponibile per la valutazione della qualità dell'aria con un sistema modellistico che assimila costantemente le informazioni strumentali.

2. Teoria

2.1 Basi fenomenologiche

I modelli di dispersione degli inquinanti in aria (o meglio le catene modellistiche) si incaricano di ricostruire e/o di prevedere la distribuzione delle principali specie inquinanti in un dominio spaziale di interesse. Possono essere di diverse tipologie ma, se ci si limita a considerare quelli di maggiore complessità e di impiego più generale (i modelli fotochimici di tipo Euleriano), essi sostanzialmente procedono per passi temporali successivi ricostruendo, in primo luogo, i campi meteorologici tridimensionali ed i campi delle variabili che caratterizzano la turbolenza dei bassi strati della Troposfera, e successivamente si incaricano di trasportare, disperdere e far reagire chimicamente fra loro gli inquinanti emessi dalle varie sorgenti inquinanti presenti nel dominio. Pertanto, la concentrazione di un generico inquinante in un istante e in un punto dello spazio è una funzione di numerose variabili fisico/chimiche che, in linea di principio, possono essere suddivise in tre

macro-insiemi distinti: emissioni, caratteristiche chimiche e caratteristiche meteorologiche. La dipendenza funzionale della generica concentrazione dalle sorgenti emmissive (che descrivono quali, quanto ed in che modo le sostanze vengono iniettate in atmosfera ad ogni istante), la dipendenza dai parametri meteorologici e micrometeorologici (che descrivono in che modo le sostanze emesse vengono trasportate e disperse in aria) e dalle proprietà chimiche (che regolano la capacità delle sostanze immesse di reagire e trasformarsi chimicamente) è regolata da un insieme di leggi differenziali fisiche e chimiche che compongono la base teorica di partenza di un modello fotochimico. Quindi, se E è l'insieme di variabili legate alle emissioni, R l'insieme delle variabili legate alla chimica dell'atmosfera e M l'insieme delle variabili che descrivono la meteorologia e la micrometeorologia, la concentrazione di un inquinante i in un punto $\mathbf{x}_j = (x_j, y_j, z_j)$ al tempo t può essere descritta genericamente dalla relazione:

$$C_i(\mathbf{x}_j) = f(E, R, M; t) \quad (1)$$

dove la funzione $f(\dots)$ sarà inevitabilmente di una complessità formidabile. D'altro canto, la concentrazione di una generica sostanza in un punto è in qualche modo legata alla concentrazione della stessa sostanza, o di altre sostanze, in punti più o meno vicini al punto \mathbf{x}_j tramite relazioni che per forza di cose *contengono e conservano*, anche se in maniera generale, informazioni sulle leggi fisico/chimiche che realmente concorrono ai processi di formazione e trasformazione di tali sostanze in atmosfera e responsabili delle concentrazioni osservate. Per essere più specifici, ci si limiti a considerare, come detto, solo la concentrazione media giornaliera del particolato sottile PM_{10} . Se si considera un punto $\mathbf{x}_i = (x_i, y_i)$ di un dominio spaziale D (il territorio di interesse) in cui sia presente una stazione di monitoraggio i in grado di misurare la concentrazione media giornaliera di PM_{10} , è possibile conoscere i valori $C_i(t_k)$ di tale variabile in ognuno dei t_k periodi di mediazione (i vari giorni di un anno) per l'intero periodo di osservazione (l'anno appunto). Queste concentrazioni sono il risultato delle differenti combinazioni di emissioni dalle sorgenti e delle differenti condizioni meteorologiche e micrometeorologiche che si susseguono durante l'anno osservato e la variabile *concentrazione media giornaliera di PM_{10} in quel punto* può essere considerata, in assenza di ogni tipo di relazione causa-effetto, a tutti gli effetti una *variabile stocastica*. Di fatto, tra tutte le possibili sequenze di valori medi giornalieri che avrebbero potuto realizzarsi *a priori*, in un determinato anno si è ottenuto proprio quella realizzazione che l'analizzatore ha catturato strumentalmente. Si selezioni poi un punto generico $P_0(x_0, y_0)$ del dominio spaziale D in cui non sia presente un analizzatore di PM_{10} , e si supponga che in altri N punti diversi e distinti del territorio ($N > 1$) $\mathbf{x}_i = \{x_i, y_i\}$, $i = 1, \dots, N$ tale misura sia invece realizzata. Sia, poi, $C_{0t}(t_k)$ la concentrazione media di tale inquinante al tempo t_k caratteristica di P_0 . Ovviamente il valore di $C_{0t}(t_k)$ è incognito, vista l'assenza dell'analizzatore. Al tempo t_k saranno poi disponibili le rispettive misure $C_i(t)$ nelle N postazioni in cui sono presenti gli analizzatori.

Se adottassimo la maniera deterministica tradizionale per descrivere la dispersione degli inquinanti in atmosfera basata sull'impiego di modelli fotochimici di tipo Euleriano alimentati con campi meteorologici e micrometeorologici tridimensionali variabili nel tempo, sulla base dei tassi di emissione che le varie sorgenti inquinanti presentano al tempo t_k , saremmo in grado (in genere numericamente) di determinare il campo $C(x, y; t_k)$ per ogni punto del dominio di interesse D . Il modello deterministico, al di là degli aspetti puramente matematici, asserisce che le concentrazioni nei vari punti del territorio dipendono all'istante t_k dalle stesse condizioni meteorologiche ed emmissive e quindi di fatto asserisce che è sempre concettualmente possibile ritenere che la concentrazione in un punto generico del dominio dipenda *in qualche modo* dalla concentrazione tipica in altri punti del dominio, visto che tutte queste concentrazioni sono il risultato di quanto viene emesso al tempo t_k e di come ciò che viene emesso si disperde. Quindi, selezionati N punti, la concentrazione nel punto P_0 è inevitabilmente esprimibile nel modo generico seguente:

$$C_0(t) = F(C_1, C_2, \dots, C_N) \quad (2)$$

Se, poi, si ipotizza che la funzione F sia regolare, sufficientemente *smooth* ed espandibile in Serie di Taylor, l'espansione di $F(C_1, C_2, \dots, C_N)$ arrestata al primo ordine si riduce ad un modello lineare che di fatto asserisce che la concentrazione media di PM_{10} al punto P_0 può essere vista (a meno di un termine di errore più o meno elevato a seconda delle caratteristiche analitiche della funzione F) come una opportuna combinazione lineare del valore assunto dalla concentrazione media giornaliera di PM_{10} negli altri N punti considerati:

$$C_0(t) = \sum_{i=1}^N \lambda_i C_i(t) \quad (3)$$

Il modello (3) rappresenta quindi una versione linearizzata della relazione (2). Pertanto è ragionevole suppor-

re la validità del modello lineare (3) purché la funzione F sia ragionevolmente *smooth*. In termini pratici, si può asserire che il modello lineare (3) può essere ritenuto valido se:

- il territorio in cui sono collocate le N centraline ed il punto P_0 è poco esteso;
- non ci sono fenomeni tali da rendere altamente irregolare la distribuzione spaziale della concentrazione.

Dato che per il momento si sta considerando la concentrazione media giornaliera di PM_{10} , queste ipotesi possono essere ritenute valide nel territorio laziale almeno in prima approssimazione.

Se, però, non è possibile adottare la metodologia deterministica tradizionale, ci si può chiedere se la concentrazione vera $C_{0t}(t_k)$ sia *stimabile* da un'opportuna combinazione lineare delle N misure $C_i(t_k)$. In pratica, non conoscendo $C_{0t}(t_k)$, cerchiamo una sua *stima* $C_0(t_k)$ la più vicina possibile al valore vero (che non è noto) impiegando un modello lineare formalmente identico alla relazione (3) che consiste sostanzialmente nell'ipotizzare che la stima in P_0 al tempo t_k sia data da una opportuna combinazione lineare del valore misurato al tempo t_k nelle N postazioni in cui viene realizzata la misura.

Il modello lineare (3) è un *modello di variabilità spaziale* della variabile C e negli N pesi λ_i è concentrata tutta la conoscenza sul fenomeno. Se il modello (3) deriva da una semplificazione di un modello di dispersione, i pesi (che avranno forme analitiche estremamente complesse) saranno di tipo deterministico, mentre se il modello (3) deriva da pure considerazioni semiempiriche e statistiche, i pesi dovranno riflettere le caratteristiche statistiche osservate negli N punti di misura durante il periodo di osservazione senza poter tenere conto di alcun tipo di legge fisica o chimica.

L'impiego di modelli lineari di variabilità spaziale del tipo (3) che cercano di stimare il valore di una variabile generica C in un punto P_0 (*postazione pivot*) noti i valori di C in N punti distinti di un dominio D (*postazioni slave*) è molto comune in geologia, in idrologia e nelle prospezioni minerarie e la Geostatistica è la disciplina che studia proprio questo tipo di problemi. La Letteratura dedicata alla Geostatistica è molto vasta e in [18], [19] e [20] viene fornita una presentazione organica della materia cui si rimanda per gli eventuali approfondimenti. Qui di seguito viene delineato il modello lineare di tipo geostatistico impiegato nel presente lavoro: esso è il più semplice modello di questo tipo e viene indicato normalmente col nome di *ordinary Kriging*.

2.2 Struttura matematica del metodo

Si consideri un dominio spaziale bidimensionale di dimensioni finite D ed un insieme di N punti interni a D in cui per ogni periodo di mediazione t_k (il giorno) sia disponibile una misura della concentrazione media giornaliera C di PM_{10} . In sostanza, per ogni punto \mathbf{x}_i ad ogni istante t_k è nota la concentrazione $C_i(t_k)$. Ovviamente, per un tale dominio è possibile definire un valore medio m della concentrazione che rappresenta la concentrazione media giornaliera di PM_{10} dell'intero dominio e per l'intero anno ed ipotizziamo che m sia un valore costante. Non è necessario domandarsi quanto valga, è sufficiente ipotizzarne l'esistenza. Va notato che, a rigore, l'esistenza di un valore m costante non è completamente supportata dalle osservazioni sperimentali, infatti si nota facilmente come tale valore sia variabile da punto a punto anche se la sua variabilità non è in genere molto elevata. Tuttavia, per semplicità operativa, nel presente lavoro si ipotizza che m sia costante in tutto il dominio D anche se la Geostatistica presenta metodi che considerano la variabilità spaziale di m , a prezzo però di notevoli complicazioni aggiuntive. Questa è una delle due ipotesi piuttosto drastiche che vengono fatte nello sviluppo del presente modello. La seconda ipotesi, altrettanto drastica, è quella della isotropia spaziale. In sostanza si ipotizza che l'influenza reciproca tra due punti di misura dipenda dalla distanza che li separa ma non dalla direzione del vettore che li unisce. Ciò comporta che i pesi del modello lineare non abbiano una sostanziale dipendenza dalla direzione prevalente del moto delle masse d'aria. Questa approssimazione è ragionevole soprattutto quando si considerano concentrazioni medie giornaliere in un territorio come quello laziale fortemente caratterizzato dalla presenza della costa e quindi in presenza di un frequente regime di brezza. Per formalizzare in maniera più dettagliata il modello (3), o meglio i pesi λ_i che in esso compaiono, è necessario fare le considerazioni che seguono.

Una proprietà desiderabile per uno stimatore è il fatto di non possedere un errore sistematico intrinseco (*bias*). Per ipotesi, il valore medio di dominio m è costante e quindi, perché lo stimatore lineare sia *unbiased*, è necessario che il valor medio sull'intero periodo di osservazione della differenza fra la stima C_0 in P_0 ed il relativo valore vero C_{0t} sia nullo. Indicando con $E[...]$ l'operatore *media d'insieme* (cioè il valore atteso della Statistica), ciò comporta che:

$$E[C_0 - C_{0t}] = \sum_{i=1}^N \lambda_i m - m = \left(\sum_{i=1}^N \lambda_i - 1 \right) m = 0 \quad (4)$$

da cui risulta che tra i pesi presenti nella (3) deve sussistere la relazione:

$$\sum_{i=1}^N \lambda_i = 1 \quad (5)$$

Come è ben noto, uno stimatore lineare per cui la somma dei pesi è unitaria è un *interpolatore esatto*, il che significa che se lo si applica ad uno degli N punti da esso considerati, la stima ottenuta eguaglia la misura.

Una seconda considerazione importante è che l'intera fenomenologia della dispersione degli inquinanti nel PBL è racchiusa nella *covarianza spaziale* tra due generici punti (*two-point covariance*) $R(\mathbf{x}_i, \mathbf{x}_j)$ dove \mathbf{x}_i e \mathbf{x}_j sono due generici punti nel dominio bidimensionale D . Dato che si è ipotizzata una isotropia spaziale in D , la covarianza non dipenderà dalle coordinate dei due punti ma solo dalla distanza h che li separa:

$$h = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (6)$$

per cui bisogna che:

$$R(h) = E[(C(\mathbf{x}_i) - m) \cdot (C(\mathbf{x}_j) - m)] \quad (7)$$

La covarianza è uno degli indicatori statistici dedicati alla descrizione quantitativa della similarità tra due misure ottenute in siti distinti. Per come è definita, normalmente i punti di misura più vicini sono più simili tra loro di quelli più distanti e quindi è naturale che la covarianza spaziale diminuisca con l'aumentare della distanza. Purtroppo la presenza della media globale m in $R(h)$ rende tale grandezza di scomodo utilizzo e per questo viene sostituita dal *semivariogramma* definito come:

$$\gamma(h) = \frac{1}{2} E[(C(\mathbf{x}_i) - C(\mathbf{x}_j))^2] \quad (8)$$

Come si nota immediatamente, il semivariogramma, a differenza della covarianza spaziale, quantifica la dissimilarità tra due punti di misura. Idealmente il $\gamma(h)$ dovrebbe essere nullo quando la distanza è nulla e dovrebbe crescere con essa fino a raggiungere un valore circa costante a distanze superiori ad un certo valore (*range*). Quando le misure sono separate da una distanza superiore al range, il semivariogramma dovrebbe mantenersi circa costante e le due misure dovrebbero risultare sostanzialmente non correlate spazialmente tra loro. Con un po' di algebra [20] è facile verificare come tra $R(h)$ e $\gamma(h)$ sussista la relazione seguente:

$$\gamma(h) = -R(h) + R(0) \quad (9)$$

dove $R(0)$ è la varianza di C in D . E' importante notare come la covarianza spaziale ed il semivariogramma costituiscano il vero modello (nel nostro caso isotropico) della distribuzione spaziale e temporale della concentrazione media giornaliera di PM_{10} che verrà sfruttato completamente nel seguito. Queste grandezze costituiscono in un modello statistico la controparte delle relazioni fisico/chimiche in un modello deterministico.

A questo punto è possibile affrontare direttamente la determinazione dei pesi presenti nel modello lineare (3) utilizzando tutte le informazioni sperimentali disponibili. Per poter giungere a ciò è necessario che la differenza tra la stima della concentrazione fatta dal modello lineare in un punto P_0 , cioè $C_0(t)$, differisca il meno possibile dal valore vero $C_{0t}(t)$ che caratterizza tale punto. In termini matematici il funzionale:

$$MSE = E[(C_0 - C_{0t})^2] = -\sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(h_{ij}) + 2 \sum_{i=1}^N \lambda_i \gamma_{i0} \quad (10a)$$

che esprime l'errore quadratico medio (*MSE, mean square error*) della stima deve essere il minimo possibile e deve rispettare anche il vincolo:

$$\sum_{i=1}^N \lambda_i = 1 \quad (10b)$$

Questo problema rientra nella problematica della ricerca di uno stimatore ottimale della classe BLUE (*Best*

Linear Unbiased Estimator) che matematicamente è un tipico problema di ottimizzazione non lineare vincolata con vincoli di uguaglianza, problema che si risolve impiegando il ben noto metodo dei moltiplicatori di Lagrange. In termini pratici, rimandando a [20] per i dettagli, questo problema di ottimizzazione vincolata può essere facilmente riformulato in termini di algebra vettoriale. Infatti, se si definiscono i due vettori seguenti:

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_N \\ \nu \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -\gamma(P_0, P_1) \\ -\gamma(P_0, P_2) \\ \dots \\ -\gamma(P_0, P_N) \\ 1 \end{bmatrix} \quad (11a)$$

dove ν è il moltiplicatore di Lagrange, e la matrice dei coefficienti:

$$\mathbf{A} = \begin{bmatrix} 0 & -\gamma(P_1, P_2) & \dots & -\gamma(P_1, P_N) & 1 \\ -\gamma(P_2, P_1) & 0 & \dots & -\gamma(P_2, P_N) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ -\gamma(P_N, P_1) & -\gamma(P_N, P_2) & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \quad (11b)$$

la soluzione del problema di ottimizzazione vincolata sopra delineato risulta essere la soluzione del sistema lineare seguente:

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{b} \quad (12)$$

che, una volta risolto, fornisce i desiderati valori dei pesi λ_i nonché il valore del moltiplicatore di Lagrange ν . La soluzione del sistema (12) non pone troppe difficoltà essendo normalmente disponibili ben collaudati algoritmi di Calcolo Numerico [21]. In particolare, è consigliabile adottare l'algoritmo basato sulla decomposizione LU o, ancor meglio, il Singular Value Decomposition Method.

Ottenuta la relazione (12) e noto il semivariogramma γ per le varie coppie di punti, lo stimatore BLUE per la concentrazione media giornaliera di PM_{10} rappresentato dalla relazione lineare (3) risulta completamente definito ed applicabile. Ciò che resta da determinare è quale sia la sua incertezza intrinseca. Tale incertezza altro non è che l'errore quadratico medio MSE che [20] può essere anche espresso in termini di moltiplicatore di Lagrange, di pesi e di semivariogramma nella forma seguente:

$$MSE = E[(C_{0t} - C_0)^2] = -\nu + \sum_{i=1}^N \lambda_i \gamma(P_0, P_i) \quad (13)$$

formulazione calcolabile molto più facilmente.

Nelle normali applicazioni geostatistiche in cui l'obiettivo è la ricostruzione di un campo spaziale, il semivariogramma viene prima definito grossolanamente sulla base delle misure disponibili e successivamente da tale semivariogramma sperimentale viene ottenuta un'espressione analitica funzione della distanza h mediante tecniche di *best-fitting* [20]. Nel caso che si sta considerando non è necessario fare ciò, visto che le serie storiche di concentrazione media giornaliera di PM_{10} negli N punti che si considerano nel modello lineare (3) permettono di stimare una volta per tutte il semivariogramma per ogni coppia di punti. In particolare, se i e j sono due punti del dominio D in cui è disponibile una serie storica della concentrazione media giornaliera di PM_{10} , il semivariogramma relativo potrà essere stimato direttamente dai valori di concentrazione media oraria con la versione campionaria della relazione (8), cioè con:

$$\gamma_{ij} = \frac{1}{2N_k} \sum_{k=1}^{N_k} [C_i(t_k) - C_j(t_k)]^2 \quad (14)$$

dove N_k è il numero di giorni considerati. Questo procedimento di fatto limita la drasticità insita nell'ipotesi di costanza del valor medio di dominio e di isotropia spaziale. Ciò che non è immediato è come determinare il semivariogramma tra il punto P_0 e gli altri N punti e come aggirare i problemi legati ai dati mancanti nelle serie storiche: ciò lo si vedrà nei paragrafi successivi.

3. Ambito di applicazione del metodo

3.1 Il contesto territoriale

L'ambito territoriale di applicazione dello stimatore lineare qui sviluppato è l'intero territorio della regione Lazio in cui opera il Sistema di Monitoraggio della Qualità dell'Aria di Arpa Lazio. Tale sistema è costituito da una rete di monitoraggio fissa le cui misure vengono assimilate in continuo da due catene modellistiche distinte, entrambe costituite dalla combinazione del modello meteorologico prognostico RAMS [22] e del modello fotochimico FARM [23] [24] che operano in sequenza, completate poi da opportuni sistemi di Data Assimilation che si incaricano di sfruttare il contenuto informativo delle varie misure rilevate dalla rete strumentale. La prima delle due catene modellistiche è dedicata alla ricostruzione *near-realtime* della situazione in atto (realizzando la *fotografia* dei campi di concentrazione con un ritardo di 3 ore rispetto al tempo reale), mentre la seconda si incarica di realizzare le previsioni dei campi di concentrazione fino a 5 giorni.

La Fig.1 è una rappresentazione sintetica del territorio laziale che è caratterizzato da un'estesa zona appenninica e da una vasta zona costiera lungo la quale sono localizzate le poche zone pianeggianti.

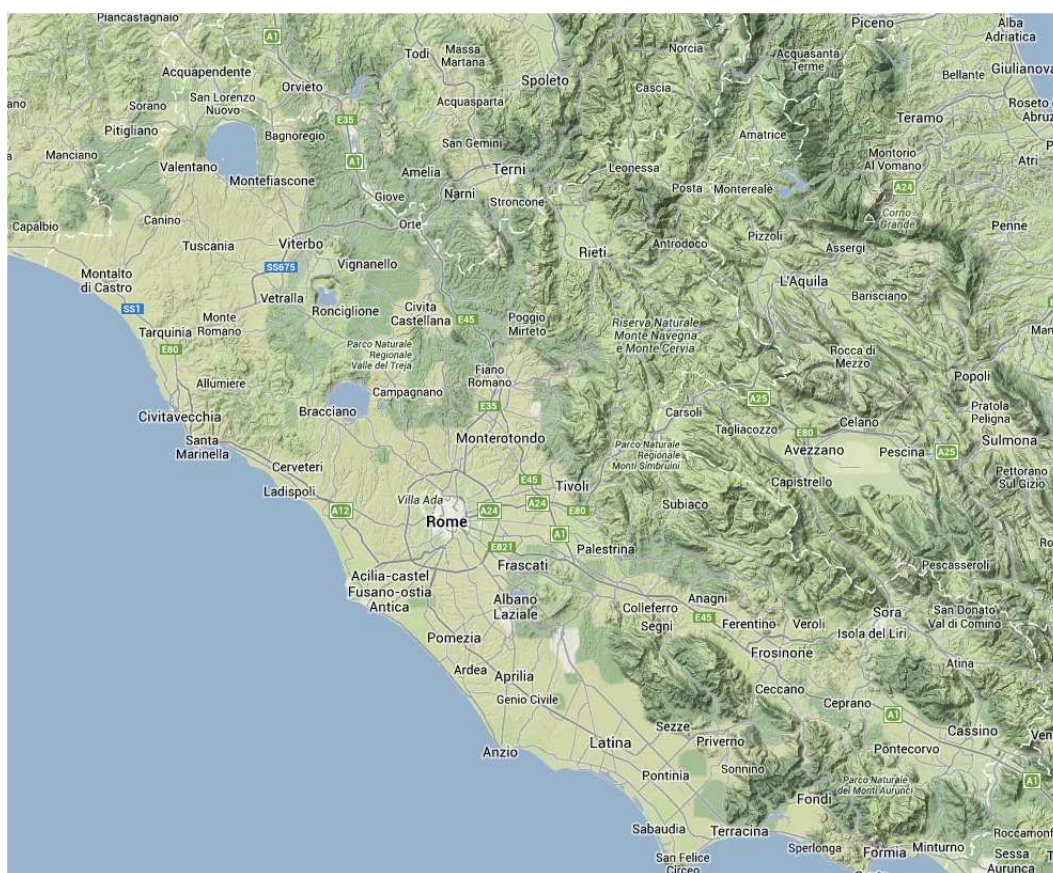


Fig.1: caratteristiche geografiche del territorio laziale

Viste le peculiarità geografiche (Fig.1) e le caratteristiche demografiche ed emissive che caratterizzano la regione, l'intero territorio è stato ripartito, secondo quanto prescritto dalla Direttiva 2008/50/CE e dal D.Lgs. 155/2010 che la recepisce, in un *agglomerato* e tre *zone* (Fig.2).

L'Area Metropolitana di Roma è l'unico agglomerato regionale ed è costituito dall'intera città di Roma (la parte storica e le vaste periferie che la circondano) e da un rilevante numero di comuni di varie dimensioni. Caratteristica comune a tutto questo agglomerato è il fatto che il territorio è sostanzialmente pianeggiante. Dal punto di vista morfologico, è caratterizzato da vaste aree altamente edificate in cui si inseriscono parchi urbani di ampie dimensioni e vaste aree agricole. L'Area Metropolitana di Roma è l'area più densamente popolata della regione e ciò determina una fortissima presenza di traffico autoveicolare che probabilmente è la maggiore fonte di pressione ambientale della zona a cui si aggiungono gli aeroporti di Ciampino e di Fiumicino, il porto di Fiumicino (di prossima espansione), alcune infrastrutture dedicate al ciclo dei rifiuti e un numero ridotto di installazioni industriali significative.

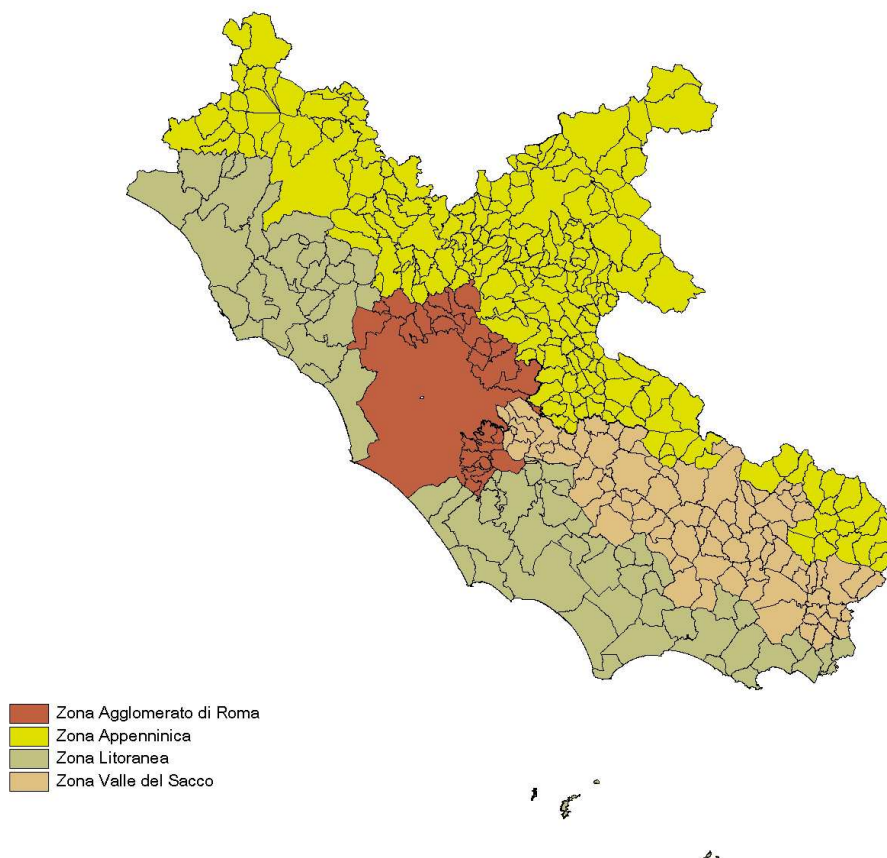


Fig. 2: la zonizzazione del territorio laziale

Le tre zone in cui è stato ulteriormente suddiviso il territorio laziale sono:

- la *zona litoranea* costituita da tutta la fascia costiera tirrenica della regione. In tale zona sono presenti numerose fonti di pressione ambientale di notevole importanza come le centrali termoelettriche di Montalto di Castro, di Civitavecchia (Torvaldaliga Nord a carbone e Torvaldaliga sud a turbogas), il porto di Civitavecchia (uno dei maggiori porti commerciali del Mediterraneo) ed il porto di Gaeta (di prossima espansione). Accanto a queste porzioni di territorio a vocazione prevalentemente industriale sono presenti anche aree di notevole interesse naturalistico in cui sopravvive la vegetazione tipica delle zone costiere tirreniche (macchia mediterranea). Un esempio significativo di ciò è rappresentato dalla Tenuta Presidenziale di Castel Porziano.
- la *zona della Valle del Sacco*, che è costituita da una vallata di grandi dimensioni, orograficamente molto complessa in cui sono presenti insediamenti abitativi di medie dimensioni (per esempio gli abitati di Frosinone e di Cassino) ed un numero rilevante di insediamenti industriali di medie dimensioni.
- la *zona appenninica* che comprende l'intero territorio collinare del viterbese, il territorio collinare e montano della Bassa Sabina e del reatino e la fascia appenninica che delimita a NordEst la Valle del Sacco.

3.2 La Base Dati considerata

Come detto, nel territorio regionale opera una rete di monitoraggio fissa della qualità dell'aria costituita, ad oggi, da 39 centraline, in ciascuna delle quali è sempre operativo un analizzatore di particolato PM_{10} . In Tab. 1 sono elencate le varie postazioni di misura della rete raggruppate secondo la zonizzazione regionale, con le rispettive coordinate. A queste centraline se ne aggiunge un'altra, ad oggi non inserita ufficialmente tra le postazioni di monitoraggio della rete regionale, localizzata al centro di Roma (in via Boncompagni, nei pressi di Piazza Fiume) e posta alla sommità della sede di rappresentanza di Arpa Lazio.

Stazione	Zona (D.G.R. 217/2012)	Latitudine	Longitudine	quota (slm)
Arenula	Agglomerato di Roma	41.89	12.48	31
Preneste	Agglomerato di Roma	41.89	12.54	37
Francia	Agglomerato di Roma	41.95	12.47	43
Magna Grecia	Agglomerato di Roma	41.88	12.51	49
Cinecitta	Agglomerato di Roma	41.86	12.57	53
Guidonia	Agglomerato di Roma	42.00	12.73	89
Villa Ada	Agglomerato di Roma	41.93	12.51	50
Castel di Guido	Agglomerato di Roma	41.89	12.27	61
Cavaliere	Agglomerato di Roma	41.93	12.66	48
Ciampino	Agglomerato di Roma	41.80	12.61	134
Fermi	Agglomerato di Roma	41.86	12.47	26
Bufalotta	Agglomerato di Roma	41.95	12.53	41
Cipro	Agglomerato di Roma	41.91	12.45	31
Tiburtina	Agglomerato di Roma	41.91	12.55	32
Malagrotta	Agglomerato di Roma	41.87	12.35	50
Colleferro-Oberdan	Valle del Sacco	41.73	13.00	219
Colleferro-Europa	Valle del Sacco	41.73	13.01	223
Alatri	Valle del Sacco	41.73	13.34	445
Anagni	Valle del Sacco	41.75	13.15	401
Cassino	Valle del Sacco	41.49	13.83	41
Ceccano	Valle del Sacco	41.57	13.34	130
Ferentino	Valle del Sacco	41.69	13.25	316
Fontechiari	Valle del Sacco	41.67	13.67	388
Frosinone Scalo	Valle del Sacco	41.62	13.33	161
FR-Mazzini	Valle del Sacco	41.64	13.35	251
Civ-Porto	Litoranea	42.10	11.79	6
Civ-Albani	Litoranea	42.10	11.80	34
Civ-Morandi	Litoranea	42.09	11.81	22
Allumiere	Litoranea	42.16	11.91	542
Civitavecchia	Litoranea	42.09	11.80	26
Aprilia 2	Litoranea	41.60	12.65	83
Latina-Scalo	Litoranea	41.53	12.95	18
Via Romagnoli	Litoranea	41.47	12.89	23
Via Tasso	Litoranea	41.46	12.91	21
Civita Castellana	Appenninica	42.29	12.41	139
Viterbo	Appenninica	42.42	12.11	338
Acquapendente	Appenninica	42.74	11.88	435
Leonessa	Appenninica	42.57	12.96	948
Rieti 1	Appenninica	42.40	12.86	397

Tab.1: le stazioni di monitoraggio della rete di qualità dell'aria del Lazio

Come si può notare, le postazioni di misura della rete sono maggiormente concentrate nell'Area Metropolitana di Roma e nella Valle del Sacco, mentre la copertura strumentale del resto della regione è decisamente inferiore. Anche se sarebbe logico considerare l'intero territorio regionale, nel presente studio, per semplicità, vengono prese in considerazione con dettaglio solo l'Area Metropolitana di Roma e la Zona della Valle del Sacco (le due porzioni di territorio laziale che costituiscono altrettanti paradigmi logici: una zona ad alta urbanizzazione ed una zona ad alta presenza industriale diffusa) e le serie storiche di concentrazione media giornaliera di PM_{10} rilevate dalle postazioni ivi operanti sono quelle relative all'anno 2012.

Il modello lineare BLUE (3) prevede di poter stimare la concentrazione media giornaliera di PM_{10} di un punto P_0 e in un generico giorno k come un'opportuna combinazione lineare delle analoghe concentrazioni medie rilevate in altri N punti distinti del dominio territoriale e verrebbe naturale pensare che questi N punti siano quelli più vicini a P_0 . Questa sensazione intuitiva sfortunatamente non è sempre suffragata dalla realtà [25]. Infatti, se si considera come esempio l'Area Metropolitana di Roma e le postazioni di misura ivi

operative, il quadro che emerge è piuttosto complesso e per nulla intuitivo. Per quantificare tutto ciò, è opportuno condurre sulla base delle misure raccolte un'analisi esplorativa volta a definire il *grado di correlazione* (parentela) *spaziale* che intercorre tra due distinti punti P_1 e P_2 del dominio spaziale impiegando a tale scopo l'indice di concordanza di Lin [26] definito come:

$$\rho_c = \frac{2\rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \quad (15)$$

dove μ_1 e μ_2 sono il valor medio della concentrazione media giornaliera di PM_{10} nei due punti P_1 e P_2 , σ_1 e σ_2 sono le relative deviazioni standard e ρ_{12} è il coefficiente di correlazione lineare tra le misure di concentrazione media giornaliera rilevate nei due punti (noto come coefficiente di correlazione di Pearson). L'indice di concordanza di Lin si è dimostrato ben più robusto della convenzionale correlazione p . Se lo si applica a tutte le possibili coppie di stazioni che operano entro l'Area Metropolitana di Roma (Fig.3) e si spazializza il valore assunto da tale indice con le normali tecniche di interpolazione geostatistica, ciò che si ottiene è molto singolare.

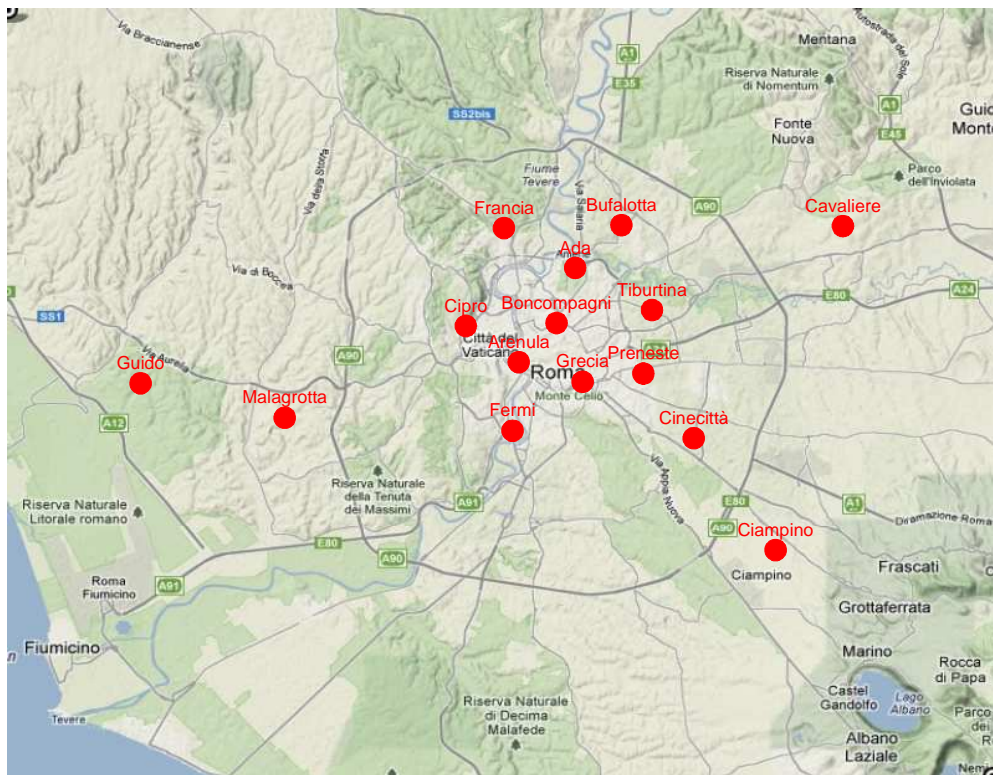


Fig.3: l'Area Metropolitana di Roma e le stazioni di monitoraggio presenti

Come esempio, si considerino cinque casi particolari relativi ad altrettante postazioni di misura:

- la postazione di Villa Ada (stazione di background urbano) posta in una vasta radura entro uno dei grandi parchi storici di Roma;
- la postazione di Castel Guido (stazione di background rurale) posta in un'area verde e quasi incontaminata ad Ovest di Roma verso il mare;
- la postazione di Via Magna Grecia (stazione da traffico) posta entro uno stretto canyon urbano vicino alla Basilica di S. Giovanni;
- la postazione di Via Enrico Fermi (stazione da traffico) sita in prossimità di un viale a traffico intenso, ma non inserita in un canyon urbano
- postazione di Via Boncompagni, stazione atipica, posta nei pressi di Piazza Fiume (centro di Roma) alla sommità di un edificio di cinque piani (sede di rappresentanza di Arpa Lazio), quindi al limite superiore del *canyon* urbano.

Consideriamo inizialmente la situazione a Castel di Guido. Dalla distribuzione dell'indice di concordanza di Lin (Fig.4a) si nota che questa stazione non *sente* la città di Roma (generalmente si colloca sopravvento alla città e quindi le masse d'aria che la raggiungono non sono ancora perturbate dalla presenza della città e dalle relative emissioni) ed è molto *parente* della stazione di Tenuta del Cavaliere (sito rurale posto ad Est di Roma e quindi normalmente sottovento alla città). Di fatto questa stazione acquisisce informazioni dalle zone marine ad Ovest di Roma (da cui provengono preferenzialmente le masse d'aria) e le distribuisce sulla città, vedendosele perturbare profondamente. Queste perturbazioni diminuiscono una volta che le masse d'aria abbandonano la città e nel contempo riemergono le condizioni di background rurale a Tenuta del Cavaliere. Pertanto, si viene a costituire una concordanza tra zone lontane tra loro e non contigue, cioè una *teleconnessione* seguendo la terminologia usata dai climatologi (Von Storch e Zwiers, 1999).

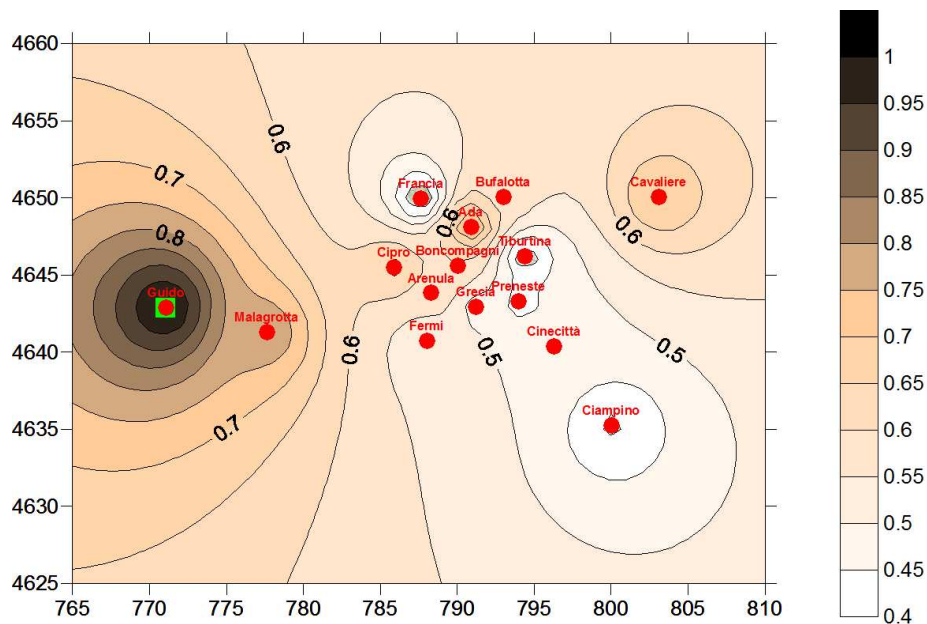


Fig.4a: correlazione spaziale tra la postazione di Castel di Guido e la restante parte dell'Area Metropolitana di Roma

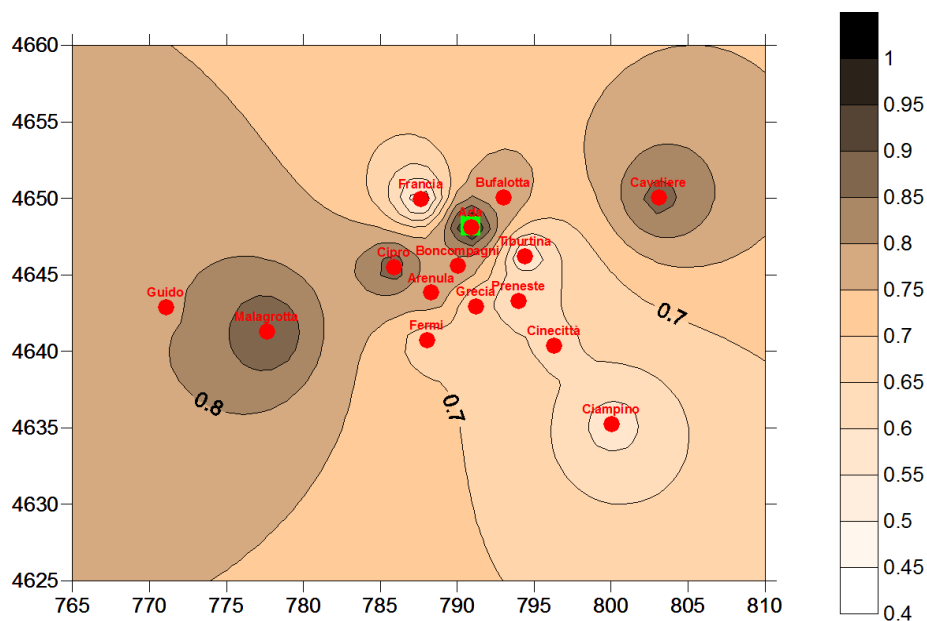


Fig.4b: correlazione spaziale tra la postazione di Villa Ada e la restante parte dell'Area Metropolitana di Roma

Consideriamo ora la situazione di Villa Ada (Fig.4b) che è situata in uno dei più antichi e vasti parchi

presenti nel centro storico di Roma, entro una vasta radura che si apre nel folto bosco di pini marittimi che caratterizza il parco. Analizzando l'indice di Lin che si ottiene confrontando le misure rilevate in questa postazione con le analoghe rilevate nelle altre postazioni della rete, si nota come questa postazione, pur ricevendo informazioni da Castel di Guido (in parte perturbate dal transito delle masse d'aria sull'Area Metropolitana di Roma), presenti caratteristiche specifiche proprie. Ciò è probabilmente dovuto alla presenza della vegetazione ed alle dimensioni del parco che ne fanno un ambiente a sé. Infatti, tale postazione ignora completamente tutto ciò che succede sull'asse NW-SE, come si vede dalla sua poca concordanza con la stazione di Ciampino. In pratica potremmo affermare che questa stazione sta in un *canyon vegetale poroso* e si limita a registrarne gli effetti (benefici).

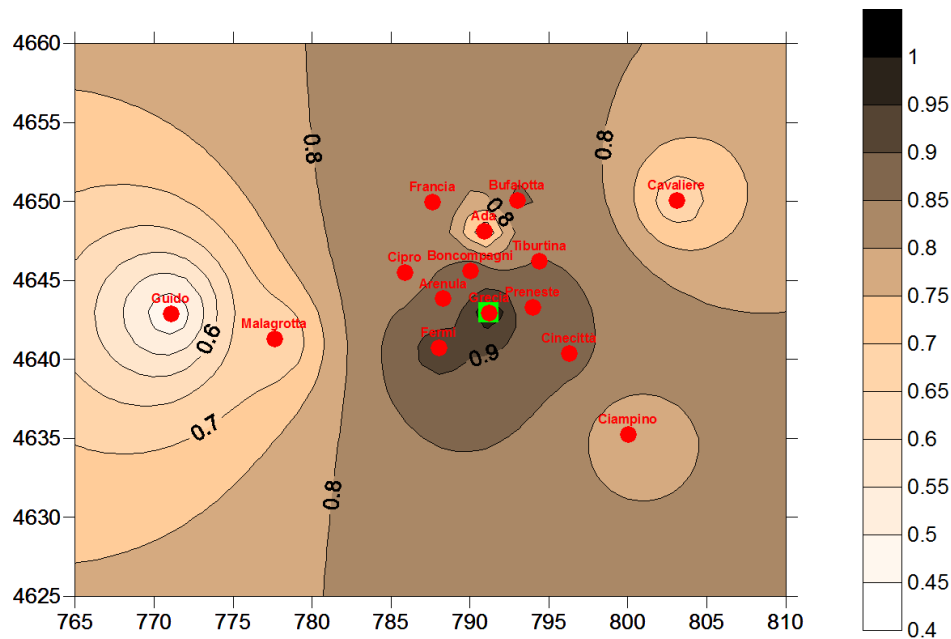


Fig.4c: correlazione spaziale tra la postazione di Via Magna Grecia e la restante parte dell'Area Metropolitana di Roma

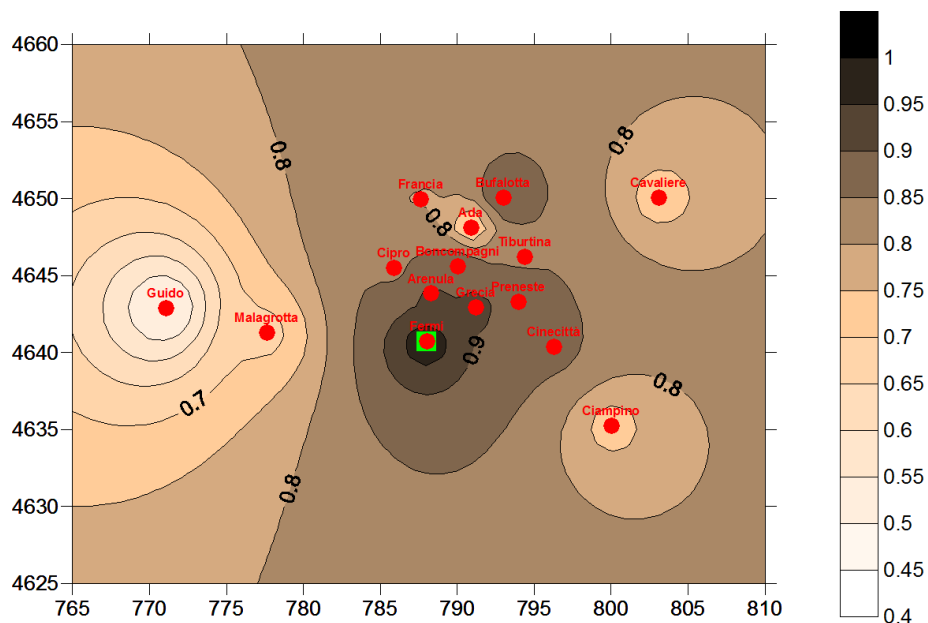


Fig.4d: correlazione spaziale tra la postazione di Via Enrico Fermi e la restante parte dell'Area Metropolitana di Roma

Analizzando, invece la Fig. 4c, risulta evidente come la stazione di Magna Grecia ignori quasi completamente le informazioni di background (quelle registrate dalle postazioni di Castel di Guido e di Tenuta del

Cavaliere) e rappresenta un mondo a sé, quello di un *canyon urbano* con scambi ridotti delle masse d'aria con le zone circostanti. Ovviamente risente del background rurale, ma il motore della variazione giornaliera di PM_{10} è prevalentemente costituito dalla variazione locali delle emissioni dovute al traffico autoveicolare.

Se però si considera la stazione di Via Enrico Fermi (Fig.4d), che è sempre una stazione da traffico ma non collocata in un canyon urbano, si nota come anch'essa sia per una certa parte un mondo a sé dipendente dalle locali emissioni del traffico veicolare. Tuttavia la sua esposizione meno confinata (non essendo posta entro canyon urbano) fa sì che la qualità dell'aria locale tenga conto meglio del background rurale, ma anche delle informazioni provenienti da vaste zone della città anche notevolmente distanti, anche sull'asse NW-SE, come per esempio le informazioni di Ciampino. E' interessante notare come, paradossalmente, la stazione da traffico di Fermi presenti un grado di concordanza con le altre stazioni della rete, anche lontane e con caratteristiche differenti, maggiore rispetto a quanto succeda per la stazione di Villa Ada. Queste sono ulteriori evidenti situazioni di teleconnessione.

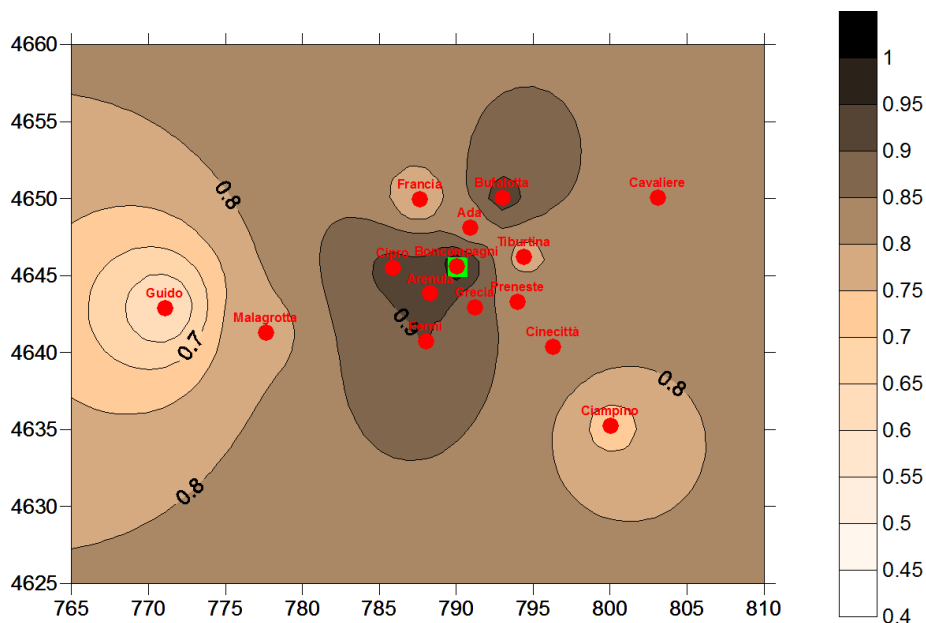


Fig.4e: correlazione spaziale tra la postazione di Via Boncompagni e la restante parte dell'Area Metropolitana di Roma

L'analisi della distribuzione spaziale dell'indice di concordanza di Lin realizzata per la stazione di via Boncompagni (Fig.4e) riserva una grossa sorpresa. Essendo questa stazione localizzata al *blending height*, cioè ad una quota in cui le informazioni provenienti dalle varie parti della città si mischiano tra loro, riceve le informazioni da tutta la città, anche dalle parti più lontane, anche dalle stazioni da traffico e da quelle poste entro i canyon urbani, e le miscela tra loro e con le informazioni provenienti dal background rurale. Di fatto concorda con tutte le stazioni della rete ed in sostanza fornisce una visione media e ponderata dello stato della concentrazione media giornaliera di PM_{10} di tutta l'Area Metropolitana di Roma (meno ottimistica delle stazioni di background, meno pessimistica delle stazioni da traffico). Di fatto, le misure di PM_{10} rilevate presso questa postazione possono essere considerate a ragione la situazione media della città di Roma.

Da tutte queste considerazioni emerge come l'interconnessione che esiste tra la concentrazione media giornaliera di PM_{10} tra le varie stazioni della rete di monitoraggio sia piuttosto complessa e per nulla scontata, soprattutto in ambito urbano e ciò determina la necessità di mettere a punto un algoritmo ottimale per l'individuazione delle postazioni *slave* da utilizzare nell'ambito dello stimatore lineare (3), sempre basandosi sulle serie storiche rilevate nell'anno 2012 dalle stazioni della rete di monitoraggio regionale che operano nell'Area Metropolitana di Roma e nella Valle del Sacco.

4. Caratterizzazione dello stimatore lineare

La (3) è l'espressione matematica dello stimatore BLUE per la concentrazione media di PM_{10} . Una volta introdotti gli aspetti matematici che lo caratterizzano, è necessario ora definirne le modalità di applicazione e l'incertezza intrinseca. Per giungere a ciò, è necessario fare alcune importanti considerazioni.

In primo luogo la relazione (3) asserisce che la stima della concentrazione media giornaliera di PM_{10} in un punto P_0 (postazione *pivot*) nel giorno k è pari alla somma pesata delle analoghe concentrazioni rilevate in N postazioni *slave*. La prima regola che si è posta nell'utilizzo pratico dello stimatore (3) è che sia la postazione *pivot* P_0 che le N postazioni *slave* si collochino tutte nella medesima area territoriale (nello specifico, o tutte nell'Area Metropolitana di Roma o tutte nella Zona della Valle del Sacco).

La seconda regola è che sia P_0 che le postazioni *slave* siano postazioni della rete fissa regionale della qualità dell'aria o sedi di campagne di monitoraggio periodico realizzate con mezzi mobili. Ciò comporta la disponibilità (nell'anno 2012, visti i limiti imposti al presente studio) di serie storiche più o meno complete per tutte le postazioni coinvolte (la postazione *pivot* e le postazioni *slave*) e quindi la possibilità di determinare i semivariogrammi necessari impiegando per questo lo stimatore campionario (14).

La terza regola si occupa di determinare quante postazioni *slave* vengono coinvolte dallo stimatore BLUE, cioè quanto vale N . E' abbastanza intuitivo che maggiore è N , minore è l'incertezza intrinseca dello stimatore, visto che con N aumenta l'insieme di informazioni che vengono trasferite alla stima da parte di ciascuna delle postazioni *slave*. Tuttavia per ragioni pratiche N non può essere molto elevato. Nel presente studio si è ipotizzato che $N = 4$ sia un numero di stazioni *slave* ragionevole da considerare.

Sia data una postazione *pivot* appartenente ad una delle due zone e siano M le postazioni presenti nella stessa zona (escludendo nel computo la stazione *pivot* stessa). Le possibili stazioni *slave* dovranno essere cercate tra le M stazioni rimanenti presenti nella zona considerata e quindi il numero di possibili quadruple di postazioni *slave* cercate dovrà essere pari al numero di possibili combinazioni di M oggetti di classe 4, cioè:

$$N_{comb} = \frac{M!}{4!(M-4)!} \quad [16]$$

Il numero di possibili quadruple associate ad una specifica postazione *pivot* è in generale elevato e ciascuna di esse potrebbe essere utilizzata dallo stimatore BLUE. Ciò che distingue l'uso di una quadrupla piuttosto che un'altra tra quelle possibili è la *bontà* della stima. La quarta regola si incarica quindi di individuare per ogni possibile quadrupla associata ad una postazione *pivot* un indicatore che rappresenti quanto la stima della concentrazione media giornaliera di PM_{10} sia prossima al valore reale per un intero periodo di osservazione. Nel presente lavoro sono stati impiegate le serie storiche rilevate nell'anno 2012 e questa procedura dovrà, a priori, essere ripetuta per ogni periodo di osservazione cui si è interessati. Per poter rendere operativa questa regola è necessario definire quanto una stima per la concentrazione media di PM_{10} in corrispondenza di una postazione *pivot* sia prossima al relativo valore *vero*. Per quantificare ciò si potrebbe far uso di molti e differenti indicatori statistici, anche se risulta naturale impiegare l'indicatore *MSE*, cioè lo scarto quadratico medio tra stima e valor vero dato dalla relazione (12), il più semplice e più naturale che si possa pensare. In sostanza, considerata una postazione *pivot*, si calcola il valore di *MSE* per ciascuna delle possibili quadruple ad essa associate e le si ordina per valore di *MSE* crescente: la prima delle quadruple è quella ottimale mentre l'ultima è quella che fornisce stime peggiori. Più in dettaglio, l'algoritmo impiegato per costruire una gerarchia di quadruple di postazioni *slave* associata ad una postazione *pivot* P_0 risulta essere costituito dai passi seguenti.

Algoritmo 1

1. Passo 1. Si selezioni una postazione *pivot* P_0 appartenente ad una delle due zone (Area Metropolitana di Roma o Valle del Sacco). Si indichi con M il numero di postazioni di monitoraggio attive in questa zona (esclusa la postazione P_0 , naturalmente);
2. Passo 2. Si determinino tutte le possibili quadruple di stazioni *slave*. Le stazioni presenti in tutte le quadruple dovranno appartenere alla stessa zona a cui appartiene P_0 ed il loro numero N_{comb} è dato dalla relazione (16);
3. Passo 3. Si consideri una delle quadruple di stazioni *slave* per la stazione *pivot* P_0 . Si calcoli sulla base delle serie storiche rilevate nell'anno 2012 il semivariogramma secondo quanto richiesto dalle relazioni (11a) e (11.b) e si risolva il sistema lineare (12) impiegando il *Singular Value Decomposition Method*. Una volta ottenuto il vettore dei pesi λ , si determini l'errore quadratico medio *MSE* o direttamente o mediante la (13). In questo passo vengono escluse tutte le giornate in cui non è operativa almeno una delle stazioni *slave* oppure la postazione *pivot*;

4. Passo 4. Si iteri il Passo 3 per tutte le N_{comb} quadruple associate alla postazione *pivot* P_0 . Per ognuna di queste combinazioni si calcoli l'errore quadratico medio relativo;
5. Passo 5. Si ordinino le N_{comb} quadruple in ordine crescente di MSE . La quadrupla col minore MSE è la quadrupla ideale da utilizzare quando lo stimatore viene applicato alla postazione *pivot* P_0 e la radice quadrata $RMSE$ dell'errore quadratico medio MSE è l'incertezza intrinseca minima dello stimatore (3) quando è applicato alla postazione *pivot* P_0 ;
6. Passo 6. Si iterano i Passi 1-5 per tutte le postazioni presenti nella zona considerata.

L'algoritmo produce quindi una gerarchia di quadruple per ogni postazione *pivot* e le stime prodotte impiegando la prima quadrupla della gerarchia saranno caratterizzate dalla minor incertezza intrinseca possibile. In sostanza, se per ogni postazione di misura di una delle due zone considerate si ignorassero le misure e si considerassero solo le stime relative, l'errore medio che si commetterebbe nel considerare la stima al posto della misura sarebbe pari al valore $RMSE$ individuato. Questo algoritmo è stato applicato a tutte le postazioni dell'Area Metropolitana di Roma e della Zona della Valle del Sacco ed in *Tab.2* e *Tab.3* vengono presentati i risultati ottenuti. Prima di commentare questi risultati conviene ricordare che essi sono stati ottenuti sulla base delle serie storiche relative all'anno 2012. Quanto fatto per l'anno 2012 potrà ovviamente essere ripetuto per gli anni futuri o per anni passati cui si sia interessati, ottenendo per ciascuno di questi periodi temporali gerarchie e pesi non necessariamente identici. La variazione annuale dei pesi (ed anche eventualmente della quadrupla ottimale) rispecchierà la variabilità annuale delle condizioni meteorologiche ed anche eventuali variazioni nelle sorgenti emmissive. Detto ciò, vale la pena fare alcune considerazioni in proposito.

Per prima cosa è interessante notare come ognuna delle quattro stazioni romane influenzate dal traffico urbano (Via Magna Grecia, Via Enrico Fermi, Viale Francia e Via Tiburtina) presentino tra le stazioni *slave* almeno una stazione di fondo (considerando la stazione di Via Boncompagni come tale). Tale comportamento potrebbe essere dovuto al fatto che lo stimatore BLUE sembra essere in grado di *leggere* gli eventi di inquinamento atmosferico come la sovrapposizione tra eventi di fondo ed eventi locali. La stazione di Via Boncompagni invece (localizzata alla sommità dell'edificio della sede di rappresentanza di Arpa Lazio al centro di Roma), viene rappresentata da tre stazioni di fondo (Castel di Guido, Tenuta del Cavaliere e Roma Bufalotta) e da una stazione da traffico (via Magna Grecia sita in un profondo canyon urbano). In questo caso lo stimatore indica chiaramente come questa postazione, localizzata al limite superiore del *canopy* urbano, in qualche modo risente sia delle condizioni di background che delle situazioni locali (anche le più estreme come quelle dei canyon urbani) miscelandole tra loro e creando così una sorta di valor medio cittadino.

Dall'analisi delle stazioni localizzate all'interno della Valle del Sacco (*Tab.2*) emergono dei valori di $RMSE$ complessivamente superiori a quanto osservato nel caso di Roma e ciò è probabilmente dovuto alla complessità orografica ed emissiva della zona. Da qui emerge la necessità di avere maggiori informazioni sulla distribuzione spaziale della concentrazione media giornaliera di PM_{10} . Non potendo però aumentare il numero delle postazioni fisse della rete regionale, sarà indispensabile impiegare mezzi differenti e di ciò si tratterà al punto 7. Ritornando a quanto presentato in *Tab.3*, emergono due situazioni interessanti. In primo luogo, la postazione di Fontechiari (località posta sul contrafforte appenninico che delimita a NE la Valle del Sacco), storicamente ritenuta come descrittiva del fondo regionale laziale, cattura pochissime informazioni locali. Questa è probabilmente la ragione per cui compare poche volte (in 2 casi su 10) tra le stazioni *slave*. Ciò potrebbe essere la dimostrazione del fatto che la Valle del Sacco è un'area sostanzialmente disaccoppiata, dal punto di vista meteorologico ed emissivo, dal resto del territorio regionale. In sostanza, questa stazione è utile nella definizione del background regionale ma per nulla rappresentativa della realtà specifica della Valle del Sacco. Al contrario, la postazione di Ferentino compare nel 70% dei casi come stazione *slave*. Alla luce della distribuzione spaziale delle varie postazioni nell'area, tale comportamento dimostra come essa sia in grado di rilevare la maggior parte delle informazioni della zona. Se si considera poi la postazione di Frosinone Scalo (posta in fondovalle ai piedi del pendio su cui si sviluppa la città di Frosinone), storicamente la *peggiore* della regione, si vede come le relative misure di concentrazione media giornaliera di PM_{10} vengano ricostruite sulla base delle analoghe misure rilevate dalle stazioni *slave* di Frosinone Via Mazzini, di Ferentino, di Ceccano e di Anagni. Questa quadrupla è estremamente interessante poiché:

- come si è visto la postazione di Ferentino riesce a riprodurre le informazioni *complessive* della Valle del Sacco: potremmo quindi ritenerla rappresentativa della situazione media della valle ed ipotizzare che essa fornisca alla serie storica di Frosinone Scalo il *fondo locale*;

- la postazione di Frosinone Via Mazzini introduce poi nella serie storica di Frosinone Scalo le peculiarità specifiche della città di Frosinone (traffico autoveicolare, riscaldamento urbano, ecc.);
- la postazione di Ceccano, essendo localizzata in una zona caratterizzata da frequenti fenomeni di ristagno delle masse d'aria e quindi di accumulo degli inquinanti, parrebbe fornire a Frosinone Scalo anche il contributo derivante da questa situazione di blocco delle masse d'aria;
- la postazione di Anagni infine sembra seguire la stessa modulazione temporale di Frosinone Scalo ma pesantemente *deamplificata*. Quindi la postazione di Anagni sembrerebbe fornire le *regole* con cui deve variare di giorno in giorno la concentrazione media giornaliera di PM_{10} .

Alla luce dell'analisi effettuate si evidenzia come l'area di Roma sia compiutamente descritta dalla rete di monitoraggio della qualità dell'aria, mentre l'area della Valle del Sacco evidenzia una complessità tale da richiedere un maggiore apporto informativo per una sua descrizione realistica, motivando così l'impiego dello stimatore lineare combinato con campagne realizzate dai mezzi mobili come vedremo al punto 7.

La procedura sopra delineata ha consentito di associare ad ogni postazione *pivot* una gerarchia di quadruple *slave* a *RMSE* crescente, quindi ad ottimalità decrescente. Stabilito ciò, per verificare la *bontà* della stima, per ogni postazione P_0 della rete di monitoraggio (di cui, come detto, si dispone per l'anno 2012 della serie storica delle misure medie giornaliere di PM_{10} , pur con inevitabili valori mancanti), si applica ogni giorno lo stimatore BLUE ottenendo quindi l'analoga serie storica delle stime. Dal confronto statistico delle serie storiche misurate e stimate si ottiene immediatamente il grado di bontà effettivo dello stimatore. Più precisamente si opera secondo l'algoritmo seguente:

Algoritmo 2

1. Passo 1. Si selezioni una generica postazione *pivot* P_0 appartenente ad una delle due porzioni territoriali considerate. A tale postazione l'Algoritmo 1 ha associato una gerarchia a *RMSE* crescente di quadruple di postazioni *slave* cui è associata la relativa quadrupla di pesi;
2. Passo 2. Si consideri un giorno k dell'anno. Se per il giorno considerato non esiste la misura della concentrazione media giornaliera di PM_{10} a P_0 , si passi al giorno successivo $k+1$. Se invece esiste, si consideri la quadrupla a *RMSE* migliore. Se per essa e per il giorno k sono disponibili le misure di concentrazione media giornaliera di PM_{10} in tutte le postazioni *slave*, si determini la stima della concentrazione in P_0 ed il relativo errore quadratico impiegando lo stimatore lineare. Se invece almeno una delle postazioni *slave* non dispone di misura, si passa alle quadruple successive della gerarchia fino ad individuare quella quadrupla per cui sono disponibili tutte le misure. Una volta individuata tale quadrupla, si determini la stima in P_0 ed il relativo errore quadratico.
3. Passo 3. Si iteri il passo precedente per tutti i giorni dell'anno considerato;
4. Passo 4. Sulla base degli errori quadratici giornalieri ottenuti col passo precedente, si calcoli il valore di *MSE* e di *RMSE*. Il valore di *RMSE* individuato può essere considerata l'incertezza intrinseca dello stimatore per l'anno considerato e per la postazione *pivot* selezionata;
5. Passo 5. Si iteri i passi precedenti per tutte le possibili postazioni *pivot*.

Nelle Tab. 4 (per l'area Metropolitana di Roma) e Tab.5 (per la Zona della Valle del Sacco) sono stati presentati i risultati ottenuti dall'applicazione dell'algoritmo presentato relativamente all'anno 2012. In esse, per ogni postazione sono riportate le informazioni seguenti:

- valor medio annuo della concentrazione media giornaliera di PM_{10} misurato e calcolato con lo stimatore lineare (3);
- numero annuo di superamenti del valore di $50 \mu g/m^3$ della concentrazione media giornaliera misurata e calcolata che è uno dei limiti previsti dalla normativa vigente per la concentrazione media giornaliera di PM_{10} ;
- valore del parametro *RMSE*, dell'indice di concordanza di Lin e del coefficiente di correlazione di Pearson.

Come si può notare, le stime del valor medio annuo della concentrazione giornaliera di PM_{10} e del numero dei superamenti del valore di $50 \mu g/m^3$ sono decisamente prossime ai valori misurati. Se poi si considera l'incertezza intrinseca allo stimatore lineare (*RMSE*), si può notare come il modello lineare riesca a riprodurre le concentrazioni medie giornaliere di PM_{10} con un'incertezza intrinseca complessiva (*RMSE*) che,

nell'Area Metropolitana di Roma, varia tra $3 \mu\text{g}/\text{m}^3$ (Bufalotta) e $5 \mu\text{g}/\text{m}^3$ (Francia) ad eccezione di Ciampino dove si osserva un valore pari a $8 \mu\text{g}/\text{m}^3$. Nella Zona della valle del Sacco lo stimatore produce dei valori di RMSE superiori a quelli osservati nell'area di Roma e compresi tra $5 \mu\text{g}/\text{m}^3$, a Colleferro-Europa, e $14 \mu\text{g}/\text{m}^3$ a Cassino. L'indicatore RMSE è estremamente pessimistico: più avanti vedremo come, in effetti, l'errore percentuale medio rispetto al valore stimato sia decisamente inferiore.

pivot	Ada	Aren.	Bonc.	Bufal.	Caval.	Ciamp.	Cine.	Cipro	Fermi	Francia	Grecia	Guido	Malagrotta	Preneste	Tibur.	Guidonia	RMSE
Ada				0.2932				0.4865	-0.2171			0.4374					3.42
Arenula							0.1098		0.4941			0.2598				0.1363	3.22
Boncompagni				0.1459	0.1235						0.3352	0.3954					3.13
Bufalotta	0.3643		0.3121		0.1209										0.2028		3.07
Cavaliere				0.3485		0.1208						0.1768				0.3539	3.62
Ciampino			-0.0163		0.4539							0.0363			0.5260		7.12
Cinecittà			0.2762									-0.1464		0.4974	0.3728		4.63
Cipro	0.4571		-0.0379								0.3453		0.2356				3.47
Fermi	-0.2008	0.5848									0.5906	0.0254					3.79
Francia		0.3784									0.4680	-0.2357			0.3892		4.39
Grecia			0.2365					0.2410	0.2534						0.2691		2.89
Guido	0.4964		0.7130								-0.4350		0.2256				4.34
Malagrotta			0.1142				0.1815	0.5323				0.1720					3.82
Preneste			0.1661				0.2594					0.0645			0.5100		4.03
Tiburtina						0.1562				0.4066		-0.1906		0.6278			4.59
Guidonia				-0.1556	0.6126			0.4647				0.0783					4.23

Tab.2: per ogni stazione pivot dell'area Metropolitana di Roma (righe) è indicata la quadrupla di stazioni slave ottimale (colonne), i relativi pesi e l'incertezza intrinseca minima

pivot	Alatri	Anagni	Cassino	Ceccano	Coll. Europa	Coll. Oberdan	Ferentino	Fontechiari	FR-Mazzini	FR-scalo	RMSE
Alatri		0.4045		0.229		6.20E-02	0.3046				7.37
Anagni					0.477		0.6686		0.2365	-0.382	5.82
Cassino						0.3936	2.30E-03		0.2699	0.3341	12.68
Ceccano	0.3338				0.4401			-0.3681		0.5942	10.11
Coll. Eur.	4.46E-02			0.1589		0.6616	0.1349				5.14
Coll. Obe.					0.6623		2.72E-02		0.4556	-0.1451	5.17
Ferentino	0.1224	0.6318		-0.1227						0.3685	5.99
Fontechiari					0.1594	0.4933			0.8465	-0.4992	9.23
FR-Mazzini	0.1989		0.2005				0.3427	0.258			6.23
FR-scalo		-1.3634		0.5787			0.8888		0.8958		10.33

Tab.3: per ogni stazione pivot della zona della valle del Sacco (righe) è indicata la quadrupla di stazioni slave ottimale (colonne), i relativi pesi e l'incertezza intrinseca minima

2012	media annua		numero di superamenti		statistica		
stazione	misurato	calcolato	misurato	calcolato	R	Lin	RMSE
Ada	24	24	9	6	0.94	0.94	3.58
Arenula	30	30	18	17	0.95	0.95	3.42
Boncompagni	28	27	12	11	0.95	0.95	3.34
Bufalotta	28	28	16	15	0.96	0.96	3.11
Cavaliere	28	27	12	13	0.93	0.93	3.98
Ciampino	32	32	43	37	0.85	0.84	8.41
Cinecittà	35	35	53	43	0.95	0.94	5.01
Cipro	27	27	19	17	0.96	0.96	3.6
Fermi	33	32	30	31	0.94	0.94	4.17
Francia	36	35	57	42	0.93	0.93	5.19
Grecia	32	31	29	30	0.97	0.97	3.12
Guido	24	24	5	8	0.89	0.89	4.34
Malagrotta	28	28	23	20	0.94	0.94	4.2
Preneste	34	34	45	39	0.96	0.96	4.46
Tiburtina	37	37	50	56	0.96	0.96	4.61
Guidonia	27	27	15	11	0.92	0.92	4.42

Tab. 4: performances tipiche dello stimatore lineare (3) applicato all'Area Metropolitana di Roma

2012	media annua		numero di superamenti		statistica		
stazione	misurato	calcolato	misurato	calcolato	R	Lin	RMSE
Alatri	36	35	77	67	0.92	0.92	7.87
Anagni	29	28	28	26	0.89	0.89	6.19
Cassino	39	37	75	78	0.85	0.84	13.88
Ceccano	48	49	118	125	0.96	0.96	10.96
Coll. Eur.	34	34	54	60	0.95	0.95	5.47
Coll. Obe.	30	30	28	41	0.92	0.92	5.68
Ferentino	34	35	63	67	0.95	0.95	6.26
Fontechiari	19	22	2	3	0.58	0.54	9.45
FR-Mazzini	31	31	57	55	0.93	0.93	6.49
FR-scalo	49	48	120	118	0.96	0.96	11.35

Tab. 5: performances tipiche dello stimatore lineare (3) applicato alla Zona della Valle del Sacco

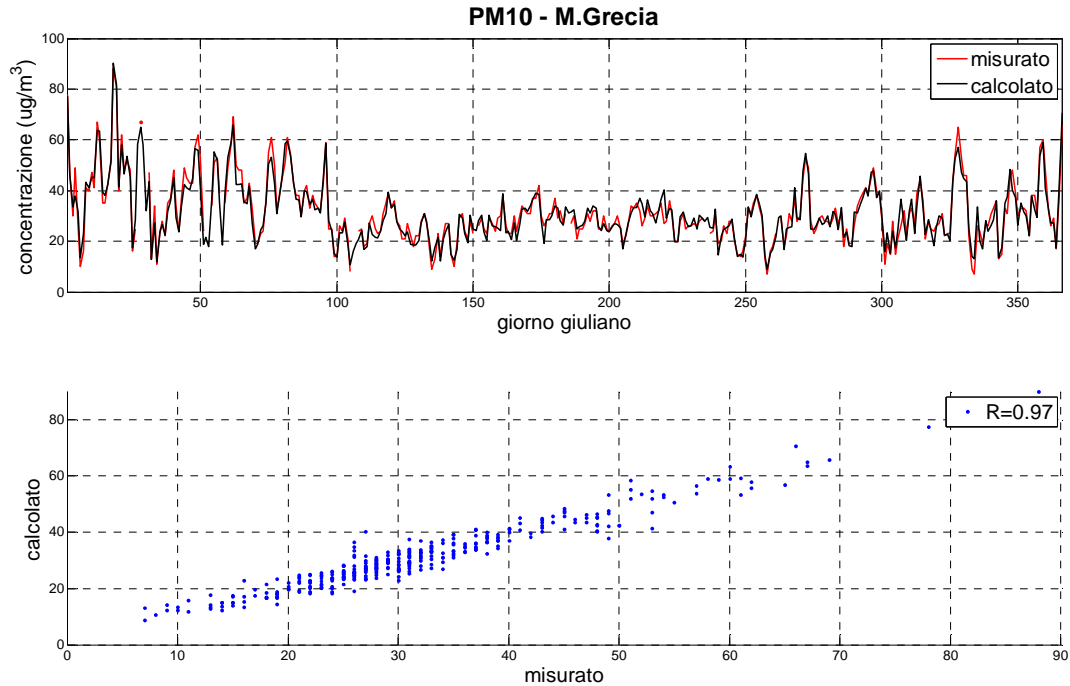


Fig. 5a: confronto tra stime giornaliere e valori misurati a Roma via Magna Grecia

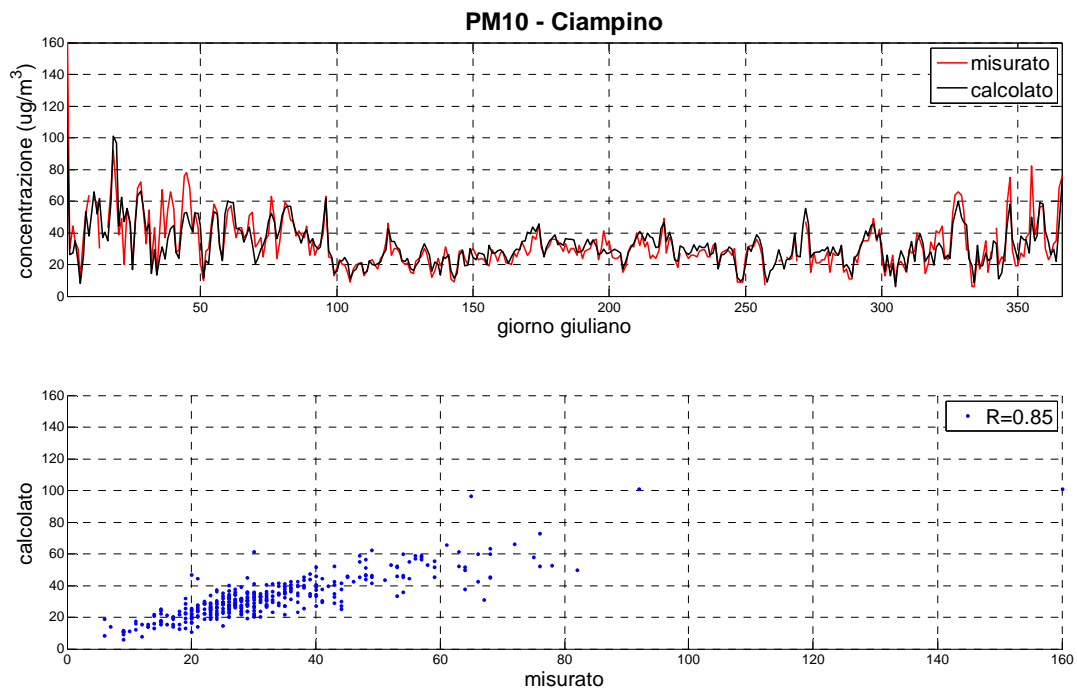


Fig. 5b: confronto tra stime giornaliere e valori misurati a Ciampino

Per meglio apprezzare la bontà delle stime operate dallo stimatore lineare (3), è conveniente considerare gli andamenti giornalieri delle stime di PM_{10} confrontate con i rispettivi valori misurati:

- alla postazione di Roma Magna Grecia che è la postazione dell'Agglomerato di Roma per cui il valore di $RMSE$ è minimo (Fig. 5a) e che è quella postazione per cui le stime del modello BLUE più si avvicinano ai valori misurati;
- alla postazione di Ciampino che è la postazione dell'Agglomerato di Roma per cui il valore di $RMSE$ è massimo (Fig. 5b), quindi quella in cui lo stimatore BLUE presenta le *performances* peggiori;
- alla postazione di Colleferro via Europa che è la postazione della Zona Valle del Sacco per cui il va-

- lore di $RMSE$ è minimo (*Fig. 5c*)
- alla postazione di Cassino che è la postazione della Zona Valle del Sacco per cui il valore di $RMSE$ è massima (*Fig. 5d*).

Come si nota immediatamente, lo stimatore lineare descrive con eccellente realismo le serie storiche misurate sia nelle postazioni con basso $RMSE$ che in quelle ad elevato $RMSE$.

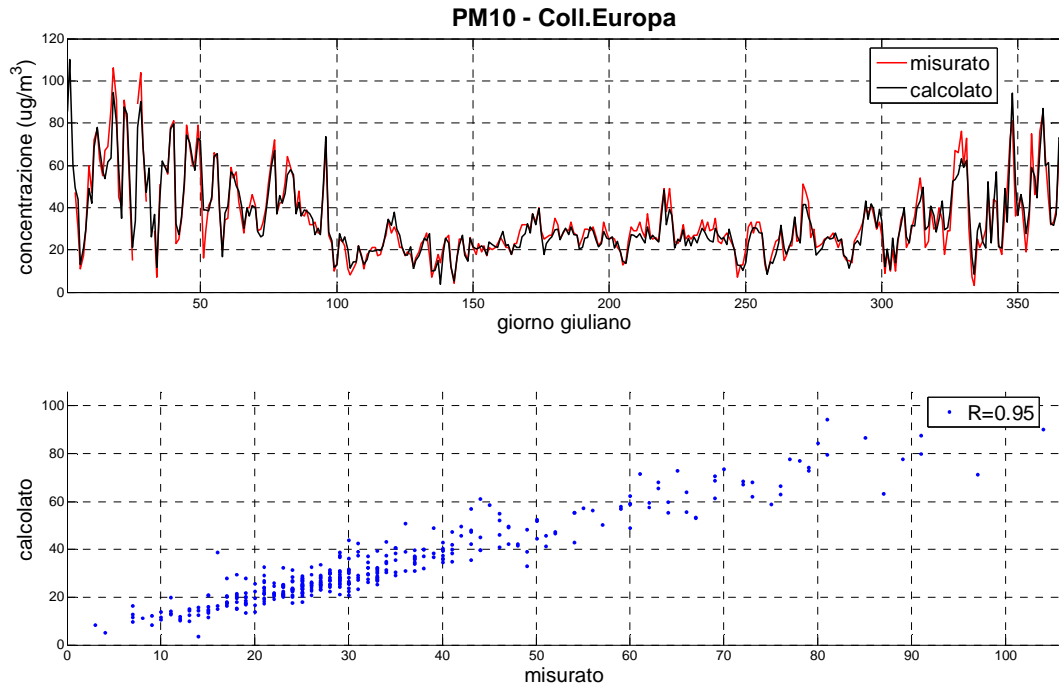


Fig. 5c: confronto tra stime giornaliere e valori misurati a Colleferro via Europa

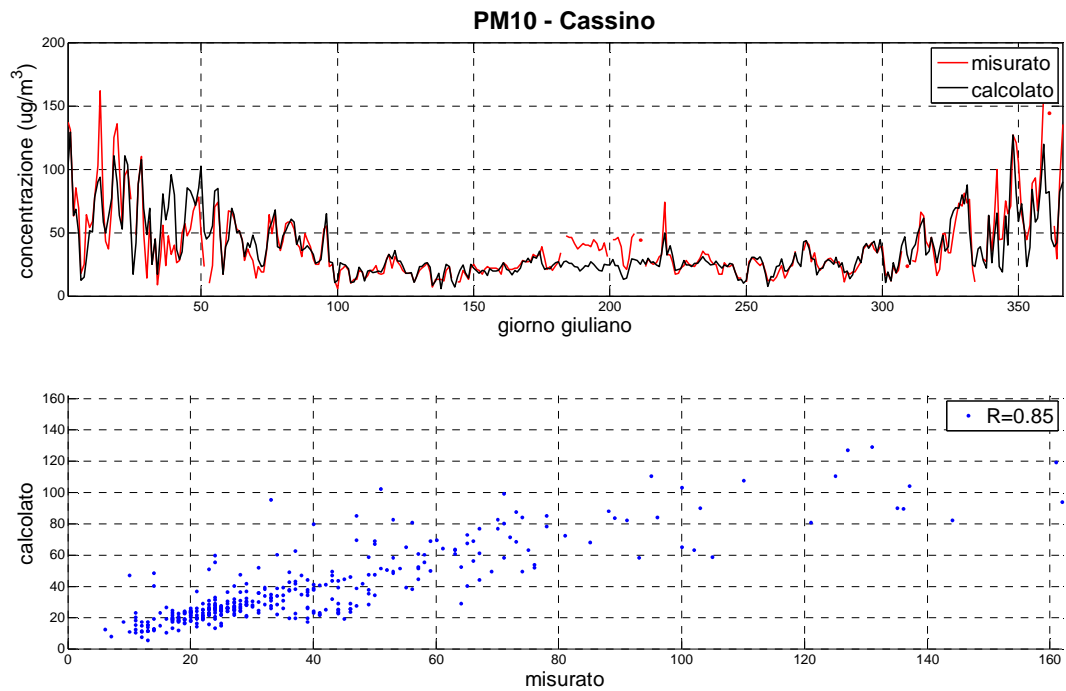


Fig. 5d: confronto tra stime giornaliere e valori misurati a Cassino

Per concludere la caratterizzazione dello stimatore lineare (3), è utile valutare alcuni indici statistici ed alcuni indici di *performance* frequentemente impiegati nella valutazione del comportamento predittivo di modelli e stimatori.

stazione		MB	RMSE	NME	NMB
Roma	Preneste	-0.01	4.47	8.01	-0.04
	Francia	1.39	5.20	8.81	3.81
	M.Grecia	0.35	3.12	5.96	1.11
	Cinecitta	0.32	6.62	9.67	0.91
	Guidonia	0.11	5.30	9.78	0.39
	Ada	-0.39	3.59	9.32	-1.61
	Guido	-0.46	4.38	12.07	-1.95
	Cavaliere	0.36	3.98	8.65	1.28
	Ciampino	-0.04	8.42	15.69	-0.13
	Fermi	0.60	4.18	7.68	1.83
	Bufalotta	-0.51	3.36	7.00	-1.86
	Cipro	-0.53	3.60	8.85	-1.95
	Tiburtina	0.35	4.62	7.97	0.95
	Arenula	0.12	3.43	6.62	0.41
	Malagrotta	0.06	4.21	9.16	0.21
	Boncompagni	0.34	3.34	7.29	1.21
Valle del Sacco	Coll.Oberdan	-0.21	5.72	12.06	-0.69
	Coll.Europa	0.37	5.48	10.05	1.09
	Ceccano	-0.58	10.98	14.18	-1.19
	Ferentino	-0.16	6.27	12.12	-0.45
	Anagni	0.49	6.20	13.48	1.68
	Alatri	0.45	7.88	13.07	1.26
	Fontechiari	-3.46	9.46	37.27	-18.25
	Cassino	1.26	13.90	20.73	3.26
	FR-scalo	2.03	11.37	15.33	4.09
	FR-Mazzini	-0.32	6.50	12.16	-1.03

Tab. 6: valore degli indici statistici

Gli indici statistici, che si incaricano di quantificare secondo criteri differenti l'incertezza intrinseca allo stimatore BLUE sono numerosi e quelli considerati qui di seguito sono i seguenti:

- ✓ Incetezza intrinseca (RMSE) definito come:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_m^i - C_o^i)^2}$$

- ✓ Errore sistematico medio (MB) definito come:

$$MB = \frac{1}{N} \sum_{i=1}^N (C_m^i - C_o^i)$$

- ✓ Errore assoluto normalizzato (NME) definito come:

$$NME = \frac{\sum_{i=1}^N |C_m^i - C_o^i|}{\sum_{i=1}^N C_o^i} 100\%$$

- ✓ Errore sistematico normalizzato (NMB) definito come:

$$NMB = \frac{\sum_{i=1}^N (C_m^i - C_o^i)}{\sum_{i=1}^N C_o^i} 100\%$$

Impiegando le serie storiche rilevate nell'anno 2012, il valore assunto da questi indicatori statistici è quello riportato in *Tab. 6*.

stazione		Accuracy (%)	FAR (%)	POD (%)	CSI (%)
Roma	<i>Preneste</i>	90	10	78	71
	<i>Francia</i>	77	12	65	60
	<i>M.Grecia</i>	98	10	93	84
	<i>Cinecitta</i>	78	14	70	63
	<i>Guidonia</i>	94	27	53	44
	<i>Ada</i>	100	0	67	67
	<i>Guido</i>	99	50	80	44
	<i>Cavaliere</i>	97	31	75	56
	<i>Ciampino</i>	72	24	65	54
	<i>Fermi</i>	89	23	80	65
	<i>Bufalotta</i>	98	13	81	72
	<i>Cipro</i>	98	12	79	71
	<i>Tiburtina</i>	83	21	88	71
	<i>Arenula</i>	89	35	61	46
	<i>Malagrotta</i>	90	25	65	54
	<i>Boncompagni</i>	100	0	92	92
Valle del Sacco	<i>Coll.Oberdan</i>	88	39	89	57
	<i>Coll.Europa</i>	79	22	87	70
	<i>Ceccano</i>	74	12	93	83
	<i>Ferentino</i>	82	16	89	76
	<i>Anagni</i>	85	27	68	54
	<i>Alatri</i>	78	9	79	73
	<i>Fontechiari</i>	98	100	0	0
	<i>Cassino</i>	69	18	85	72
	<i>FR.Scalo</i>	71	9	89	82
	<i>FR-Mazzini</i>	71	20	77	65

Tab. 7: valore degli indicatori di performance

Per valutare la bontà operativa dello stimatore lineare (3) ed in particolare per verificare la capacità del modello di ricostruire gli eventi critici rilevati dalla rete di monitoraggio di qualità dell'aria, sono stati impiegati i seguenti indicatori di performance frequentemente utilizzati per valutare la capacità predittiva di modelli sia deterministiche che statistici:

- ✓ Accuratezza: percentuale di ricostruzioni corrette

$$Accuracy (\%) = 100 \frac{A + D}{A + B + C + D}$$

- ✓ Tasso di falso allarme (FAR): percentuale di superamenti del valore di $50 \mu\text{g}/\text{m}^3$ calcolati che non si sono verificati nella realtà.

$$FAR (\%) = 100 \frac{B}{B + D}$$

- ✓ Probabilità di corretta ricostruzione (POD): percentuale di superamenti osservati che vengono calcolati correttamente.

$$POD (\%) = 100 \frac{D}{C + D}$$

- ✓ Indice di successo di eventi critici (CSI): percentuale di corretta ricostruzione di eventi critici al netto delle situazioni di non superamento.

$$CSI (\%) = 100 \frac{D}{B + C + D}$$

in cui le variabili presenti hanno il significato seguente:

- A: numero di casi in cui il modello calcola correttamente un non-superamento;
- B: numero di casi in cui il modello calcola un superamento che non avviene nella realtà;
- C: numero di casi in cui il modello non calcola un superamento che invece avviene nella realtà;
- D: numero di casi in cui il modello calcola un superamento misurato nella realtà.

L'applicazione dello stimatore lineare all'anno 2012 ha condotto alla situazione illustrata nella *Tab. 7*.

5. Supporto alla validazione delle misure della rete

Il tipico analizzatore di PM_{10} che opera all'interno delle postazioni di monitoraggio fisse della rete regionale della qualità dell'aria è sostanzialmente un apparato che opera nel modo seguente:

- aspira l'aria ambiente dall'esterno della centralina dopo averla privata del contenuto di polvere e/o aerosol a granulometria superiore a $10 \mu m$. Ciò si ottiene ponendo un'opportuna testa a monte del condotto di aspirazione;
- convoglia l'aria aspirata sopra un filtro (o un nastro) che trattiene il particolato contenuto nell'aria stessa in maniera permanente;
- misura, opo 24 ore di aspirazione, l quantitativo totale di particolato raccolto sul filtro. Ciò normalmente si realizza impiegando una sorgente radioattiva di ^{14}C ;
- calcola la concentrazione media giornaliera di PM_{10} sulla base della quantità in peso raccolta sul filtro e sulla base della portata d'aria aspirata durante l'intero periodo di campionamento.

Una volta calcolata la concentrazione media giornaliera di PM_{10} , l'elettronica dell'analizzatore la invia al centro di raccolta dati e la rende disponibile. La misura così ottenuta, prima di essere effettivamente archiviata ed utilizzata, deve essere *validata* dagli operatori della rete che devono verificare se il valore di concentrazione prodotto dall'analizzatore rappresenta effettivamente, oltre ogni ragionevole dubbio, la concentrazione media giornaliera di PM_{10} presente nell'aria nel punto di campionamento e ciò non è impresa da poco. In effetti, una misura di particolato potrebbe essere affetta da un errore inaccettabile o essere semplicemente il prodotto di comportamenti anomali dell'analizzatore a causa di numerosi e diversi motivi. Anche se alcune di queste situazioni anomale vengono individuate dalla diagnostica interna dell'analizzatore (che si preoccupa di segnalare come non valida la misura in questione), esiste comunque una miriade di situazioni che fanno sì che la misura prodotta dall'analizzatore, pur essendo formalmente corretta, di fatto risulti ben lontana dal valore reale. Molte possono essere le possibili anomalie ed il processo di *validazione*, realizzato manualmente dagli operatori della rete, tenta, sulla base dell'esperienza accumulata, di discriminare tra una misura accettabile (*valida*) ed una misura inaccettabile (*non valida*). E' del tutto evidente che l'operazione di validazione è una procedura quasi totalmente manuale e soggettiva. Lo sviluppo dello stimatore BLUE (3) può in questo caso fornire un prezioso contributo oggettivo al processo di validazione. Qui di seguito viene proposta, come prima applicazione pratica dello stimatore, una metodologia che lo vede come un supporto oggettivo e sistematico alla validazione delle misure.

Si consideri a questo scopo una generica postazione P_0 dove opera in continuo un analizzatore di PM_{10} . Per essa e per le altre postazioni operative nella stessa porzione di territorio in cui si trova P_0 (Area Metropolitana di Roma o Zona della Valle del Sacco) sono ovviamente disponibili le serie storiche rilevate e validate. Per fissare le idee, consideriamo P_0 che opera in un giorno k del presente anno (2013) e per P_0 e per le altre postazioni sono disponibili le serie storiche dell'intero anno 2012 e dei giorni fino al giorno $k-1$ dell'anno 2013. Operando secondo quanto riportato al punto 4, sulla base delle misure rilevate nell'anno 2012 viene associata una gerarchia di quadruple di postazioni *slave* con i rispettivi pesi λ_i da utilizzare nello

stimatore (3). Pertanto, per ogni giorno k dell'anno 2013 lo stimatore BLUE (3) fornisce per P_0 una stima che è basata sulla statistica più recente (quella del 2012) della distribuzione quotidiana di PM_{10} sul territorio. Se la distribuzione annuale delle emissioni e delle caratteristiche meteorologiche e dispersive dell'atmosfera non cambia drasticamente da un anno all'altro (cosa ben poco probabile), l'interdipendenza statistica tra le concentrazioni di PM_{10} dei diversi punti del territorio (rappresentata concretamente dalla specifica gerarchia di quadruple associate a P_0 e dai pesi relativi) è ragionevole che si mantenga pressoché inalterata. Quindi è ragionevole ritenere che le stime in P_0 relative al 2013 realizzate dallo stimatore (3) con la gerarchia di quadruple ed i rispettivi pesi determinati sulla base delle serie storiche del 2012 siano decisamente accurate. Partendo da questa ragionevole ipotesi di lavoro, una possibile procedura di supporto alla validazione delle misure prodotte dall'analizzatore operativo in P_0 potrebbe essere articolata nei passi seguenti:

Algoritmo 3

- Passo 1. Si scelga una postazione P_0 per cui risulta disponibile la misura della concentrazione media giornaliera di PM_{10} $C_0(k)$ relativa al giorno k strumentalmente valida (per la quale, cioè, la diagnostica dell'analizzatore non ha riscontrato alcun tipo di anomalia);
- Passo 2. Si stimi la concentrazione media di PM_{10} $C_{0s}(k)$ del giorno k impiegando lo stimatore BLUE (3) in cui la gerarchia delle quadruple di postazioni *slave* ed i relativi pesi λ_i sono stati determinati sulla base delle serie storiche dell'anno 2012;
- Passo 3. si individui un intervallo di concentrazione con l'estremo superiore pari a $C_{0s}(k) + \Delta$ e l'estremo inferiore sia pari a $C_{0s}(k) - \Delta$, entro cui è ragionevole che si collochi il valore *vero* della concentrazione con un dato valore di confidenza;
- Passo 4. Se $C_0(k)$ risulta interno all'intervallo precedente, allora $C_0(k)$ può essere ritenuto valido; se invece tale valore si colloca all'esterno di questo intervallo, $C_0(k)$ viene ritenuta una *misura dubbia* e devono essere attivate da parte degli operatori della rete controlli specifici in proposito;
- Passo 5. Si ripetono tutti i passi precedenti per tutte le postazioni della rete di monitoraggio;
- Passo 6. Si ripetono tutti i passi precedenti per ogni giorno k dell'anno 2013
- Passo 7. Nel 2014 si ripeterà l'intera procedura dopo aver ricostruito, sulla base delle serie storiche valide rilevate nell'anno 2013, la gerarchia di quadruple per ogni postazione di rete (secondo la metodologia riportata al punto 4) ed i relativi pesi.

Come è immediato constatare, l'intera procedura, totalmente oggettiva, è estremamente semplice ed è facilmente automatizzabile nel software del Centro Regionale della Qualità dell'Aria, diminuendo di molto la complessità e la soggettività della procedura di validazione quotidiana delle misure di PM_{10} .

L'unico elemento ancora da determinare è l'ampiezza dell'intervallo Δ . Per giungere a ciò, consideriamo una generica postazione P_0 . Per essa nell'anno 2012 (e pure negli anni precedenti) risulta disponibile sia la serie storica delle misure validate di concentrazione media giornaliera di PM_{10} che le rispettive stime realizzate con lo stimatore (3). Pertanto, per ogni giorno k dell'anno è immediato individuare lo *scarto relativo* ε_k tra misura e stima, definito come:

$$\varepsilon_k = \frac{C_0(k) - C_{0s}(k)}{C_{0s}(k)} \quad [17]$$

che è quindi una realizzazione di una variabile stocastica avente media nulla, visto che lo stimatore (3) è per definizione *unbiased*, e che potrà assumere valori sia negativi che positivi. La densità di probabilità di questa variabile stocastica potrebbe essere qualsiasi, ma visti i vincoli sopra elencati, una possibile scelta potrebbe essere una densità di probabilità di tipo gaussiano. In effetti, come si può vedere nella Fig.6, i dati relativi all'anno 2012 e a tutta la rete confermano sostanzialmente questa ipotesi di lavoro. Come detto, il valore medio annuo dello scarto relativo risulta nullo per costruzione, mentre la relativa varianza risulta pari a:

$$\sigma_\varepsilon^2 = \frac{1}{N-1} \sum_{k=1}^N \varepsilon_k^2 \quad [18]$$

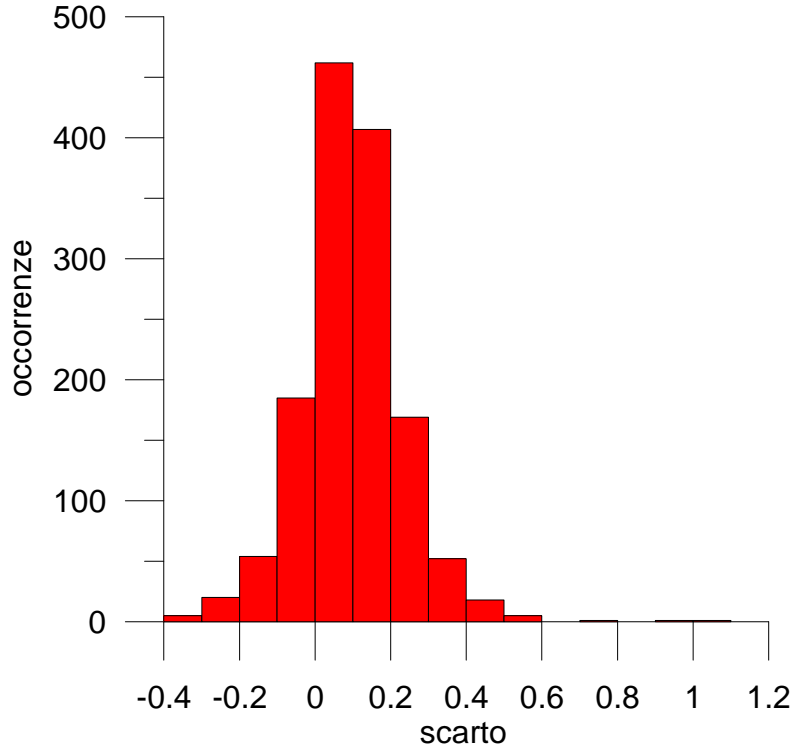


Fig.6: istogramma dello scarto relativo per alcune postazioni della rete di monitoraggio nell'anno 1012

Come è noto [28], per una distribuzione gaussiana l'intervallo di scarti relativi entro cui si collocherà con una prefissata probabilità P uno specifico scarto incontrato nel giorno k è quell'intervallo centrato attorno allo zero e con estremo superiore ed inferiore rispettivamente pari a $\pm z^* \sigma_\varepsilon$ dove, come noto, z^* dipende dal livello di confidenza P considerato (Tab. 8).

Probabilità (%)	z^*
90.0	1.645
95.0	1.960
99.0	2.576
99.7	3.000

Tab. 8: valore assunto dal parametro z^* di una distribuzione gaussiana

Pertanto, il confronto richiesto dal Passo 3 dell'algoritmo di validazione potrebbe essere fatto direttamente sullo scarto relativo rilevato, comunque operativamente risulta più semplice utilizzare direttamente nel confronto la misura $C_0(k)$. In questo caso è immediato dimostrare che se si assume un livello di confidenza P (selezionando quindi il valore appropriato di z^*), l'ampiezza Δ dell'intervallo entro cui il valore misurato dovrà stare per essere ritenuto valido sarà il seguente:

$$\pm \Delta = \pm C_{0s}(k) \cdot z^* \sigma_\varepsilon \quad [19]$$

Come si nota, l'ampiezza dell'intervallo varia in maniera direttamente proporzionale al valore della stima $C_{0s}(k)$. Ciò evidenzia come il parametro $RMSE$ sia molto pessimistico, infatti l'ampiezza dell'intervallo attorno al valore stimato entro cui si va a collocare la misura con probabilità P non è pari a $RMSE$ ma è proporzionale alla stima stessa, quindi è piccolo per stime piccole e ampie per valori elevati.

Per dimostrare l'applicabilità pratica della procedura di Quality Control, essa è stata applicata alle serie storiche di concentrazione media giornaliera di PM_{10} rilevate dalle postazioni della rete di monitoraggio. In particolare, si è simulato di applicare questa metodologia di confronto giorno per giorno per tutto l'anno 2012 impiegando per lo stimatore BLUE i pesi e le gerarchie di quadruple ottenute dall'analisi delle serie storiche relative al 2011. Il risultato è stato eccellente, come si può notare in Fig.7 dove è riportato il confronto tra le misure (punti rossi) e la banda di accettabilità (curve nere) individuata dal metodo di validazione, con un livello di confidenza pari al 95%, per la postazione di Magna Grecia.

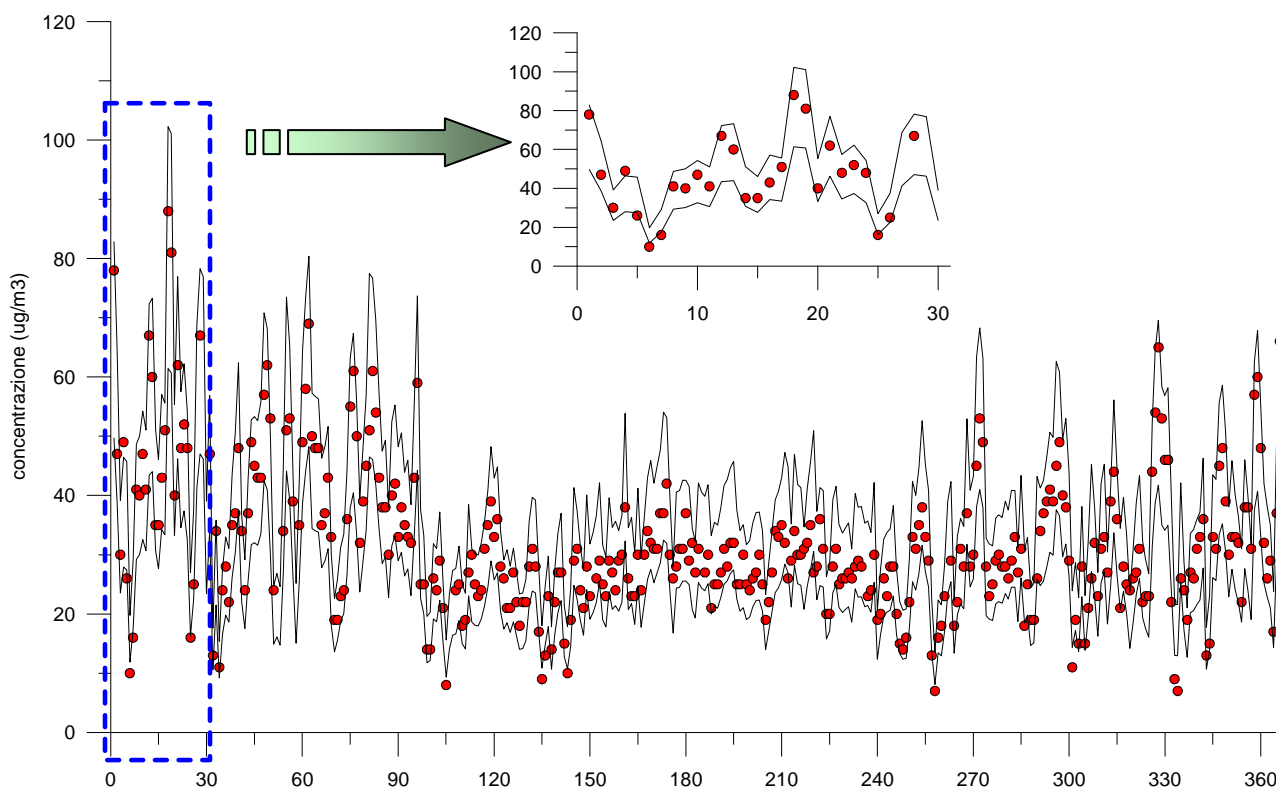


Fig. 7: applicazione della procedura di Quality Control alla postazione di Magna Grecia nell'anno 2012.

6. Gap filling

In ogni rete di monitoraggio della qualità dell'aria le serie storiche del PM_{10} (ed anche degli altri inquinanti) non sono praticamente mai complete ma, al contrario presentano *dati mancanti* più o meno numerosi a seconda del tipo di analizzatore che produce la serie storica ed al tipo di manutenzione messa in atto. La mancanza di dati, e quindi il *rendimento della rete di monitoraggio*, a parità di analizzatore e di tipo di manutenzione varia in generale di anno in anno introducendo una mancanza di informazione nel processo di valutazione della qualità dell'aria richiesto dalla normativa vigente.

La causa della mancanza di dati è il risultato di malfunzionamenti che si riscontrano negli analizzatori, malfunzionamenti che possono essere di vario genere. Spesso sono guasti di tipo meccanico o elettrico, a volte sono il risultato di problemi al software di gestione del flusso di dati, altre volte sono problemi legati ad eventi esterni che influiscono negativamente sulla misura (per esempio la temperatura dell'ambiente in cui opera l'analizzatore e quindi avarie al sistema di condizionamento della cabina di monitoraggio), ed infine a volte sono semplicemente di origine incognita. Il verificarsi di queste situazioni, che comunque possono verificarsi nonostante un'attenta, capillare e frequente manutenzione della strumentistica sia hardware che software, produce delle discontinuità nelle serie temporali. Dal punto di vista tecnico tale problema risulta irrisolvibile in quanto la strumentazione, non essendo *ideale* e quindi *non perfetta*, è e sarà sempre soggetta ad inconvenienti di questo tipo.

Talvolta il malfunzionamento di un analizzatore per un lasso di tempo lungo produce dei cosiddetti *buchi temporali* che in qualche misura possono alterare e, in alcuni casi rendere impossibile, una corretta valutazione della qualità dell'aria anche in termini normativi (numero di superamenti, medie annue, ecc.). Questo problema comune anche ad altri tipi di reti (si pensi per esempio alle reti meteorologiche) ha reso necessario lo sviluppo di opportune metodologie di *gap filling*, cioè di metodologie di vario genere in grado di sostituire i dati mancanti di una serie storica di una data variabile con stime della stessa di cui sia nota l'incertezza intrinseca.

Limitandoci a considerare esclusivamente le serie storiche della concentrazione media giornaliera di PM_{10} , è immediato renderci conto che lo stimatore BLUE proposto può essere agevolmente impiegato per

operare il *gap filling* delle relative serie storiche. Infatti, la procedura di *gap filling* realizzata impiegando lo stimatore BLUE (3) operativamente si configura come un *corollario* della procedura delineata a supporto della validazione delle misure. Più precisamente, una volta realizzata la stima della concentrazione media giornaliera di PM_{10} per il generico giorno k , se la misura non è disponibile e non è valida, al suo posto nella serie storica viene posta la stima relativa cui si attribuisce il *flag stima*. Al termine dell'anno sarà quindi disponibile una serie storica completa per ogni postazione della rete da cui sarà possibile *valutare* la concentrazione media annua di PM_{10} ed il numero di superamenti della soglia di $50 \mu g/m^3$, oltre al valore della concentrazione media annua e dei superi calcolati tenendo conto esclusivamente delle misure valide.

Per verificare la consistenza e l'adeguatezza della procedura di *gap filling* sono state considerate 4 postazioni della rete di monitoraggio di tipo diverso: la postazione di Roma via Boncompagni (fondo urbano), la postazione di Roma via Magna Grecia e Francia (traffico) e la postazione di Castel di Guido (fondo rurale). Di queste 4 postazioni sono state considerate le serie storiche di concentrazione media giornaliera di PM_{10} rilevate nell'anno 2012 da cui sono stati rimossi tutti i giorni in cui la misura di particolato non era disponibile. In questo modo le tre serie storiche risultano artificialmente complete.

Prima di procedere alla verifica della procedura, è stata realizzata una statistica dei malfunzionamenti riscontrati nell'anno 2012 in tutti gli analizzatori PM_{10} presenti nelle postazioni della rete presenti nell'Area Metropolitana di Roma e nella Valle del Sacco. In pratica è stata registrata la durata (cioè i giorni consecutivi) per cui la misura di concentrazione di particolato risultava indisponibile nelle diverse postazioni di monitoraggio. Complessivamente, nell'anno 2012 sono stati registrati in totale una media di 21 giorni di indisponibilità di dati per analizzatore, pari a circa il 6 % di dati teorici. Nella Fig.8 è presentato un istogramma nella cui ascissa sono riportati i giorni consecutivi per cui la misura di concentrazione era indisponibile mentre in ordinata era riportato il numero di occorrenze di ciò. Sempre nella Fig.8 è presentata la rispettiva distribuzione cumulata. Come si può vedere, la distribuzione dei guasti non è simmetrica visto che non può mai avere valori negativi e presenta una notevole somiglianza con una distribuzione geometrica. Analizzando la figura, si vede come per circa il 70% dei casi la mancanza di dati è pari ad un giorno, confermando come la manutenzione correttiva della rete sia tempestiva ed efficace. Solo un 20% dei casi presenta una durata di due giorni consecutivi, mentre durate superiori sono decisamente meno probabili. Di fatto l'istogramma presentato in figura e la sua cumulata costituiscono il modello statistico di guasto tipico della rete di monitoraggio nell'anno 2012.

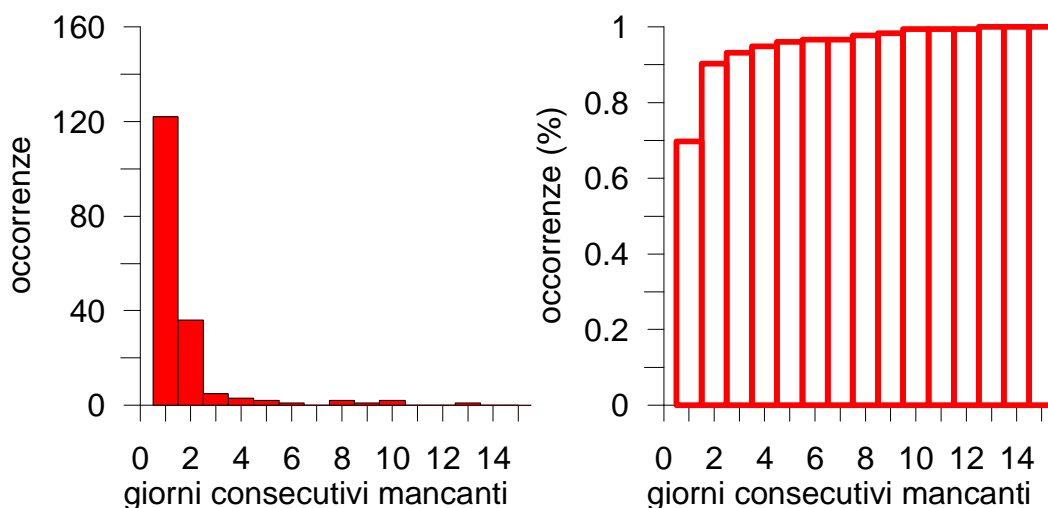


Fig.8: istogramma della frequenza della durata dell'indisponibilità delle misure (a sinistra) e la rispettiva cumulata (a destra)

Per verificare le *performances* della procedura di *gap-filling* descritta, è necessario fruttare le serie storiche note e mettere a punto una procedura di simulazione basata su di esse. Più in particolare, si è messa a punto una procedura di simulazione di tipo Monte Carlo che consente, a partire dalla serie storica (artificialmente completa) di una delle quattro postazioni prese a riferimento, di ricostruire delle serie (tutte differenti tra loro) affette da malfunzionamenti la cui statistica è del tutto simile a quella effettivamente riscontrata nelle serie storiche 2012 e riprodotta negli istogrammi di Fig. 8.

Operativamente, si consideri una qualsiasi delle 4 postazioni considerate. La procedura di simulazione Monte Carlo può essere scomposta sinteticamente in due macro-passi operativi: il primo passo si incarica di

generare M serie storiche distinte, affette da malfunzionamenti distribuiti casualmente nell'anno secondo il modello statistico di Fig.8, ed il secondo passo si incarica di calcolare il valore medio annuo di concentrazione di particolato ed il numero di superamenti della soglia di $50 \mu\text{g}/\text{m}^3$ per ciascuna delle serie artificiali ottenute. Mentre il secondo macro-passo è ovvio, è opportuno descrivere nel dettaglio il primo macro-passo della procedura che consiste dei passi elementari seguenti:

- **Passo 1.** Si consideri il giorno k (per il primo giorno considerato $k = 1$). Si estragga da un generatore di numeri casuali $U(0,1)$ distribuiti uniformemente tra 0 e 1 un valore a ;
- **Passo 2.** Se $a < 0.5$, allora la misura di concentrazione media giornaliera di particolato viene mantenuta. Si ritorna al Passo 1 con $k := k + 1$;
- **Passo 3.** Se, invece, $a \geq 0.5$, allora si assume che nel giorno k si sia riscontrato un malfunzionamento che ha portato all'assenza di una misura valida;
- **Passo 4.** In questo caso si deve definire per quanti giorni consecutivi è persistito il malfunzionamento. Per fare ciò, si estrae a caso da una distribuzione $U(0,1)$ un nuovo numero *random* b . Considerando la cumulata di Fig.8, si determina il numero N di giorni successivi di persistenza del malfunzionamento cui corrisponde una probabilità cumulata pari a b .
- **Passo 5.** Si rende non valida la misura di concentrazione del particolato per il giorno k e per i successivi $N-1$ giorni. Si pone $k := k + N - 1$ e si ritorna al Passo 1 fino all'esaurimento di tutti i giorni dell'anno.

Per ciascuna delle quattro postazioni questa procedura è stata ripetuta per $M = 100$ volte ottenendo M serie storiche sintetiche distinte. Ciò è equivalsa a generare M serie storiche annuali differenti per ciascuna postazione, tutte affette da malfunzionamenti dell'analizzatore e la presenza di questi malfunzionamenti durante l'anno rispecchiava la reale statistica riscontrata dalla rete di monitoraggio nell'anno 2012. In realtà, la procedura di verifica delle capacità dello stimatore a realizzare un accurato *gap-filling* è estremamente conservativa, infatti nella procedura si è ipotizzato che la probabilità che in un dato giorno non sia disponibile la misura di PM_{10} sia pari al 50%, anche se per la rete laziale nel 2012 la probabilità era del 2%.

Nelle Figg.9 sono riportati i risultati ottenuti da una delle iterazioni estratte casualmente per le quattro stazioni di monitoraggio considerate. Come si può vedere, la procedura di *gap filling* proposta risulta decisamente adeguata allo scopo.

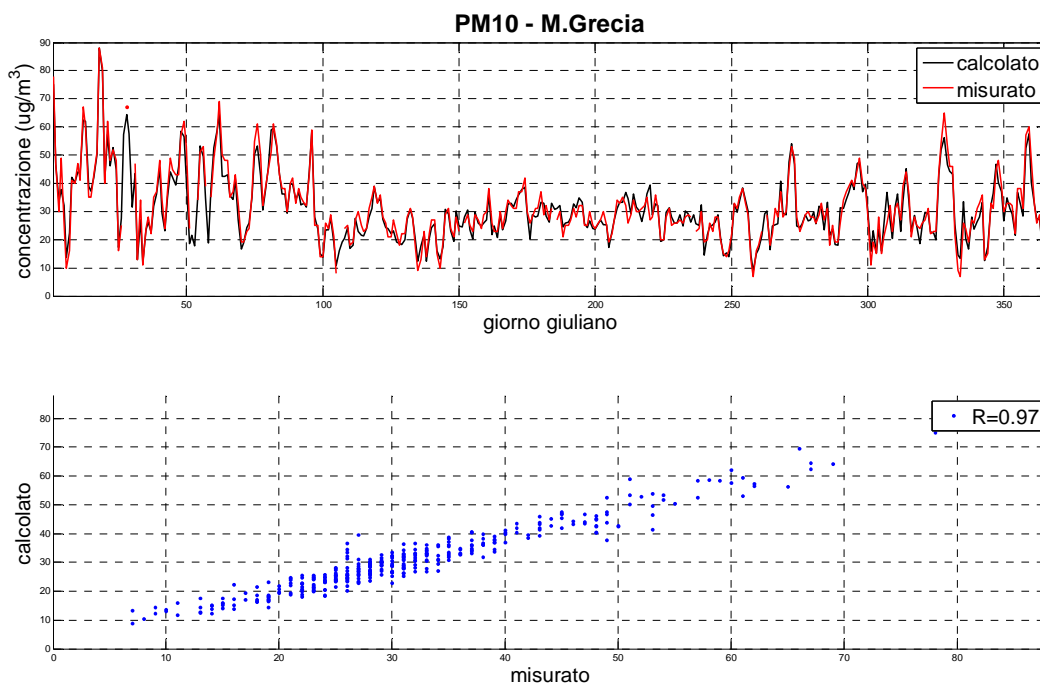


Fig. 9a: confronto tra i valori di una delle serie decimate casualmente ed i valori misurati a Roma-Magna Grecia

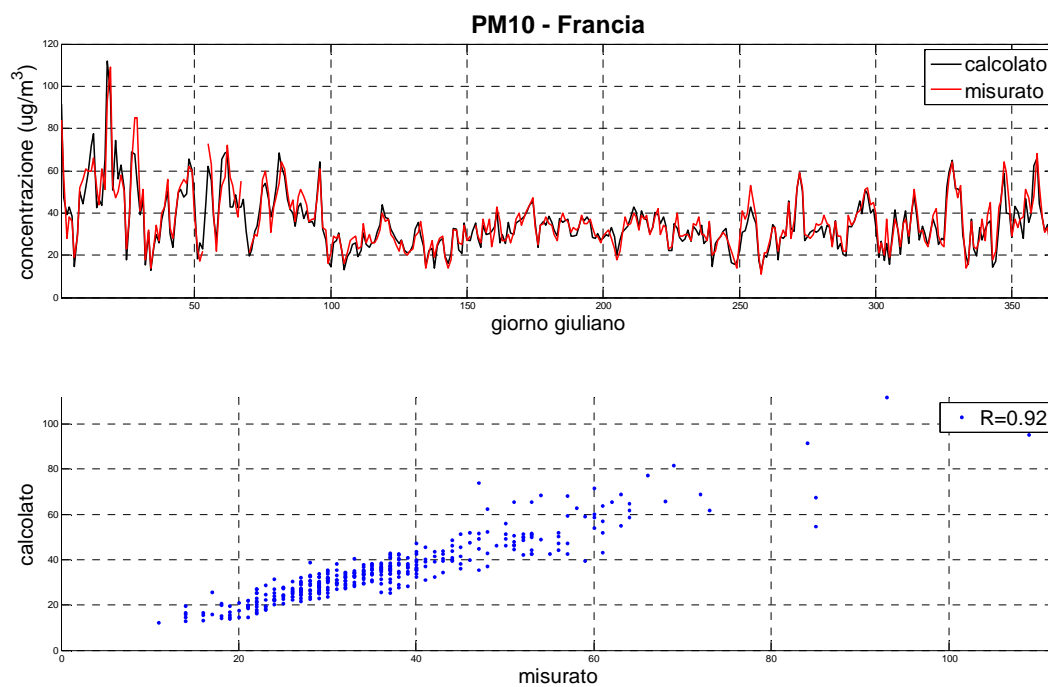


Fig. 9b: confronto tra i valori di una delle serie decimate casualmente ed i valori misurati a Roma –Corso Francia

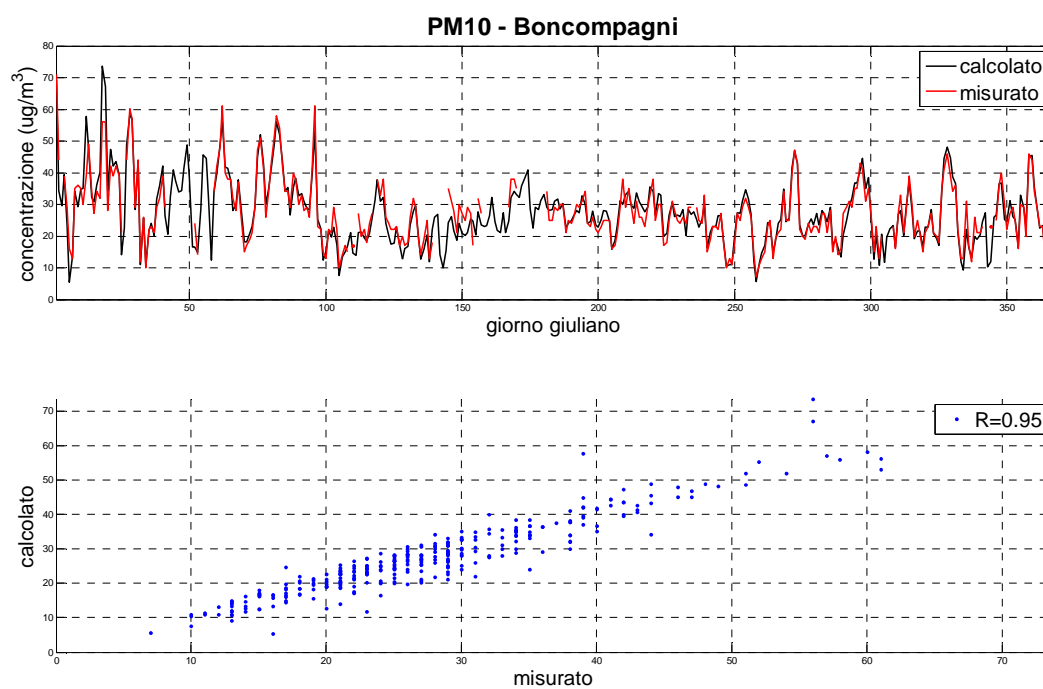


Fig. 9c: confronto tra i valori di una delle serie decimate casualmente ed i valori misurati a Roma – Via Boncompagni

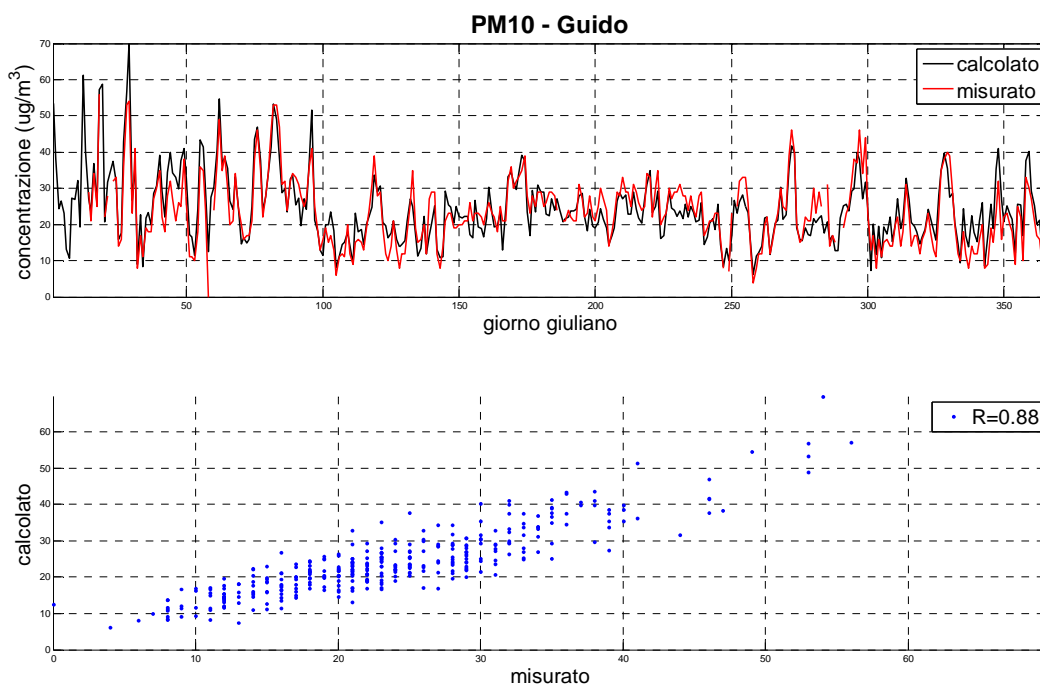


Fig. 9d: confronto tra i valori di concentrazione di una delle serie decimate casualmente ed i valori misurati a Castel di Guido

La visione delle Figg. 9 fornisce solo una visione qualitativa della bontà della procedura di gap-filling. Maggiore oggettività è data dal calcolo del valore medio annuo di PM_{10} per ciascuna serie storica decimata casualmente e poi ricostruita per mezzo dello stimatore (3) e dall'individuazione della distribuzione di tali valori medi riportata in Fig. 10 per le stazioni di via Boncompagni e di Via Magna Grecia. Come si nota, queste distribuzioni sono sostanzialmente gaussiane, centrate esattamente sul valore medio di concentrazione media annua effettivamente riscontrato nella serie storica non decimata e caratterizzate da un'ampiezza (deviazione standard) molto contenuta. Per le altre stazioni della rete si è ottenuto un risultato del tutto simile. Inoltre il numero di superamenti del valore di soglia della concentrazione media giornaliera caratteristica delle serie decimate differiva dal valore caratteristico della serie non decimata al massimo di più o meno una unità.

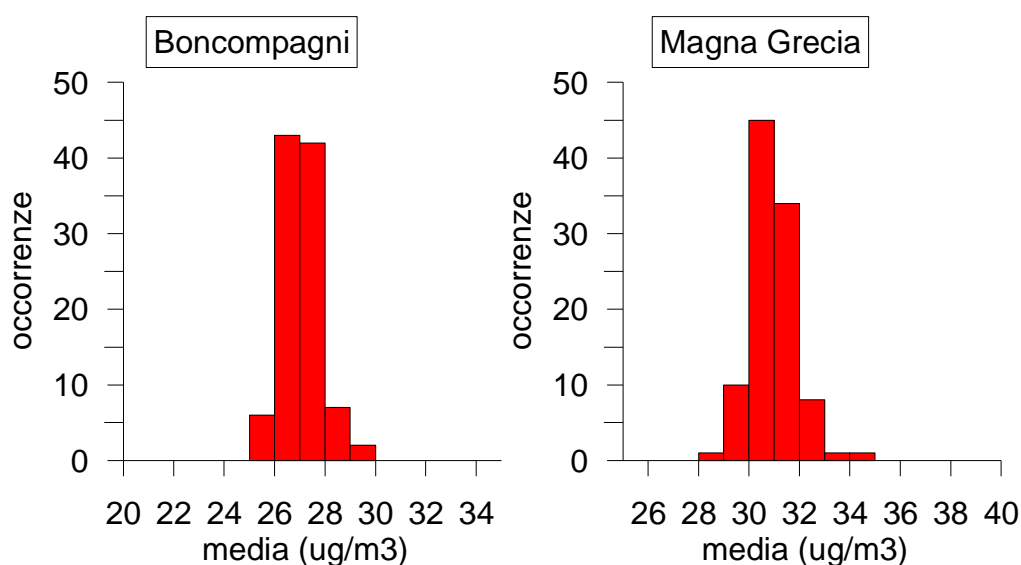


Fig.10: alcune distribuzioni del valore medio annuo di concentrazione di PM_{10} presentato dalle serie decimate generate nella simulazione Monte Carlo

7. Misure discontinue e postazioni virtuali

Un sistema di monitoraggio della qualità dell'aria, secondo la normativa vigente, è in generale costituito da due diverse tipologie di misure: le misure realizzate dagli analizzatori posti entro le postazioni di monitoraggio fisse che costituiscono la rete e le misure realizzate con laboratori mobili in cui normalmente sono presenti gli stessi analizzatori che operano nelle postazioni fisse. Entrambi gli strumenti di monitoraggio rivestono ad oggi un ruolo fondamentale nel controllo della qualità dell'aria. La rete fissa, attraverso la quale è possibile effettuare le misure senza rilevanti discontinuità temporali in un certo punto dello spazio, è considerata lo strumento base per il controllo della qualità dell'aria poiché descrive l'evoluzione temporale dei livelli orari o giornalieri di concentrazione durante l'intero arco di vita della stazione di monitoraggio. La *staticità* della rete fissa rappresenta quindi il punto di forza della misura ma, paradossalmente, anche la sua debolezza. Di fatto tale strumento fornisce informazioni dettagliate sui livelli di concentrazione delle sostanze inquinanti rappresentative di un'area intorno al punto di misura stesso la cui estensione è strettamente dipendente dal contesto urbano, orografico ed emissivo in cui è localizzata la stazione di monitoraggio. Per quanto numerose possano essere le postazioni fisse di una rete di monitoraggio, è pressoché impossibile considerare questo come l'unico strumento per la valutazione della qualità dell'aria.

La norma nazionale (il D.Lgs. 155/2010 che recepisce la direttiva 2008/50/CE), che oggi regola le metodologie di controllo della qualità dell'aria, richiede di effettuare la valutazione della qualità dell'aria integrando le misure ottenute dalla rete fissa di monitoraggio con le misure effettuate mediante l'utilizzo di laboratori mobili, in grado di fornire, in determinati punti dello spazio, delle informazioni considerate *indicative* sulla stato di qualità dell'aria. La capacità di un laboratorio mobile di coprire nell'arco di un anno diversi punti del territorio rappresenta pertanto un'informazione aggiuntiva e complementare all'informazione fornita costantemente dalla rete fissa per avere una conoscenza più capillare sul territorio dei livelli di qualità dell'aria. D'altro canto, l'elasticità di un laboratorio mobile, e cioè la sua capacità di *muoversi* sul territorio, rappresenta anche il suo punto critico poiché per sua stessa natura non è mai in grado di fornire informazioni per intervalli di tempo confrontabili con quelli di una rete fissa che poi definiscono anche i periodi su cui calcolare gli indicatori statistici (valore medio annuo e numero di superi del valore soglia). Nasce a questo punto l'esigenza di sfruttare appieno le potenzialità di un laboratorio mobile con l'obiettivo di rendere in qualche modo compatibili misure che per caratteristiche temporali sono profondamente differenti. Il primo modo per rendere immediatamente fruibili le misure di concentrazione media giornaliera di PM_{10} ottenute grazie all'impiego dei mezzi mobili è data direttamente dalla normativa vigente (D.Lgs. 155/2010). In effetti, nel caso di misure discontinue nel tempo, ma sufficientemente omogenee durante l'arco dell'anno, se il periodo minimo di copertura annuale è almeno il 14% (8 settimane distribuite uniformemente durante l'arco dell'anno), ai sensi del D.Lgs.155/2010 alla Nota (1) della Tabella 1 dell'Allegato I è possibile determinare se si ha un superamento o meno della soglia dei $50 \mu g/m^3$ non contando i singoli eventi di supero ma seguendo quanto recita il citato decreto:

“Se le misurazioni discontinue sono utilizzate per valutare il rispetto del valore limite del PM_{10} , occorre valutare il 90.4 percentile (che deve essere inferiore o uguale a $50 \mu g/m^3$) anziché il numero di superamenti, il quale è fortemente influenzato dalla copertura dei dati.”

Al di là di ciò, è comunque ipotizzabile l'impiego dello stimatore BLUE (3) per ricostruire l'intera serie storica annuale delle concentrazioni medie giornaliere di particolato sottile. La metodologia che è ragionevole impiegare potrebbe essere così articolata:

- si eseguano con un mezzo mobile una serie di campagne sperimentali in un dato punto del territorio (che quindi sarà una postazione *pivot* per lo stimatore BLUE), raccogliendo le concentrazioni medie giornaliere di particolato e si selezionino le porzioni di serie storiche delle misure di concentrazione media giornaliera di particolato rilevate dalle postazioni della rete fissa di monitoraggio relative ai periodi di campagna;
- si applichi la procedura riportata al punto 4, ottenendo per questo punto del territorio sede della campagna una gerarchia di quadruple di postazioni *slave* con i relativi pesi λ_i ;
- si consideri l'intero periodo in cui si sono svolte le campagne considerate (se per esempio sono state realizzate 4 campagne durante l'intero anno, è l'intero anno il periodo da considerare);
- sulla base delle serie storiche relative alle varie postazioni *slave* si applichi giorno dopo giorno per l'intero periodo considerato lo stimatore BLUE (3);

- alla fine si ottiene una serie storica costituita dalle misure realizzate dal mezzo mobile per tutti i giorni in cui era presente nel sito considerato e dalle stime per tutti i giorni in cui il mezzo mobile era assente.

Anche se del tutto ragionevole, questa procedura operativa va valutata sulla base di criteri oggettivi. Per fare ciò, si è simulata la realizzazione di campagne sperimentali prendendo in considerazione le serie storiche dell'anno 2012 relative a diverse postazioni della rete di monitoraggio e decimandole casualmente in modo che fossero disponibili solo le misure relative a 8 settimane distribuite omogeneamente durante l'anno. Su queste serie storiche decimate è stato applicato lo stimatore BLUE (3) ottenendo il *gap filling* dei dati mancanti e producendo quindi serie storiche complete per le quali è stato possibile determinare il valore medio annuo della concentrazione media giornaliera di PM_{10} ed il numero di superamenti del valore di soglia. Il risultato dell'applicazione della procedura descritta precedentemente al caso delle stazioni di Via Magna Grecia, Corso Francia, via Boncompagni e Castel di Guido sono riportate nelle *Figg. 11*. Nella parte superiore di ciascuna figura è riportato in rosso l'andamento delle concentrazioni medie giornaliere di PM_{10} effettivamente misurate nella postazione considerata cui si sovrappone in nero l'andamento della stima realizzata dallo stimatore BLUE. Per meglio apprezzare le *performances* della procedura, nella parte inferiore di ciascun grafico è presentato lo *scatter-plot* tra la serie temporale di PM_{10} riempita applicando il metodo statistico (asse delle ordinate) e la concentrazione realmente misurata nel 2012 (asse delle ascisse). Come si nota visivamente il comportamento della procedura è più che soddisfacente.

Per rendere quantitativo tutto ciò, per alcune postazioni significative della rete regionale sono stati calcolati gli standard di qualità previsti dalla normativa vigente, cioè la media annuale della concentrazione media giornaliera di PM_{10} ed il numero di superamenti del valore di soglia in tre casi distinti: ciò si ottiene considerando le serie storiche realmente misurate nelle postazioni considerate, ciò che si ottiene considerando solo le misure sopravvissute alla decimazione applicando quanto consentito dal D.Lgs. 155/2010 per le misure discontinue e ciò che si ottiene dall'utilizzo diretto delle serie storiche ricostruite impiegando lo stimatore BLUE. Il risultato ottenuto è riportato complessivamente nella *Tab. 9*. Come si può vedere, le serie ricostruite presentano indici di qualità dell'aria in generale più vicini a quelli presentati dalla serie di concentrazioni non decimate, mentre la verifica del numero di superi della serie incompleta spesso è di molto inferiore rispetto al valore reale e in ciò sta la disposizione normativa ricordata secondo cui in questi casi il rispetto dello standard va fatto impiegando il 90.4° percentile.

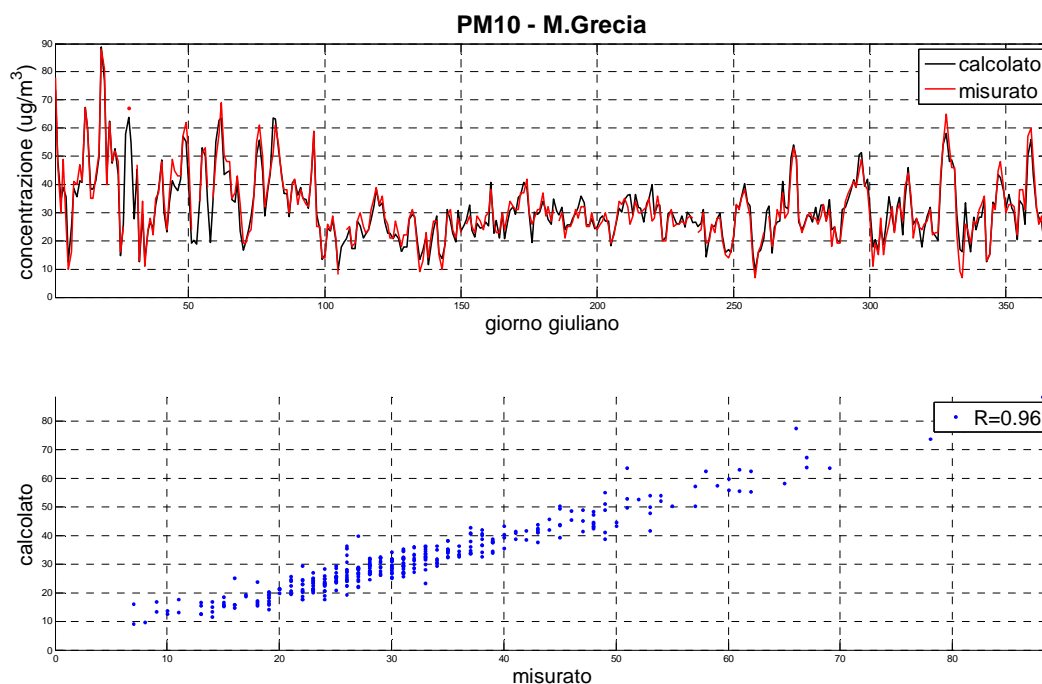


Fig. 11a: confronto tra la serie storica di Roma – Via Magna Grecia e la sua ricostruzione a partire dalla decimazione della serie originaria in cui sopravvivono solo 8 settimane l'anno distribuite uniformemente

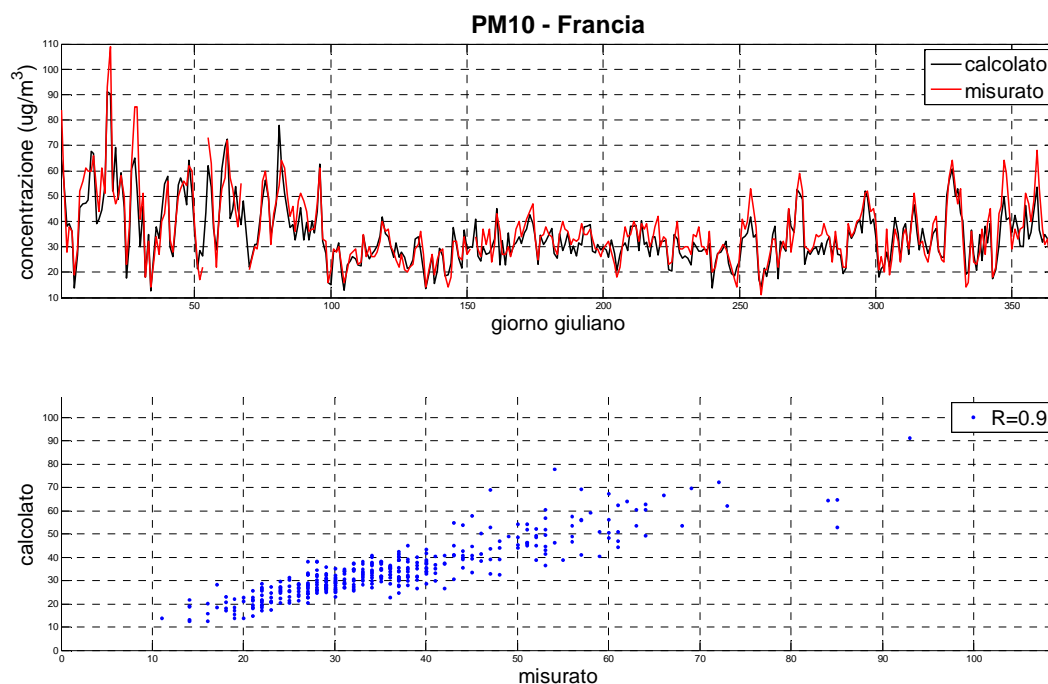


Fig. 11b: confronto tra la serie storica di Roma – Corso Francia e la sua ricostruzione a partire dalla decimazione della serie originaria in cui sopravvivono solo 8 settimane l'anno distribuite uniformemente

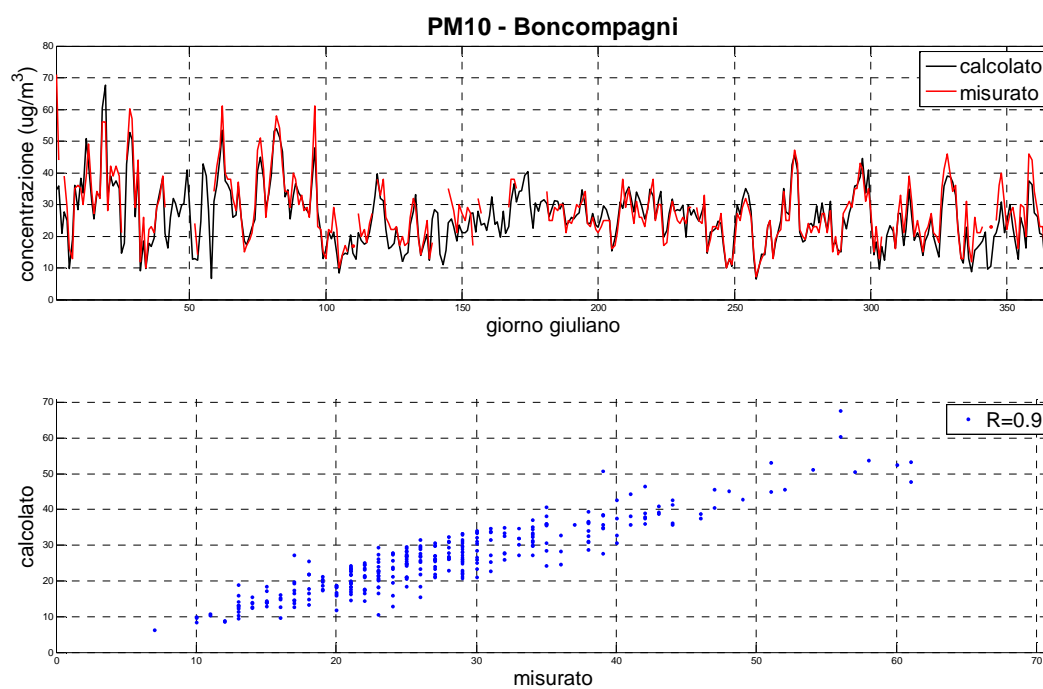


Fig. 11c: confronto tra la serie storica di Roma – Via Boncompagni e la sua ricostruzione a partire dalla decimazione della serie originaria in cui sopravvivono solo 8 settimane l'anno distribuite uniformemente

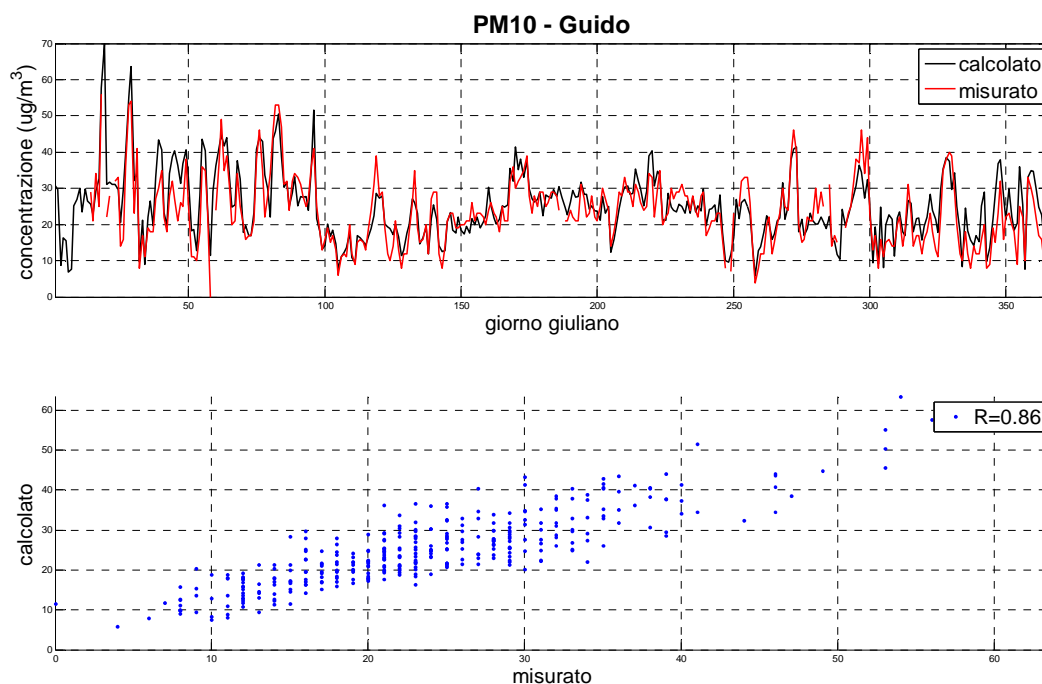


Fig. 11d: confronto tra la serie storica di Castel di Guido e la sua ricostruzione a partire dalla decimazione della serie originaria in cui sopravvivono solo 8 settimane l'anno distribuite uniformemente

stazione	osservazioni		serie decimata		BLUE+osservazioni	
	media ($\mu\text{g}/\text{m}^3$)	numero di superamenti	media ($\mu\text{g}/\text{m}^3$)	numero di superamenti	media ($\mu\text{g}/\text{m}^3$)	numero di superamenti
Grecia	32	29	33	7	32	28
Francia	36	57	37	10	35	42
Boncompagni	28	12	27	2	26	7
Ada	24	9	25	2	25	8
Guido	24	5	22	1	25	5
Preneste	34	45	35	8	33	28
Cinecittà	35	53	36	10	34	45
Cavaliere	28	12	25	2	24	4
Fermi	33	30	34	7	33	37
Tiburtina	37	50	39	11	36	56
Cipro	27	19	28	4	28	17
Bufalotta	28	16	29	5	29	21
Arenula	30	18	30	4	29	13
Malagrotta	28	23	29	6	28	20

Tab. 9: standard di qualità dell'aria per la concentrazione media giornaliera di PM_{10} nel caso a) delle serie storiche originarie, b) delle serie decimate e c) delle serie decimate e ricostruite con lo stimatore BLUE

Sulla base dei risultati ottenuti, risulta decisamente realistico estendere questa procedura al caso delle campagne sperimentali realizzate con i mezzi mobili, purché esse vengano ripetute periodicamente nell'arco dell'anno, vengano effettuate in periodi distribuiti uniformemente nell'arco dell'anno e vengano ripetute periodicamente, anno dopo anno, negli stessi punti del territorio. In questo caso, dopo un periodo di transizione che si può stimare in almeno quattro settimane di campagne sperimentali realizzate in un dato sito, si calcolano sulla base della procedura riportata al punto 4 la gerarchia di quadruple delle postazioni *slave* ed i relativi pesi. Ciò comporta che sia possibile attivare la procedura sopra descritta per ricostruire la serie storica

relativa al sito considerato, che verrà mantenuta sempre attiva fornendo costantemente una stima di concentrazione media giornaliera. Allo scadere di ogni anno si riaggiognerà la gerarchia delle postazioni *slave* ed i relativi pesi per tener conto di eventuali variazioni nel quadro emissivo locale e non e delle variazioni del quadro meteorologico e micrometeorologico. Il risultato netto sarà che pur non avendo aggiunto nuove postazioni fisse alla rete di monitoraggio regionale, nei fatti ad essa si aggiungeranno tante *postazioni virtuali* quanti saranno i siti sedi delle campagne sperimentali periodiche con i mezzi mobili, incrementando notevolmente le informazioni disponibili per la valutazione della qualità dell'aria del territorio. Nella regione Lazio è stato realizzato un piano di monitoraggio periodico con i mezzi mobili allo scopo di aggiungere alla rete fissa di monitoraggio almeno una decina di postazioni virtuali localizzate in punti del territorio in cui era necessario incrementare l'informazione della qualità dell'aria. Nell'anno 2014 sarà possibile verificare i risultati ottenuti.

8. Conclusioni e sviluppi futuri

L'esigenza di poter disporre di un'adeguata informazione per valutare la qualità dell'aria di un territorio regionale dettata dalla normativa vigente è fortemente limitata dal numero di postazioni della rete fissa di monitoraggio della qualità dell'aria che non possono essere incrementate ulteriormente visti i costi di investimento e di gestione che comportano. Per superare queste limitazioni, nel presente lavoro è stato esplorato l'uso di uno stimatore lineare di tipo geostatistico in diverse situazioni operative reali, limitandone l'applicazione alla sola concentrazione media giornaliera di PM_{10} .

In primo luogo è stata presentata la base teorica su cui si fonda tale stimatore e sono stati presentati i dettagli matematici indispensabili per una sua applicazione pratica. Questo stimatore, che fornisce la stima della concentrazione media di PM_{10} in un dato punto del territorio ed in un dato giorno, si configura come una semplice media pesata della concentrazione media giornaliera di PM_{10} rilevata in altri punti del territorio. Come si è visto, l'impiego pratico dello stimatore richiede l'estrazione dalle serie storiche disponibili delle informazioni relative alla fenomenologia della dispersione degli inquinanti attraverso un'opportuna analisi statistica che si condensa nella stima del semivariogramma, cioè nell'indicatore di similarità statistica tra coppie di postazioni di monitoraggio. Una volta note queste variabili statistiche, un apparato matematico puramente algebrico conduce alla determinazione dei pesi richiesti dallo stimatore.

Sono stati esplorati tre possibili utilizzi pratici di questo stimatore nella valutazione della qualità dell'aria del territorio laziale limitatamente al PM_{10} . Il primo impiego è rivolto al supporto nell'azione quotidiana di validazione delle misure realizzata dagli operatori della rete di monitoraggio. Accanto alle azioni messe in atto dagli operatori della rete, la maggior parte delle quali di tipo soggettivo, viene proposta e valutata una procedura totalmente automatizzabile che fornisce agli operatori un intervallo probabile (ad un determinato livello di confidenza) entro cui la misura dovrebbe collocarsi rispetto alla stima operata dallo stimatore perché possa essere ritenuta priva di anomalie. Nel Centro Regionale della Qualità dell'aria è previsto a breve di realizzare un sistema informatico di supporto agli operatori della rete che li assista nella validazione quotidiana delle misure di PM_{10} .

La seconda applicazione è dedicata al *gap filling*, cioè alla ricostruzione dei dati mancanti nelle serie storiche rilevate dagli analizzatori operativi nelle varie postazioni della rete. L'applicazione di una tale procedura operativa da parte del Centro Regionale della Qualità dell'Aria consente di aumentare di fatto il rendimento della rete e di poter disporre di serie storiche virtualmente complete, utili per realizzare efficacemente e realisticamente la valutazione della distribuzione nello spazio e nel tempo della concentrazione media giornaliera del PM_{10} .

Da ultimo, l'impiego dello stimatore geostatistico a servizio delle campagne sperimentali realizzate periodicamente dai mezzi mobili in siti del territorio laziale privi di postazioni di monitoraggio fisse consente di costruire serie storiche anche per questi punti. Esse conterranno vere e proprie misure quando il mezzo mobile è operativo nel sito e conterranno le migliori stime possibili quando il mezzo mobile opera altrove. Di fatto in questo modo si è introdotto nella rete di monitoraggio un nuovo insieme di *postazioni di monitoraggio virtuali* che sorvegliano punti particolarmente critici del territorio regionale privi di misure strumentali continue, incrementando significativamente le informazioni ambientali raccolte che verranno assimilate nelle catene modellistiche che garantiscono la valutazione e la previsione della qualità dell'aria per l'intero territorio regionale, come richiesto dalla normativa vigente.

Lo stimatore qui presentato si riferisce esclusivamente alla concentrazione media giornaliera di PM_{10}

e gli sviluppi futuri previsti sono tutti focalizzati sulla sua applicazione ad inquinanti diversi e a tempi di mediazione differenti. L'estensione più logica e più immediata dello stimatore quella relativa al $PM_{2.5}$, agli IPA ed ai metalli, dato per tutti questi inquinanti la normativa richiede che si realizzino medie giornaliere. Tuttavia, l'estensione più interessante e più immediata prevista è quella di realizzare uno stimatore geostatistico dedicato alla stima della concentrazione media oraria di NO_2 e di O_3 . E' ragionevole pensare che nello sviluppo di un tale stimatore si renda necessario superare alcune delle drastiche ipotesi di lavoro introdotte nello sviluppo dello stimatore della concentrazione media giornaliera di particolato. In particolare è ragionevole che si debba superare l'ipotesi di isotropia spaziale perché a livello orario il movimento delle masse d'aria dovrebbe comportare che i pesi relativi alla coppia stazione *pivot* stazione *slave* dipendano non solo dalla distanza tra le due stazioni ma anche dalla direzione del vettore che le unisce. Ciò dovrebbe comportare un aumento limitato della complessità nella procedura matematica dedicata alla determinazione dei pesi presenti nello stimatore. Comunque sulla base di sperimentazioni preliminari in cui è stato applicato lo stimatore nella forma attuale anche alla stima del valore orario della concentrazione di NO_2 si è visto che sorprendentemente il comportamento dello stimatore è stato buono, oltre ogni ragionevole speranza.

Bibliografia

- [1] Seaton A, MacNee W, Donaldson K, et al.: *Particulate air pollution and acute health effects*. - Lancet, 1995; 345: 176–178.
- [2] Brook RD, Rajagopalan S, Pope CA 3rd, et al. (2010): *Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association*. Circulation 2010; 1 June: 2331-2378.
- [3] Direttiva Europea 2008/50/CE 107/CE, *Qualità dell'aria ambiente e per un'aria più pulita in Europa*, Gazzetta Ufficiale dell'Unione Europea del 11/06/2008, n. L 152
- [4] Dutton J.A.(1995): *Dynamics of Atmospheric Motion* – Dover Publications, 617 pp
- [5] Stull R.B.(1989): *An Introduction to Boundary Layer Meteorology* – Kluwer Academic Publishers, 666 pp
- [6] Garratt J.R.(1992): *The Atmospheric Boundary Layer* – Cambridge University Press, 316 pp
- [7] Sorbjan (1989): *Structure of the Atmospheric Boundary Layer* – Printice Hall, 317 pp
- [8] Kaimal e Finnigan (1994): *Atmospheric Boundary Layer Flows: their structure and measurements* – Oxford University Press, 289 pp
- [9] Sozzi R., T. Georgiadis, M. Valentini (2002): *Introduzione alla Turbolenza Atmosferica: concetti, stime, misure* – Pitagora Editrice, 525 pp
- [10] Venkatram A, J.C. Wyngaard (1988): *Lectures on Air Pollution Modeling* – American Meteorological Society, 390 pp
- [11] Seinfeld e Pandis (1998): *Atmospheric Chemistry and Physics: from air pollution to climate change* – John Wiley&Sons, 1326 pp
- [12] Pielke R.A. (1984): *Mesoscale Meteorological Modeling* – Academic Press, 612 pp
- [13] Jacobson M.Z. (2005): *Fundamentals of Atmospheric Modeling* – Cambridge University Press, 813 pp
- [14] Baklanov, A et al., (2007): *Integrated systems for forecasting urban meteorology, air pollution and population exposure*. Atmos. Chem. Phys., 7, 855–874
- [15] Kalnay E. (2003): *Atmospheric Modeling, Data Assimilation and Predictability* – Cambridge University Press, 341 pp
- [16] Bennett A.F. (2002): *Inverse Modeling of the Ocean and atmosphere* – Cambridge University Press, 234 pp

- [17] D.Lgs. 155/2010, *Attuazione della direttiva 2008/50/CE relativa alla qualità dell'aria ambiente e per un'aria più pulita in Europa*, Gazzetta Ufficiale della Repubblica Italiana del 15 settembre 2010, n. 217
- [18] Goovaert P (1997): *Geostatistics for natural resources Evaluation* – Oxford University Press, 482pp
- [19] Wackernagel H (1998): *Multivariate Geostatistics* – Springer, 291 pp
- [20] Kitanidis P.K.: *Introduction to Geostatistics: applications ti hydrology* – Cambridge University Press, 249 pp
- [21] Press W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery (1992): *Numerical Recipes in FORTRAN* – Cambridge University Press, 963 pp.
- [22] Cotton WR, R.A. Pielke, R.L. Walko, G.E. Liston, C.J. Tremback, H. Jiang, R.L. McAnelly, J.Y. Harrington, M.E. Nicholls, G.C. Carrio, J.P. McFadden (2003): *RAMS 2001: current status and future directions* – Meteorology and Atmospheric Physics, 82, 5-29
- [23] Gariazzo C., C. Silibello, S. Finardi, P. Radice, A. Piersanti, G. Calori, A. Cecinato, C. Perrino, F. Nussio, M. Cagnoli, A. Pelliccioni, G.P. Gobbi, P. Di Filippo (2007): *A gas/aerosol air pollutants study over the urban area od Rome using a comprehensive chimica tran sport model* – Atmospheric Environment 41, 7286-7303
- [24] Silibello C., G. Calori, G. Brusasca, A. Giudici, E. Angelino, G. Fossati, E. Peroni, E. Buganza (2008): *Modelling of PM₁₀ concentrations over Milano urban area using two aerosol modules* – Environmental Modelling and Software 23, 333-343
- [25] Sozzi R., S. Ceradini, F. Troiano, A. Di Giulio, F. Sacco, S. Listrani, F. Barbini, A. Bolignano, M. Morelli (2012): *Il monitoraggio delle polveri sottili a Roma: la rete tradizionale e la postazione alla sommità del canopy urbano* – Quinto convegno nazionale sul particolato atmosferico PM2012, Perugia
- [26] Lin L. I-Kuei (1989): *A concordance correlation coefficient to evaluate reproducibility* _ Biometrics 45, 255-268
- [27] Von Storch H, F.W. Zwiers (1999); *Statistical Analysis in Climate Research* – Cambridge University Press, 484 pp
- [28] Moore D.S., G.P. McCabe (1993): *Introduction to the Practice of Statistics. Second Edition* – W.H. Freeman and Company, New York, 854 pp