



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria Civile,
Ambientale e Meccanica

**SVILUPPO E VERIFICA DI UNA METODOLOGIA PER LA
NORMALIZZAZIONE DELLE CONCENTRAZIONI MEDIE
ANNUE DI BISSIDO DI AZOTO RISPETTO ALLE
CONDIZIONI METEOROLOGICHE PER LA PROVINCIA DI
BOLZANO**

RELAZIONE TECNICA

Prefazione

La presente relazione tecnica finale è stata elaborata in adempimento di quanto previsto dalla convenzione stipulata in data 01 ottobre 2018 fra l'Agenzia provinciale per l'ambiente e la tutela del clima della Provincia Autonoma di Bolzano (APPA) e il Dipartimento di Ingegneria Civile, Ambientale e Meccanica (DICAM) dell'Università di Trento, per l'elaborazione di uno studio sulla normalizzazione meteorologica delle concentrazioni di NO_2 sul territorio della Provincia di Bolzano. Le relazione tecnica rendiconta inoltre gli approfondimenti richiesti da APPA e concordati durante la riunione del 15 marzo 2019, svolta presso il DICAM. Come da offerta prot. n. 30096/P/5.10 del 07 settembre 2018, lo studio dal titolo: *“Sviluppo e verifica di una metodologia per la normalizzazione delle concentrazioni medie annue di biossido di azoto rispetto alle condizioni meteorologiche per la Provincia di Bolzano”* si è sviluppato nelle seguenti azioni:

1. Analisi della letteratura tecnico/scientifica del settore per valutare metodologie già applicate in contesti simili.
2. Reperimento ed analisi dei dati disponibili per quanto riguarda le misure meteorologiche e le misure di qualità dell'aria (tale è stata svolta dall'Agenzia provinciale per l'ambiente e la tutela del clima della Provincia Autonoma di Bolzano).
3. Implementazione di alcuni algoritmi per la normalizzazione delle concentrazioni di biossido di azoto in funzione delle condizioni meteorologiche.
4. Taratura degli algoritmi per la normalizzazione delle concentrazioni di biossido di azoto e individuazione della formulazione più appropriata per i casi studio in esame.
5. Redazione della relazione tecnica finale del progetto.

Come previsto dalla convenzione lo studio è stato condotto sotto la responsabilità tecnico-scientifica del prof. Dino Zardi, coordinatore del Gruppo di Fisica

dell'Atmosfera del Dipartimento di Ingegneria Civile, Ambientale e Meccanica dell'Università degli studi di Trento, e grazie alla collaborazione del dott. Marco Falocchi, dell'ing. Lorenzo Giovannini afferenti allo stesso Gruppo.

Indice

Introduzione	1
1 Tecniche di normalizzazione meteorologica	4
1.1 Modelli statistici	5
1.2 Decision trees e random forest	5
1.3 Normalizzazione meteorologica con random forest	10
2 Datasets e caratterizzazione delle aree studio	12
2.1 Definizione dei datasets	12
2.1.1 Dati di qualità dell'aria	13
2.1.2 Dati meteorologici	13
2.2 Caratterizzazione meteo-climatologica e delle concentrazioni di NO ₂	16
3 Implementazione dell'algoritmo di normalizzazione meteorologica	32
3.1 Organizzazione della cartella di lavoro	32
3.2 Files di input	33
3.2.1 Cartella: db	33
3.2.2 File: settings_XX1.r	35
3.3 Codice principale	40
3.3.1 File: NO2Normalization_24h_XX1.r	40
3.3.2 Cartella: scripts	43
3.4 Files di output	44
4 Analisi delle concentrazioni normalizzate di NO₂	47
Conclusioni	64
Bibliografia	66

Elenco delle figure

2.1	Caratterizzazione del regime delle precipitazioni a scala mensile riferita al periodo 2012–2018 per le stazioni di Bronzolo, Bolzano Ospedale, Bolzano Via Amba Alagi, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	21
2.2	Cicli mensili della radiazione globale misurati presso le stazioni di Bronzolo, Bolzano Ospedale, Bolzano Via Amba Alagi, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	22
2.3	Cicli mensili della temperatura dell’aria misurati presso le stazioni di Bronzolo, Bolzano Ospedale, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	23
2.4	Cicli mensili del <i>temperature lapse rate</i> misurati dal profilatore termico nei primi 100 m e 500 m di atmosfera dal suolo, tra la stazione di Merano e Lauregno, tra la stazione di Bressanone e Fie allo Sciliar e tra Bressanone e Plose. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	24
2.5	Distribuzioni in frequenza dell’intensità del vento a scala stagionale per la stazione di Bronzolo. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	25
2.6	Distribuzioni in frequenza dell’intensità del vento a scala stagionale per la stazione di Bolzano Ospedale. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	25
2.7	Distribuzioni in frequenza dell’intensità del vento a scala stagionale per la stazione di Bolzano Via Amba Alagi. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	26
2.8	Distribuzioni in frequenza dell’intensità del vento a scala stagionale per la stazione di Merano. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	26

2.9	Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	27
2.10	Cicli mensili di concentrazioni di NO ₂ misurati dalle stazioni di Bolzano Via Claudia Augusta (BZ4), Bolzano Piazza Adriano (BZ5), Laives (LS1), Merano (ME1) e Bressanone (BX1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	28
2.11	Distribuzioni in frequenza delle concentrazioni di NO ₂ a scala stagionale per la stazione di Bolzano Via Claudia Augusta (BZ4). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	29
2.12	Distribuzioni in frequenza delle concentrazioni di NO ₂ a scala stagionale per la stazione di Bolzano Piazza Adriano (BZ5). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	29
2.13	Distribuzioni in frequenza delle concentrazioni di NO ₂ a scala stagionale per la stazione di Laives (LS1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	30
2.14	Distribuzioni in frequenza delle concentrazioni di NO ₂ a scala stagionale per la stazione di Merano (ME1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	30
2.15	Distribuzioni in frequenza delle concentrazioni di NO ₂ a scala stagionale per la stazione di Bressanone (BX1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).	31
4.1	Concentrazioni NO ₂ normalizzate sul periodo 2012–2018.	49
4.2	Concentrazioni presso la stazione di Bolzano Via Claudia Augusta (BZ4). Concentrazioni di NO ₂ misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO ₂ normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della <i>random forest</i> e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.	53

4.3	Concentrazioni presso la stazione di Bolzano Piazza Adriano (BZ5). Concentrazioni di NO ₂ misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO ₂ normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della <i>random forest</i> e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.	54
4.4	Concentrazioni presso la stazione di Laives (LS1). Concentrazioni di NO ₂ misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO ₂ normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della <i>random forest</i> e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.	55
4.5	Concentrazioni presso la stazione di Merano (ME1). Concentrazioni di NO ₂ misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO ₂ normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della <i>random forest</i> e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.	56
4.6	Concentrazioni presso la stazione di Bressanone (BX1). Concentrazioni di NO ₂ misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO ₂ normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della <i>random forest</i> e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.	57

4.7	Peso delle variabili esplicative nella <i>random forest</i> realizzata per analizzare le concentrazioni di NO ₂ sul periodo 2012–2018. Gli acronimi delle grandezze esplicative sono: WG intensità del vento, WR: direzione del vento, LT: temperatura dell'aria, LF umidità relativa, LD pressione atmosferica, GS radiazione globale, N precipitazione, Γ <i>temperature lapse rate</i> , “unix” data unix, “day” giorno della settimana, “julian” giorno giuliano. I numeri al pedice corrispondono alla numerazione delle stazioni al suolo elencate in Tab. 4.1 e utilizzate per ciascuna serie di NO ₂ , mentre si indicano con Γ_{100} e Γ_{500} i <i>temperature lapse rate</i> ricavati dal profilatore termico per uno strato di atmosfera con spessore rispettivamente di 100 e 500 m.	58
4.8	Dipendenza delle concentrazioni osservate di NO ₂ da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bolzano Via Claudia Augusta (BZ4). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.	59
4.9	Dipendenza delle concentrazioni osservate di NO ₂ da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bolzano Piazza Adriano (BZ5). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.	60
4.10	Dipendenza delle concentrazioni osservate di NO ₂ da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Laives (LS1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.	61
4.11	Dipendenza delle concentrazioni osservate di NO ₂ da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Merano (ME1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.	62
4.12	Dipendenza delle concentrazioni osservate di NO ₂ da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bressanone (BX1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.	63

Elenco delle tabelle

1.1	Dataset di partenza dell'albero decisionale. La variabile da classificare (x) è riportata nella prima colonna, mentre nelle restanti colonne si riportano i valori assunti dalle variabili esplicative, selezionate o disponibili per descrivere x	6
1.2	Dataset ordinato in funzione dei valori di copertura.	7
1.3	Dataset ordinato in funzione dell'umidità.	8
1.4	Dataset ordinato in funzione del vento.	8
2.1	Stazioni convenzionali di qualità dell'aria gestite da APPA e utilizzate durante lo studio. La tabella riporta la consistenza delle serie disponibili, espressa in termini di percentuale di medie giornaliere disponibili per ciascun anno considerato.	13
2.2	Stazioni meteorologiche al suolo, appartenenti alla rete di monitoraggio gestita dal Servizio Meteorologico della Provincia Autonoma di Bolzano, selezionate ai fini dello studio.	14
2.3	Consistenza delle serie di intensità del vento (WG), direzione del vento (WR), temperatura dell'aria (LT), radiazione globale (GS) e di precipitazione (N) per ciascun anno del periodo selezionato (2012–2018) e per ciascuna stazione al suolo selezionata (Tab. 2.2). Si indica con “x” una consistenza superiore al 50%, con “o” una consistenza inferiore al 50% mentre l'assenza di simbolo indica l'assenza della serie storica.	16
2.4	Altezze di precipitazione cumulate annue per le stazioni meteorologiche selezionate e statistica dei valori annui per il periodo 2012-2017 al fine di favorire il confronto con l'anno 2018.	17
4.1	Utilizzo del dataset meteorologico per la normalizzazione delle serie di NO ₂	48

Introduzione

La normativa vigente in materia di tutela della qualità dell'aria stabilisce per ogni specie inquinante alcune soglie da non superare su base annua. Tali limiti, tuttavia, non tengono conto dell'impatto di diverse situazioni meteorologiche sui processi di trasporto e di dispersione degli inquinanti, che si riflettono di conseguenza sui valori misurati delle concentrazioni di inquinanti. Ad esempio, a parità di emissioni e di misure finalizzate al loro contenimento, la persistenza di venti intensi favorisce la diluizione degli inquinanti in atmosfera, mentre la persistenza di inversioni termiche al suolo inibisce il rimescolamento favorendo l'accumulo degli inquinanti nei bassi strati di atmosfera. Risulta pertanto importante poter valutare l'efficacia di eventuali politiche per il controllo della qualità dell'aria indipendentemente dalla forzante meteorologica.

Il presente studio si prefigge l'obiettivo di mettere a punto una metodologia scientificamente basata che consenta di normalizzare le concentrazioni medie annue di biossido di azoto (NO_2) rispetto alle condizioni meteorologiche medie che caratterizzano la Provincia di Bolzano, onde discriminare quindi l'effetto delle emissioni da quello delle condizioni meteorologiche. Infatti, soprattutto in contesti montani, le concentrazioni degli inquinanti misurate al suolo sono strettamente correlate ai processi di trasporto e di diffusione turbolenta, a loro volta condizionati dall'interazione del vento con la topografia complessa (*e.g.* Tomasi et al., 2018), dallo sviluppo di circolazioni locali (*e.g.* Giovannini et al., 2014a,b, 2017) e di inversioni termiche (*e.g.* Wolf et al., 2014; Falocchi et al., 2018).

La normalizzazione meteorologica è stata condotta per le serie storiche di concentrazioni di NO_2 misurate presso le stazioni di qualità dell'aria di: Bolzano (BZ4 e BZ5), Laives (LS1), Merano (ME1), Bressanone (BX1). Al fine di caratterizzare le diverse aree studio dal punto di vista meteo-climatologico, invece, sono stati utilizzati i dati delle stazioni di: Bolzano Ospedale, BZ6 (Via Amba Alagi), Bronzolo, San Genesio per la conca di Bolzano e Laives; Bressanone, Fiè allo Sciliar e Plose per la conca di Bressanone; Merano e Lauregno per la conca di Merano. Al fine di ottenere informazioni sulla stratificazione termica, sono stati utilizzati sia i dati relativi alle differenze di temperatura tra stazioni di fondovalle e in quota, sia

i profili verticali di temperatura misurati del profilatore termico installato presso l'aeroporto di Bolzano.

I dati di concentrazioni di NO_2 ed i profili di temperatura misurati sono stati messi a disposizione dall'Agenzia provinciale per l'ambiente e la tutela del clima della Provincia Autonoma di Bolzano, mentre i dati delle stazioni meteorologiche al suolo sono stati forniti dall'Ufficio Idrografico della Provincia Autonoma di Bolzano.

La presente relazione tecnica si articola in 4 Capitoli che rendicontano le attività svolte durante lo studio tecnico, le analisi effettuate ed i risultati ottenuti.

Capitolo 1. Analisi della letteratura tecnico/scientifica del settore. Il presente Capitolo è da intendersi come introduttivo alla tecnica della normalizzazione meteorologica di serie storiche di concentrazioni, in quanto qui si definisce la tecnica e si fornisce una panoramica dei metodi che possono essere implementati per eseguirla. In particolare, nel Capitolo si riportano due diversi approcci: (a) la tecnica della regressione (*e.g.* Zeldin & Meisel, 1978) e (b) l'utilizzo di tecniche di *machine learning* per la classificazione di dati (*e.g.* Grange et al., 2018; Grange & Carslaw, 2019). Una tale descrizione, oltre a contestualizzare l'analisi, delinea i pro ed i contro dei diversi approcci, al fine di individuare la strategia che può essere più efficacemente applicata per i casi studio della Provincia di Bolzano.

Capitolo 2. Caratterizzazione meteo-climatologica e delle concentrazioni di NO_2 per le aree di studio. Il Capitolo è dedicato alla caratterizzazione delle aree studio, sia dal punto di vista meteo-climatologico sia dal punto di vista delle concentrazioni di NO_2 . L'analisi è stata condotta sul periodo di riferimento 2012–2017. In particolare, la caratterizzazione meteo-climatologica delle aree in esame è stata condotta analizzando la temperatura dell'aria ed il suo gradiente verticale, la velocità del vento, la radiazione globale e la precipitazione. Tale analisi permette la contestualizzazione di ciascun sito investigato al fine di meglio comprendere i risultati restituiti dall'algoritmo di normalizzazione meteorologica.

Capitolo 3. Implementazione dell'algoritmo di normalizzazione meteorologica. Coerentemente con quanto riportato nel Capitolo 1, il presente Capitolo descrive l'implementazione dell'algoritmo di normalizzazione meteorologica basato sulla tecnica di classificazione dei dati nota come “*random forest*”. In particolare, l'algoritmo utilizzato è implementato a partire dal pacchetto “*rmweather*” (Grange et al., 2018; Grange & Carslaw, 2019), che permette di effettuare la normalizzazione di dati di qualità dell'aria sulla base delle condizioni meteorologiche presenti (<https://github.com/skgrange/rmweather>). Tale pacchetto, implemen-

tato in ambiente R, è open-source e può quindi essere distribuito e modificato liberamente.

Capitolo 4. Analisi delle concentrazioni normalizzate di NO_2 per l'anno 2018. L'algoritmo, descritto al Capitolo precedente (Cap. 3), è qui utilizzato per valutare l'andamento temporale delle concentrazioni di biossido di azoto nell'anno 2018 ed il loro trend nell'intero periodo osservativo 2012–2018 per le stazioni di qualità dell'aria prese in esame. Nel Capitolo si riportano e si commentano i risultati ottenuti.

Conclusioni In questo Capitolo si riassumono le attività svolte ed i risultati ottenuti.

CAPITOLO 1

Tecniche di normalizzazione meteorologica

Le concentrazione di un inquinante misurato in atmosfera e la sua evoluzione temporale dipendono principalmente da due fattori:

1. il quantitativo della sostanza immessa in atmosfera, *i.e.* il regime delle emissioni;
2. l'efficienza dei processi meteorologici nel disperdere l'inquinante.

Il primo fattore è essenzialmente legato all'implementazione di politiche atte a mitigare/contenere le emissioni. Il secondo, invece, è fortemente collegato ai processi meteorologici che localmente caratterizzano una data area. Infatti, a parità di emissione, è logico attendersi che la concentrazione misurata di un inquinante sia più bassa [alta] in siti in cui la capacità di dispersione dell'atmosfera sia più alta [bassa]. Al fine quindi di poter valutare l'effettiva concentrazione media annua di un dato inquinante in un'area e quindi valutare anche l'effetto di politiche atte a mitigare tali concentrazioni, risulta fondamentale scorporare dai valori misurati il ruolo della forzante meteorologica.

La normalizzazione meteorologica (nel seguito NM) è una tecnica di analisi che permette di svincolare l'effetto della variabilità delle condizioni meteorologiche su di una serie di concentrazioni di un inquinante atmosferico. La NM può essere condotta attraverso diversi approcci, tra cui l'utilizzo di:

1. modelli statistici predittivi basati sulla determinazione di trend (*e.g.* Zeldin & Meisel, 1978);
2. metodologie di *machine learning* per la classificazione di dati (*e.g.* Grange et al., 2018; Grange & Carslaw, 2019).

Entrambi gli approcci mirano a valutare la dipendenza di una concentrazione misurata da un set di grandezze, denominate variabili esplicative, che si assume essere

rappresentative dei processi atmosferici locali e che abbiano un effetto sulla dispersione degli inquinanti. In particolare, il set di variabili esplicative è tipicamente costituito da grandezze meteorologiche (*e.g.* intensità e direzione del vento, temperatura dell'aria, radiazione globale, pressione) e da variabili temporali quali l'ora, il giorno della settimana e dell'anno, che sono utilizzate come predittori dei cicli giornalieri, settimanali e stagionali degli inquinanti.

1.1 Modelli statistici

I modelli statistici sono tipicamente basati sull'analisi dei trend di concentrazioni e di come questi possano essere correlati alle variabili esplicative individuate. Per poter applicare questi modelli è anzitutto necessario soddisfare le ipotesi su cui si basano. Inoltre, il legame concentrazione–variabile esplicativa tipicamente non è lineare, così come non è noto a priori quali siano le variabili esplicative che maggiormente descrivono la variabilità delle concentrazioni. Per questi motivi l'utilizzo dei modelli statistici è molto difficoltoso.

1.2 Decision trees e random forest

Un approccio alternativo è offerto dall'utilizzo di algoritmi sviluppati nell'ambito del *machine learning*. Il *machine learning* rappresenta una fusione di statistica, analisi dati e computazione che permette lo sviluppo di numerosi modelli predittivi che, a loro volta, si configurano come alternativi ai classici modelli statistici, basati sulla regressione e l'estrapolazione delle informazioni. In particolare, i modelli predittivi permeati dal *machine learning* non devono soddisfare le ipotesi su cui, invece, si basano i modelli statistici.

Nell'ambito della NM, l'algoritmo di *machine learning* più utilizzato è quello noto come *random forest* o *decision forest* (Grange et al., 2018; Grange & Carslaw, 2019). Nelle linee che seguono si delinea il funzionamento dell'algoritmo, rimandando alla letteratura di riferimento per ulteriori dettagli (*e.g.* Breiman, 2001; Tong et al., 2003; Grange et al., 2018; Grange & Carslaw, 2019).

In linea del tutto generale, una *random forest* è un algoritmo che, a partire dai valori assunti dalle variabili esplicative selezionate, predice il valore di una grandezza dipendente (*e.g.* una concentrazione) come media d'insieme (o media d'*ensemble*) delle previsioni restituite dai suoi singoli elementi (*decision trees* o “alberi decisionali”). Al fine quindi di comprendere il funzionamento di una *random forest*, è necessario comprendere il funzionamento dei suoi singoli elementi, ossia come un singolo albero decisionale o *decision tree* è implementato.

Tabella 1.1: Dataset di partenza dell'albero decisionale. La variabile da classificare (x) è riportata nella prima colonna, mentre nelle restanti colonne si riportano i valori assunti dalle variabili esplicative, selezionate o disponibili per descrivere x .

x	copertura	umidità	vento
0	S	A	D
0	S	A	F
1	N	A	D
1	P	A	D
1	P	B	D
0	P	B	F
1	N	B	F
0	S	A	D
1	S	B	D
1	P	B	D
1	S	B	F
1	N	A	F
1	N	B	D
0	P	A	F

Un *decision tree* è un algoritmo binario ricorsivo che permette di classificare un set di osservazioni sulla base dei valori assunti da un set di variabili esplicative. A titolo esemplificativo, si illustra l'implementazione di un albero decisionale utilizzando il dataset riportato in Tab. 1.1. Questo dataset è detto di training in quanto è utilizzato per “far crescere” l'albero decisionale e contiene nella prima colonna i valori della variabile dipendente x (da classificare), mentre le restanti colonne contengono le variabili esplicative selezionate (copertura del cielo, umidità e vento). In questo dataset, tutte le variabili assumono valori discreti. In particolare:

- x assume valori 0 o 1;
- la copertura del cielo (copertura) assume i valori: S (soleggiato), N (nuvoloso), P (pioggia);
- l'umidità è indicata come alta (A) o bassa (B);
- il vento è indicato come forte (F) o debole (D).

Ciò permette di semplificare la descrizione del processo di “crescita” dell'albero decisionale, senza però perdere in generalità dell'approccio. Nei casi reali, dove le variabili assumono tipicamente valori continui, si procede definendo opportunamente gli intervalli di variabilità delle singole grandezze.

L'algoritmo seleziona una variabile esplicativa e ne classifica i valori, riorganizzando la struttura dell'intero dataset. Ad esempio, si consideri la copertura del cielo. Come anticipato questa variabile esplicativa assume tre valori (S, N e P). Si procede quindi nel riorganizzare il dataset che pertanto risulterà suddiviso in tre

Tabella 1.2: Dataset ordinato in funzione dei valori di copertura.

x	copertura	umidità	vento
1	N	A	D
1	N	B	F
1	N	A	F
1	N	B	D
1	P	A	D
1	P	B	D
0	P	B	F
1	P	B	D
0	P	A	F
0	S	A	D
0	S	A	F
1	S	B	D
0	S	A	D
1	S	B	F

segmenti, in ciascuno dei quali la copertura assume il medesimo valore. In Tabella 1.2 si riporta l'esito di questa operazione, in cui: nel primo segmento vi sono le osservazioni effettuate in presenza di cielo coperto (N), il secondo segmento è associato ad episodi di pioggia (P), mentre nel terzo segmento vi sono i dati in cui il cielo è soleggiato (S). A questo punto si valutano i valori assunti dalla variabile dipendente nei singoli segmenti, al fine di ottenere delle regole che ne permettano di predirne il comportamento. Nel caso in cui nel medesimo segmento la variabile dipendente assume lo stesso valore, allora si dice che il nodo ha raggiunto la purezza, ossia non è necessario procedere oltre nella classificazione. Viceversa, è necessario procedere a classificare il dataset in ragione di un'altra variabile esplicativa. Tornando al caso della copertura del cielo, si osserva così che:

1. quando il cielo è nuvoloso $x = 1$, indipendentemente dalle altre variabili. In questo caso il nodo ha raggiunto la purezza e non deve essere classificato ulteriormente;
2. quando il cielo è soleggiato o nuvoloso, x assume entrambi i valori e quindi è necessario procedere ad un'ulteriore classificazione.

Si ottiene così la regola:

- se copertura = N $\Rightarrow \mathbf{x} = 1$.

L'algoritmo procede quindi a riclassificare il dataset contenuto in ciascun segmento associato a periodi con pioggia e di cielo soleggiato in funzione dei valori assunti dall'umidità. Come nel caso precedente, si suddivide il dataset in segmenti in cui l'umidità assume lo stesso valore e si valutano i valori assunti dalla variabile dipendente in ciascun segmento. L'esito di questa seconda riorganizzazione è riportato in Tab. 1.3, da cui si osserva che:

Tabella 1.3: Dataset ordinato in funzione dell'umidità.

x	copertura	umidità	vento
1	N	A	D
1	N	B	F
1	N	A	F
1	N	B	D
1	P	A	D
0	P	A	F
1	P	B	D
0	P	B	F
1	P	B	D
0	S	A	D
0	S	A	F
0	S	A	D
1	S	B	D
1	S	B	F

Tabella 1.4: Dataset ordinato in funzione del vento.

x	copertura	umidità	vento
1	N	A	D
1	N	B	F
1	N	A	F
1	N	B	D
1	P	A	D
1	P	B	D
1	P	B	D
0	P	A	F
0	P	B	F
0	S	A	D
0	S	A	F
0	S	A	D
1	S	B	D
1	S	B	F

1. quando il cielo è soleggiato (S) $x = 0$ con umidità alta, mentre $x = 1$ con umidità bassa;
2. quando piove (P), x assume entrambi i valori indipendentemente dall'umidità dell'aria.

Si ottengono così due nuove regole:

- se copertura = S e umidità = A $\Rightarrow \mathbf{x} = \mathbf{0}$;
- se copertura = S e umidità = B $\Rightarrow \mathbf{x} = \mathbf{1}$.

Poiché i valori di x in presenza di pioggia sono indipendenti dall'umidità, si procede ricorsivamente nel classificare il segmento con pioggia in funzione del vento, ottenendo le ultime due regole:

- se **copertura** = **P** e **vento** = **F** $\Rightarrow \mathbf{x} = 0$;
- se **copertura** = **P** e **vento** = **D** $\Rightarrow \mathbf{x} = 1$.

La procedura ricorsiva, implementata a partire dalla classificazione del dataset sulla base della copertura, ha quindi permesso di ottenere le seguenti regole:

1. se **copertura** = **N** $\Rightarrow \mathbf{x} = 1$;
2. se **copertura** = **S** e **umidità** = **A** $\Rightarrow \mathbf{x} = 0$;
3. se **copertura** = **S** e **umidità** = **B** $\Rightarrow \mathbf{x} = 1$;
4. se **copertura** = **P** e **vento** = **F** $\Rightarrow \mathbf{x} = 0$;
5. se **copertura** = **P** e **vento** = **D** $\Rightarrow \mathbf{x} = 1$.

Sulla base dell'albero decisionale costruito è ora possibile predire il valore della variabile x noti i valori delle variabili esplicative “copertura”, “umidità” e “vento”. Tuttavia:

1. questa procedura è valida finché x assume i medesimi valori che ha assunto in passato con le medesime condizioni (variabili esplicative) osservate;
2. le regole ottenute dipendono dall'ordine con cui sono state selezionate le variabili esplicative per costruire l'albero decisionale;
3. le regole ottenute dipendono dal numero di variabili esplicative disponibili e dal numero di variabili esplicative selezionate per costruire l'albero decisionale. In presenza di numerose variabili esplicative, non è opportuno permettere la crescita completa dell'albero. Infatti, un'eccessiva crescita dell'albero decisionale può dar luogo a stime poco robuste, soprattutto quando la purezza di un nodo è raggiunta solamente una volta con una sola osservazione sull'intero dataset.

L'algoritmo *random forest* permette di superare queste problematiche, in quanto stima il valore della grandezza dipendente sulla base dei valori predetti da più alberi decisionali. Questi algoritmi si sviluppano come segue:

1. l'algoritmo separa il dataset osservativo disponibile in due sottoinsiemi: un sotto-insieme è dedicato a far crescere la *random forest* (training), mentre il secondo è dedicato alla sua validazione. Il dataset è ovviamente costituito dalla variabile dipendente e da un certo numero N di variabili esplicative. La

percentuale che definisce la dimensione del sottoinsieme di training è tipicamente definita dall'utilizzatore. In particolare, la realizzazione dei sottoinsiemi è ottenuta estraendo in modo casuale i dati disponibili dal dataset iniziale, al fine di disporre di due sotto-insiemi con osservazioni strettamente correlate temporalmente.

2. l'algoritmo procede a far “crescere” un numero di alberi decisionali definito dall'utilizzatore. Ciascun albero decisionale è fatto però crescere su di un numero limitato P di variabili esplicative, con $P < N$. Inoltre, anche in questo caso, la scelta delle P variabili esplicative è effettuata in modo casuale.
3. La scelta dell'ordine delle variabili esplicative è definita sulla base di criteri propri del *machine learning*.
4. La capacità predittiva della “foresta” di alberi decisionali è quindi valutata utilizzando il sottoinsieme di validazione. In questa fase, le variabili esplicative del sottoinsieme di validazione sono utilizzate per ottenere da ciascun albero decisionale della foresta una previsione della variabile dipendente. L'algoritmo calcola quindi una media d'insieme delle previsioni restituite dai singoli alberi e confronta tale valore con quello realmente osservato. Di conseguenza, il valore predetto dalla *random forest* è la media d'insieme delle predizioni dei singoli alberi. Per questo motivo si è solito definire una *random forest* come un *ensemble decision tree*.

Tipicamente, gli algoritmi *random forest* restituiscono i seguenti output:

- capacità predittiva della *random forest*, confrontando i valori delle variabile dipendente presenti nel sottoinsieme di validazione con quelli stimati a partire dalle corrispondenti variabili esplicative;
- importanza delle variabili esplicative, ossia quali sono le variabili esplicative che permettono di descrivere meglio la variabilità del dataset osservativo;
- dipendenze parziali, ossia qual'è il legame che sussiste tra la variabile dipendente e ciascuna variabile esplicativa presa singolarmente e indipendentemente dalle altre.

1.3 Normalizzazione meteorologica con random forest

La *random forest* così costruita è ora utilizzata per stimare la serie di concentrazioni normalizzate meteorologicamente. In particolare, per ogni concentrazione osservata,

la *random forest* campiona casualmente per un certo numero di volte M (definito dall'utilizzatore) le variabili esplicative associate. In corrispondenza di ciascun campionamento, l'algoritmo restituisce una previsione della concentrazione che, ovviamente, dipende dalle variabili esplicative utilizzate per generare tale valore. Al termine dei campionamenti, le M realizzazioni sono aggregate utilizzando una media aritmetica. Il valore ottenuto rappresenta la concentrazione in corrispondenza di una condizione meteorologica media.

Nel presente studio, la normalizzazione meteorologica è stata condotta a partire dalla procedura open-source implementata nel pacchetto R “**rmweather**” (Grange, 2018b). Questo algoritmo di normalizzazione meteorologica è stato recentemente applicato con successo (ed i risultati pubblicati su rivista scientifica internazionale) per valutare i trend di PM10 in Svizzera (Grange et al., 2018) e per valutare l'efficacia di politiche per il controllo del biossido di zolfo (SO₂) e degli ossidi di azoto (NO_x, NO₂) nell'area di Londra (Grange & Carslaw, 2019).

CAPITOLO 2

Datasets e caratterizzazione delle aree studio

Come precedentemente descritto, al fine di implementare la tecnica della normalizzazione meteorologica è necessario disporre di serie storiche sufficientemente estese nel tempo e complete sia dell'inquinante di interesse, sia delle variabili meteorologiche da utilizzare come proxy dei processi di trasporto e diffusione degli inquinanti.

Il presente studio è condotto per valutare l'effetto dei processi atmosferici, che localmente interessano la Provincia di Bolzano, sulle concentrazioni medie annue di biossido di azoto (NO_2). Tali concentrazioni sono misurate dalle stazioni convenzionali di qualità dell'aria gestite dall'Agenzia provinciale per l'ambiente e la tutela del clima della Provincia Autonoma di Bolzano. Coerentemente con quanto riportato nella convenzione stipulata e con gli approfondimenti richiesti durante l'incontro tenutosi il 15 marzo 2019, il periodo 2012–2018 è stato selezionato come periodo su cui condurre l'analisi. In questo capitolo, si descrivono le serie storiche di NO_2 e delle grandezze meteorologiche disponibili per il periodo investigato, unitamente con le operazioni preliminari effettuate per ottenere un dataset osservativo omogeneo (Sez. 2.1). A ciò fa seguito, per ciascuna area studio, la caratterizzazione meteoroclimatologica e delle concentrazioni di NO_2 sul periodo di riferimento 2012–2017 e il confronto con l'anno 2018 (Sez. 2.2).

2.1 Definizione dei datasets

Poiché lo studio si propone di valutare l'effetto della forzante meteorologica sulle concentrazioni medie annue, l'analisi è stata condotta aggregando tutte le informazioni disponibili alla scala giornaliera. In particolare, sono stati calcolati i valori medi giornalieri di: concentrazioni di NO_2 , intensità e direzione del vento, temperatura e umidità dell'aria, pressione atmosferica e radiazione globale. Coerentemente a ciò sono stati calcolati i profili medi giornalieri della temperatura dell'aria misu-

Tabella 2.1: Stazioni convenzionali di qualità dell’aria gestite da APPA e utilizzate durante lo studio. La tabella riporta la consistenza delle serie disponibili, espressa in termini di percentuale di medie giornaliere disponibili per ciascun anno considerato.

Stazione	ID	Quota	2012	2013	2014	2015	2016	2017	2018
Bolzano (Via C. Augusta)	BZ4	247	92.9	96.2	92.6	94.8	96.2	98.9	98.1
Bolzano (Piazza Adriano)	BZ5	260	98.6	98.9	99.2	99.5	98.6	95.3	96.2
Laives (Via Galizia)	LS1	230	96.2	97.8	97.0	97.8	96.4	96.7	95.3
Merano (Via Trogmann)	ME1	297	98.4	97.8	95.9	96.4	98.9	98.9	99.2
Bressanone (Via Stazione)	BX1	564	94.8	93.7	95.9	99.5	97.8	96.2	97.8

rati dal profilatore termico. Relativamente al dato di precipitazione, invece, si è proceduto al calcolo dell’altezza di precipitazione cumulata giornaliera.

2.1.1 Dati di qualità dell’aria

Come casi studio, sono state prese in considerazione le concentrazioni medie giornaliere di NO_2 misurate dalle stazioni di qualità dell’aria installate presso i centri abitati di:

- Bolzano (BZ4 e BZ5);
- Laives (LS1);
- Merano (ME1);
- Bressanone–Varna (BX1).

In Tabella 2.1 si elencano le stazioni di qualità dell’aria, la loro collocazione e la percentuale annua di dati medi giornalieri disponibili. Dalla Tabella emerge come la presenza di dati mancanti nelle serie storiche di concentrazioni di NO_2 non inficia la significatività statistica del dataset.

2.1.2 Dati meteorologici

Al fine di caratterizzare dal punto di vista meteo–climatologico sia i processi atmosferici in prossimità del suolo, sia la struttura termica nelle diverse aree investigate, sono stati utilizzati:

- i dati meteorologici misurati dalle stazioni al suolo appartenenti alla rete di monitoraggio gestita dall’Ufficio Idrografico della Provincia Autonoma di Bolzano e la stazione meteorologica di APPA installata sul tetto dell’edificio di Via Amba Alagi. L’analisi è stata condotta sulle serie storiche di intensità del vento, temperatura dell’aria, radiazione globale e precipitazione misurate dalle stazioni meteorologiche di fondovalle più prossime alle stazioni di qualità

Tabella 2.2: Stazioni meteorologiche al suolo, appartenenti alla rete di monitoraggio gestita dal Servizio Meteorologico della Provincia Autonoma di Bolzano, selezionate ai fini dello studio.

Stazione	ID	Lat. [N]	Long. [E]	Quota (m a.s.l.)	Posizione
Bronzolo	85700MS	46.41	11.31	226	Fondovalle
Bolzano Ospedale	83200MS	46.50	11.31	254	Fondovalle
Bolzano (Via Amba Alagi)	BZ6APPA	46.50	11.34	288	Fondovalle
Merano	23200MS	46.69	11.14	330	Fondovalle
Bressanone	39100MS	46.73	11.64	590	Fondovalle
Fiè allo Sciliar	75600MS	46.51	11.51	840	Versante
San Genesio	82910MS	46.53	11.33	970	Versante
Lauregno	90000SF	46.51	11.03	2165	Versante
Plose	69900MS	46.70	11.73	2472	Versante

dell'aria e quindi da intendersi come rappresentative dei processi meteorologici al suolo in prossimità del punto di misura degli inquinanti. Inoltre, sono state individuate stazioni meteorologiche, appartenenti alla medesima rete di monitoraggio, installate in quota sui versanti al fine di poter investigare la struttura termica dell'atmosfera. In Tabella 2.2 si elencano le stazioni meteorologiche considerate, mentre in Tab. 2.3 si riportano le serie di grandezze meteorologiche disponibili per ciascuna stazione e per ciascun anno. Anche per i dati meteorologici, si può osservare come le serie storiche delle stazioni al suolo su fondovalle siano sufficientemente complete e quindi il dataset è statisticamente rappresentativo ai fini dello studio. Relativamente alle stazioni su versante, invece, si osserva che la serie di dati di temperatura dell'aria misurata presso Lauregno è significativamente incompleta per il 2013 e il 2016, mentre quella misurata a San Genesio ha inizio nel 2015. Per questo motivo si è deciso di non utilizzare nello studio il gradiente di temperatura misurato tra il fondovalle e San Genesio.

- i profili di temperatura, misurati dal profilatore termico installato presso l'aeroporto di Bolzano e gestito da APPA. In particolare, il profilatore termico (un MTP-5HE realizzato da Attex, Russia) restituisce il profilo verticale della temperatura dell'aria nei primi 1000 metri di atmosfera. Ciascun profilo di temperatura è costituito da 21 punti, equispaziati di 50 m lungo la verticale. Il profilatore termico utilizzato nello studio, restituisce ogni 10 min la temperatura dall'aria da 250 m s.l.m. (12 m dal piano campagna) a 1250 m s.l.m.. Coerentemente con quanto riportato in precedenza, si è quindi proceduto ad omogeneizzare il dataset di profili di temperatura con le altre informazioni disponibili, calcolando il profilo di temperatura medio giornaliero. Per ulteriori informazioni in merito alle caratteristiche del profilatore termico e al suo

funzionamento, si rimanda alla letteratura scientifica di riferimento, *e.g.* Gai-kovich et al. (1992); Troitskii et al. (1993); Westwater et al. (1999); Scanzani (2010); Ezau et al. (2013); Wolf et al. (2014).

Le stazioni meteorologiche rappresentative per le stazioni di qualità dell'aria selezionate sono:

- Bolzano Ospedale, Bolzano Via Amba Alagi e Bronzolo per BZ4;
- Bolzano Ospedale, Bolzano Via Amba Alagi per BZ5;
- Bronzolo per LS1;
- Merano per ME1;
- Bressanone per BX1.

La caratterizzazione della struttura termica, utile per discriminare la stabilità atmosferica, è stata condotta valutando il *temperature lapse-rate* Γ ($^{\circ}\text{C km}^{-1}$), definito come l'opposto del gradiente termico altitudinale:

$$\Gamma = -\frac{dT}{dz} \simeq -\frac{T(z_{i+1}) - T(z_i)}{z_{i+1} - z_i}, \quad (2.1)$$

dove $T(z)$ è la temperatura dell'aria T misurata alla quota z . Questa grandezza è stata calcolata come:

- la differenza di temperatura misurata tra le stazioni installate sul versante e la stazione di fondovalle di riferimento diviso la differenza di quota;
- il coefficiente angolare della retta che meglio interpola i profili verticali di temperatura misurati dal profilatore termico nei primi 100 m e 500 m di atmosfera al di sopra del suolo.

Nel primo caso, Γ è stato stimato tra le stazioni di:

- Lauregno–Merano (dislivello: 1835 m) per ME1;
- Fiè allo Sciliar–Bressanone (dislivello: 250 m) e Plose–Bressanone (dislivello: 1882 m) per BX1,

mentre i valori di Γ stimati attraverso il profilatore termico sono stati utilizzati per caratterizzare la struttura termica per le stazioni di qualità dell'aria di Bolzano (BZ4 e BZ5), Laives (LS1) e Merano (ME1). Infatti, tutte queste stazioni giacciono sul fondovalle della medesima valle ed è quindi lecito attendersi condizioni simili di stratificazione atmosferica in prossimità del suolo. Questa ipotesi, invece, è meno plausibile per il sito di Bressanone.

Tabella 2.3: Consistenza delle serie di intensità del vento (WG), direzione del vento (WR), temperatura dell'aria (LT), radiazione globale (GS) e di precipitazione (N) per ciascun anno del periodo selezionato (2012–2018) e per ciascuna stazione al suolo selezionata (Tab. 2.2). Si indica con “x” una consistenza superiore al 50%, con “o” una consistenza inferiore al 50% mentre l'assenza di simbolo indica l'assenza della serie storica.

Stazione	Variabile	2012	2013	2014	2015	2016	2017	2018
Bronzolo (85700MS)	WG	x	x	x	x	x	x	x
	WR	x	x	x	x	x	x	x
	LT	x	x	x	x	x	x	x
	GS	x	x	x	x	x	x	x
	N	x	x	x	x	x	x	x
Bolzano Ospedale (83200MS)	WG	x	x	x	x	x	x	x
	WR	x	x	x	x	x	x	x
	LT	x	x	x	x	x	x	x
	GS	x	x	x	x	x	x	x
	N	x	x	x	x	x	x	x
Bolzano Via Amba Alagi (BZ6APPA)	WG	x	x	x	x	x	x	x
	WR	x	x	x	x	x	x	x
	LT							
	GS	x	x	x	x	x	x	x
	N	x	x	x	x	x	x	x
Merano (23200MS)	WG	x	x	x	x	x	x	x
	WR	x	x	x	x	x	x	x
	LT	x	x	x	x	x	x	x
	GS	x	x	x	x	x	x	x
	N	x	x	x	x	x	x	x
Bressanone (39100MS)	WG	x	x	x	x	x	x	x
	WR	x	x	x	x	x	x	x
	LT	x	x	x	x	x	x	x
	GS	x	x	x	x	x	x	x
	N	x	x	x	x	x	x	x
Fiè allo Sciliar (75600MS)	LT	x	x	x	x	x	x	x
San Genesio (82910MS)	LT				o	x	x	x
Lauregno (90000SF)	LT		o	x	x	o	x	x
Plose (69900MS)	LT	x	x	x	x	x	x	x

2.2 Caratterizzazione meteo-climatologica e delle concentrazioni di NO₂

Come anticipato, la caratterizzazione meteo-climatologica e delle concentrazioni di NO₂ per ciascuna stazione di qualità dell'aria è stata condotta analizzando le serie di precipitazione, radiazione globale, temperatura dell'aria, intensità del vento, di *temperature lapse-rate* e di concentrazioni di NO₂ per il periodo 2012–2017, selezionato come periodo di riferimento, e per l'anno 2018, selezionato come anno di confronto.

Tabella 2.4: Altezze di precipitazione cumulate annue per le stazioni meteorologiche selezionate e statistica dei valori annui per il periodo 2012-2017 al fine di favorire il confronto con l'anno 2018.

	Bolzano Ospedale 83200MS	Bolzano Via Amba Alagi BZ6APPA	Bronzolo 85700MS	Merano 23200MS	Bressanone 39100MS
2012		695.6			
2013	932.1	841.8	824.0	909.7	867.8
2014	1064.9	1043.0	962.9	1052.1	882.5
2015	524.4	399.8	519.4	626.1	621.9
2016	709.0	533.0	675.7	838.6	809.0
2017	796.4	805.0	705.9	850.0	703.5
minimo	524.4	399.8	519.4	626.1	621.9
massimo	1064.9	1043.0	962.9	1052.1	882.5
media	805.4	722.9	737.6	855.3	776.9
deviazione standard	207.2	210.1	166.3	153.7	111.6
2018	717.7	742.4	864.1	755.8	722.0

Precipitazione. La caratterizzazione delle precipitazioni nel periodo investigato (Fig. 2.1) evidenzia un picco di precipitazione nei mesi estivi, coerentemente con la caratterizzazione climatologica delle precipitazioni dell'area in esame. Le distribuzioni ottenute mostrano una notevole variabilità delle altezze di precipitazione durante i diversi anni che però è indipendente dalla stazione meteorologica presa in esame. Ciò implica che la variabilità delle altezze di precipitazione mensili osservate è da attribuire maggiormente alla variabilità annuale delle precipitazioni, legate quindi alle circolazioni a scala sinottica e non ai soli effetti locali legati, soprattutto nei mesi estivi, alla presenza di temporali. Infatti, dall'analisi delle altezze di precipitazione cumulate annue (Tab. 2.4), si osserva come il 2014 ed il 2015 siano stati, rispettivamente, l'anno più piovoso e l'anno più secco, per tutte le stazioni investigate. Il 2018, invece, si configura come un anno con una piovosità intermedia.

Radiazione globale. L'analisi della radiazione globale (Fig. 2.2) è stata condotta in termini di cicli medio mensili. Nei grafici di Fig. 2.2 si riporta quindi il ciclo medio mensile osservato nel periodo 2012–2017 (tratti neri) con le relative deviazioni standard (bande azzurre) a cui è sovrapposta la distribuzione mensile dei valori di radiazione globale misurati durante il 2018 (linea rossa nei grafici). Dal confronto emerge che la distribuzione delle medie mensili ottenuta per il 2018 è pressoché sempre contenuta nella distribuzione media \pm deviazione standard stimata per il periodo 2012-2017, per tutte le variabili e tutte le stazioni. Ciò implica che il 2018 è stato caratterizzato da una forzante meteorologica simile a quella osservata nel periodo 2012–2017.

Temperatura dell'aria. Le distribuzioni mensili di temperatura dell'aria, rappresentate in modo analogo al dato di radiazione globale, mostrano come in tutti i siti il 2018 sia stato caratterizzato da temperature dell'aria nella media del periodo di riferimento 2012–2017, fatta eccezione dei mesi di febbraio e dicembre 2018, invece più freddi.

Temperature lapse-rate. Come introdotto, la stratificazione termica è stata valutata sulla base del *temperature lapse-rate* Γ , definito in (2.1), utilizzando la differenza di temperatura tra due stazioni meteorologiche e il profilo di temperatura misurato dal profilatore termico installato presso l'aeroporto di Bolzano. Nel primo caso, differenza tra due stazioni, il valore di Γ ottenuto, oltre che essere rappresentativo del singolo mese, è da considerarsi come rappresentativo dell'atmosfera tra le due stazioni e quindi dipende dal dislivello tra i due punti di misura. Viceversa, la stima di Γ effettuata con il profilatore termico, essendo limitata al profilo di temperatura nei primi 100 m e 500 m di atmosfera, è da considerarsi rappresentativa del campo termico in prossimità del suolo. Le distribuzioni medie mensili riportate in Fig. 2.4 mostrano che il *temperature lapse-rate* è pressoché sempre positivo durante l'anno e tende ad annullarsi o assume valori negativi nei mesi invernali. In particolare, valori negativi sono apprezzabili solamente per le distribuzioni ottenute dal profilatore termico nei primi 100 m di atmosfera e confrontando i dati di Fié allo Sciliar con Bressanone. Infatti, in questi casi, il ridotto spessore di atmosfera investigato permette di meglio apprezzare la presenza di inversioni termiche in prossimità del suolo. In particolare, i valori di Γ ottenuti per la stazione di qualità dell'aria BX1 con le temperature misurate tra Fié allo Sciliar (75600MS) e Bressanone (69900MS) mostrano una grande variabilità. Questo comportamento è attribuito alla diversa collocazione tra le due stazioni al suolo, in quanto di Fié allo Sciliar non si trova nelle immediate vicinanze di quella di Bressanone.

Le distribuzioni medie mensili stimate a partire dal profilatore termico (Fig. 2.4), invece, mostrano valori di Γ inferiori. Soprattutto in autunno ed in inverno, i valori di Γ sono mediamente negativi ed indicano come i primi 100 m di atmosfera siano sede di importanti inversioni termiche al suolo.

Concentrandosi sui valori di Γ stimati attraverso i dati del profilatore termico, si osserva che, mediamente, il *temperature lapse-rate* per l'anno 2018 è superiore a quello del periodo di riferimento nei mesi freddi (febbraio, marzo e novembre). Ciò lascia presupporre che il 2018 sia stato caratterizzato da un minor numero di inversioni termiche al suolo e/o da inversioni termiche al suolo meno intense rispetto a quelle osservate nel periodo di riferimento. Le distribuzioni dei Γ stimati con le stazioni al suolo, e quindi rappresentative dell'atmosfera all'interno delle diverse valli, confermano quanto osservato per i Γ del profilatore termico. Ossia che durante

il 2018 la colonna d'aria nelle diverse valli sia stata mediamente meno stabile che non durante il periodo di riferimento, inibendo così lo sviluppo di inversioni termiche al suolo.

Intensità del vento. Differentemente dalle precedenti grandezze meteorologiche, la caratterizzazione dell'anemometria per i siti investigati è stata condotta su base stagionale. Infatti, si è proceduto a calcolare sia per il periodo di riferimento (2012–2017) sia per il periodo di confronto (2018), la distribuzione in frequenza delle velocità del vento per le stazioni di Bronzolo (Fig. 2.5), Bolzano Ospedale (Fig. 2.6), Bolzano Via Amba Alagi (Fig. 2.7), Merano (Fig. 2.8) e Bressanone (Fig. 2.9). L'analisi permette di osservare come le distribuzioni ottenute per il 2018 siano pressoché confrontabili con quelle calcolate per il periodo di riferimento 2012–2017. In particolare, la stazione di Bronzolo ha registrato venti più intensi durante la primavera 2018 e meno intensi nell'autunno dello stesso anno. A Merano, invece, primavera e autunno 2018 sono stati caratterizzati da un regime dei venti più debole rispetto al periodo di riferimento. Una debole intensificazione dei venti si osserva nei mesi estivi del 2018 presso le stazioni di Bolzano Via Amba Alagi e Bressanone.

Concentrazioni di NO₂. Tutte le stazioni di qualità dell'aria mostrano un marcato ciclo annuale delle concentrazioni di NO₂, in cui i massimi annui si raggiungono nel periodo invernale, mentre i minimi nel periodo estivo. L'evoluzione temporale delle medie mensili per il periodo di riferimento (2012–2017) e per l'anno di confronto (Fig. 2.10) mostra che mediamente i massimi annui registrati per il 2018 sono inferiori o uguali a quelli osservati negli anni precedenti, mentre i minimi appaiono essere tra loro confrontabili. In particolare nei mesi tra autunno e inverno 2018 le concentrazioni di NO₂ appaiono più contenute rispetto a quanto misurato negli anni precedenti. Questo effetto potrebbe essere legato al persistere di condizioni con venti intensi osservate in questi mesi, che quindi hanno contribuito al miscelamento dell'atmosfera.

Al fine di poter meglio apprezzare questi comportamenti, in analogia con l'analisi delle intensità dei venti, si è proceduto a rappresentare la distribuzione di frequenza stagionale delle concentrazioni di NO₂ per ciascuna stazione di qualità dell'aria: Fig. 2.11 per BZ4, Fig. 2.12 per BZ5, Fig. 2.13 per LS1, Fig. 2.14 per ME1, e Fig. 2.15 per BX1. Come risultato, si osserva che le distribuzioni in frequenza delle concentrazioni misurate nel 2018 tendono ad assumere valori di picco in corrispondenza di concentrazioni inferiori, soprattutto nei mesi invernali ed autunnali.

I risultati ad ora ottenuti non permettono però di valutare se la diminuzione delle concentrazioni di NO₂ osservata nell'anno 2018 è legata ad una effettiva diminu-

zione delle emissioni e/o all'effetto delle forzanti meteorologiche. Infatti, il periodo invernale è stato caratterizzato da una minor presenza di inversioni termiche al suolo, responsabili della stagnazione degli inquinanti nei bassi strati di atmosfera e da una maggiore intensità dei venti, che quindi può aver contribuito a disperdere con maggiore efficienza gli NO_2 .

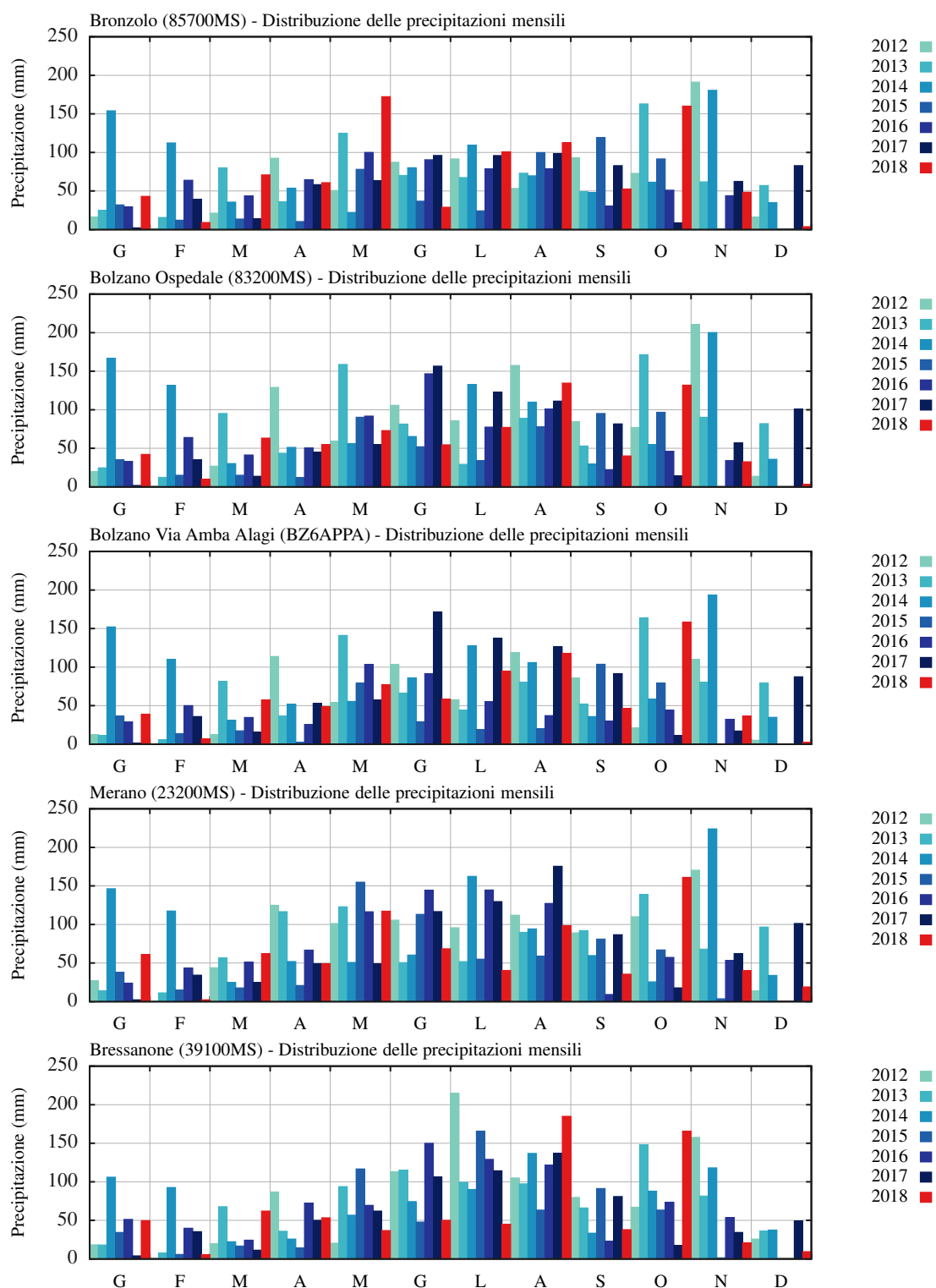


Figura 2.1: Caratterizzazione del regime delle precipitazioni a scala mensile riferita al periodo 2012–2018 per le stazioni di Bronzolo, Bolzano Ospedale, Bolzano Via Amba Alagi, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

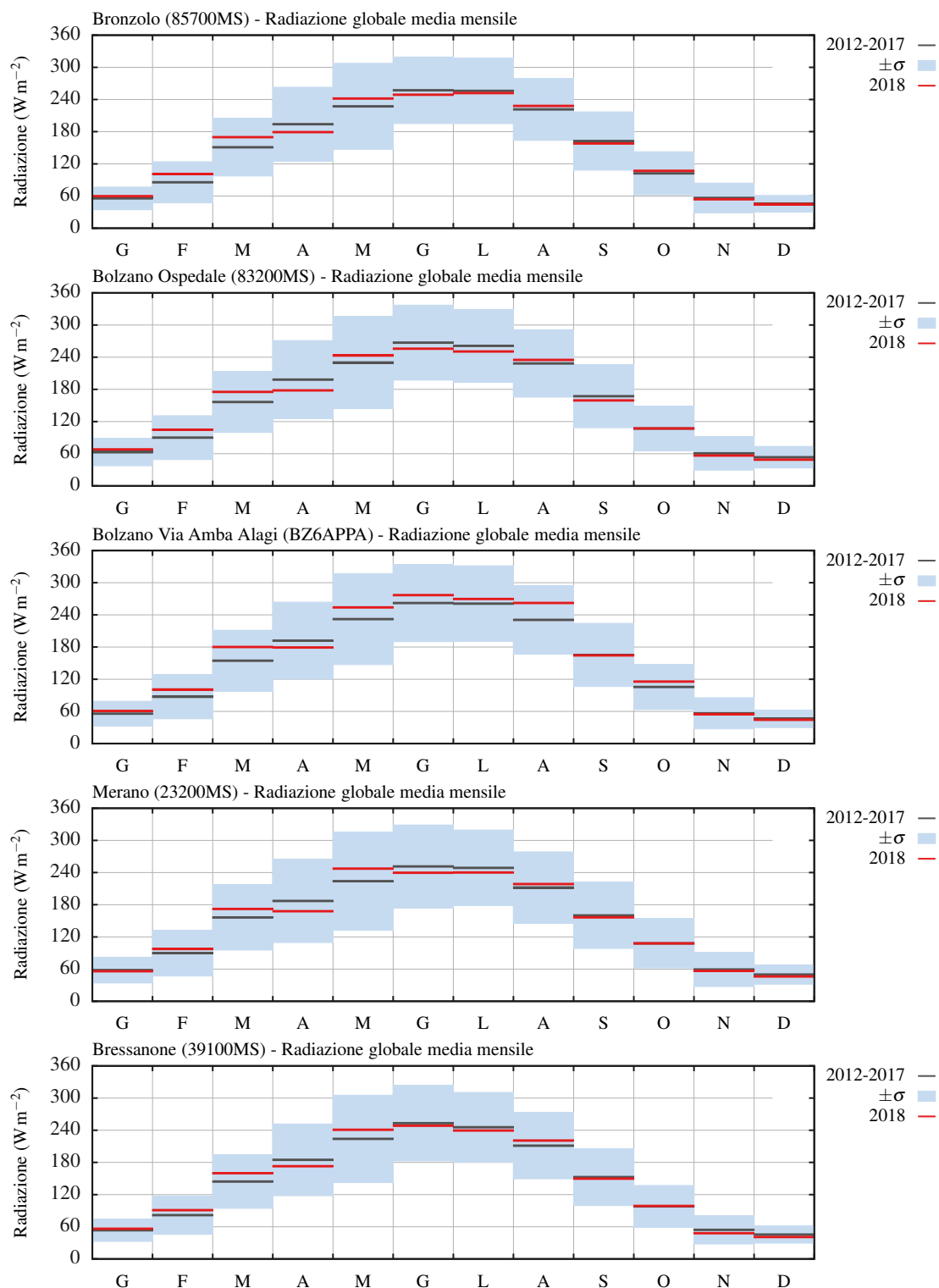


Figura 2.2: Cicli mensili della radiazione globale misurati presso le stazioni di Bronzolo, Bolzano Ospedale, Bolzano Via Amba Alagi, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

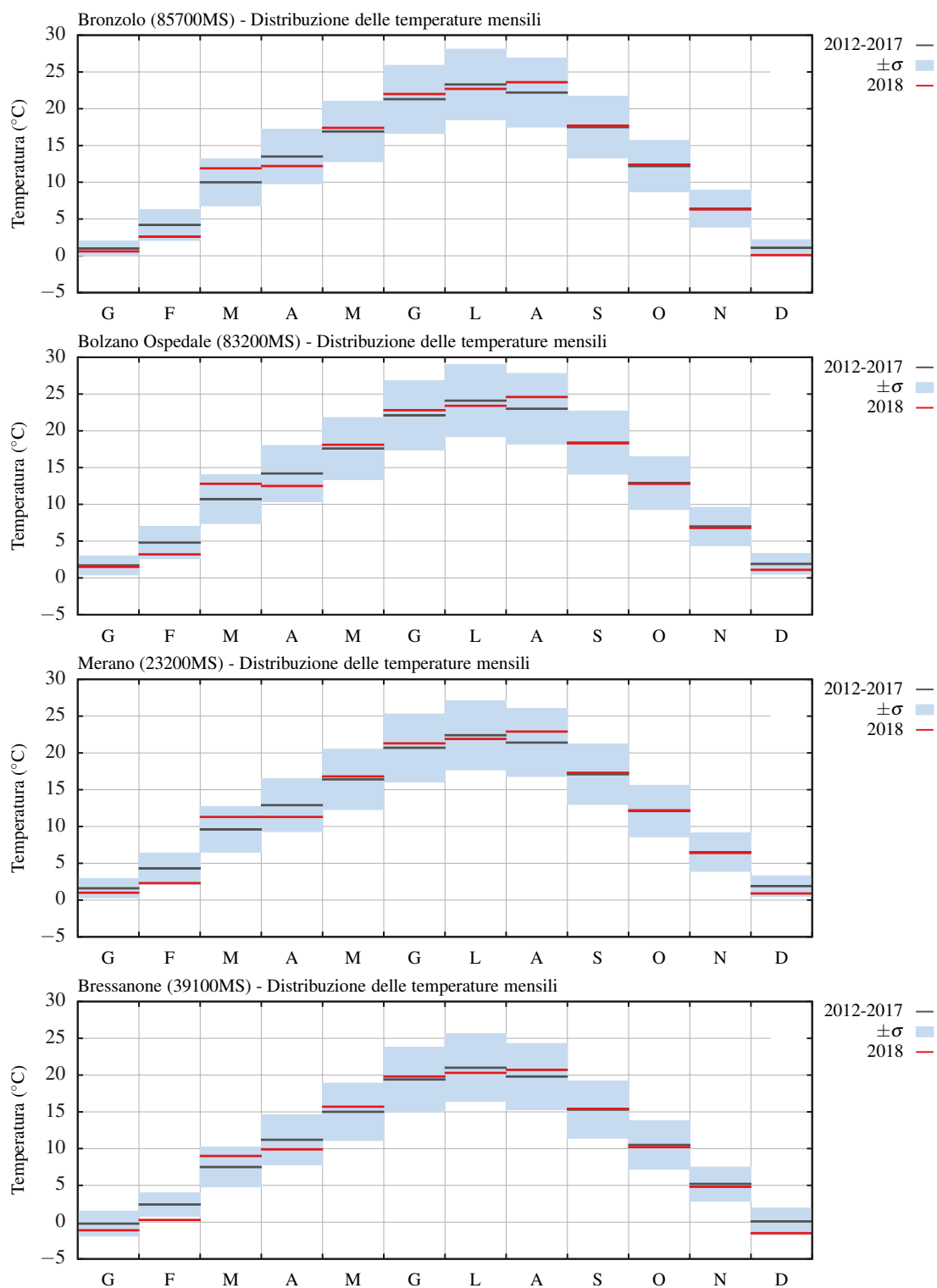


Figura 2.3: Cicli mensili della temperatura dell'aria misurati presso le stazioni di Bronzolo, Bolzano Ospedale, Merano e Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

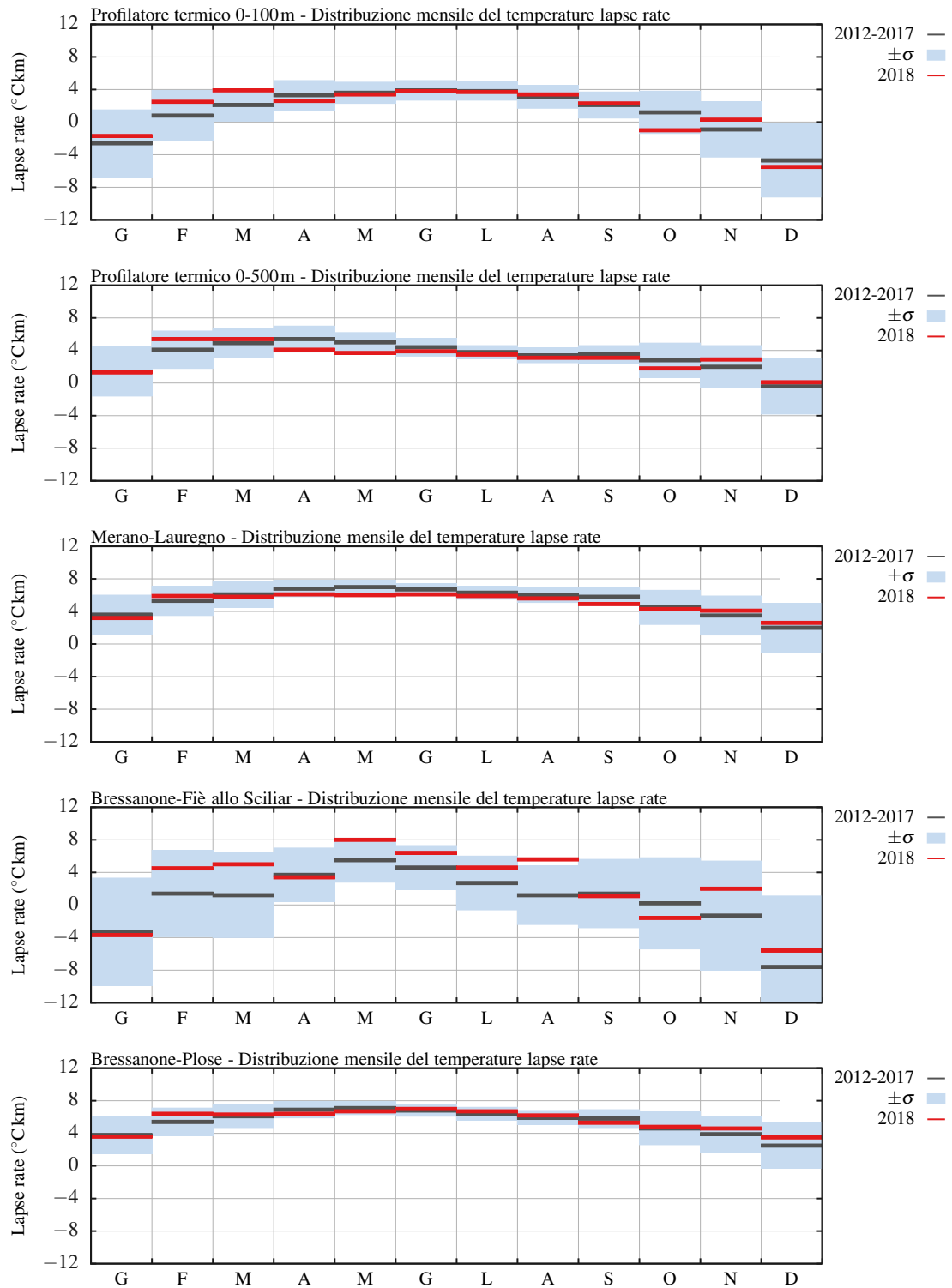


Figura 2.4: Cicli mensili del *temperature lapse rate* misurati dal profilatore termico nei primi 100 m e 500 m di atmosfera dal suolo, tra la stazione di Merano e Lauregno, tra la stazione di Bressanone e Fiè allo Sciliar e tra Bressanone e Plose. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

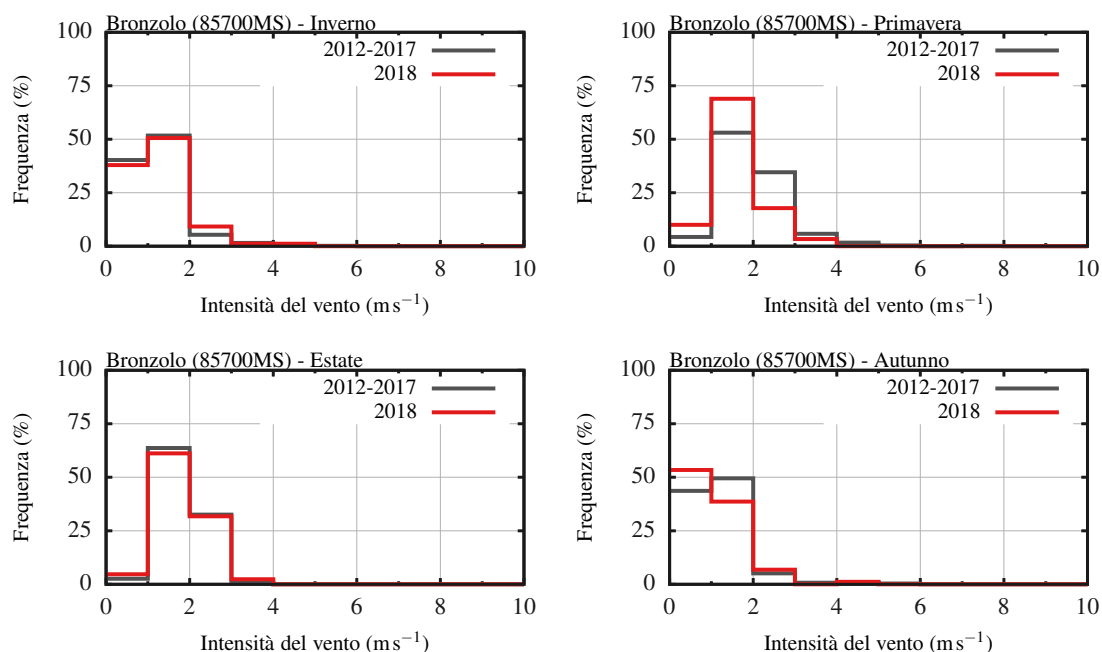


Figura 2.5: Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Bronzolo. Periodo di riferimento: 2012-2017, anno di controllo: 2018 (in rosso).

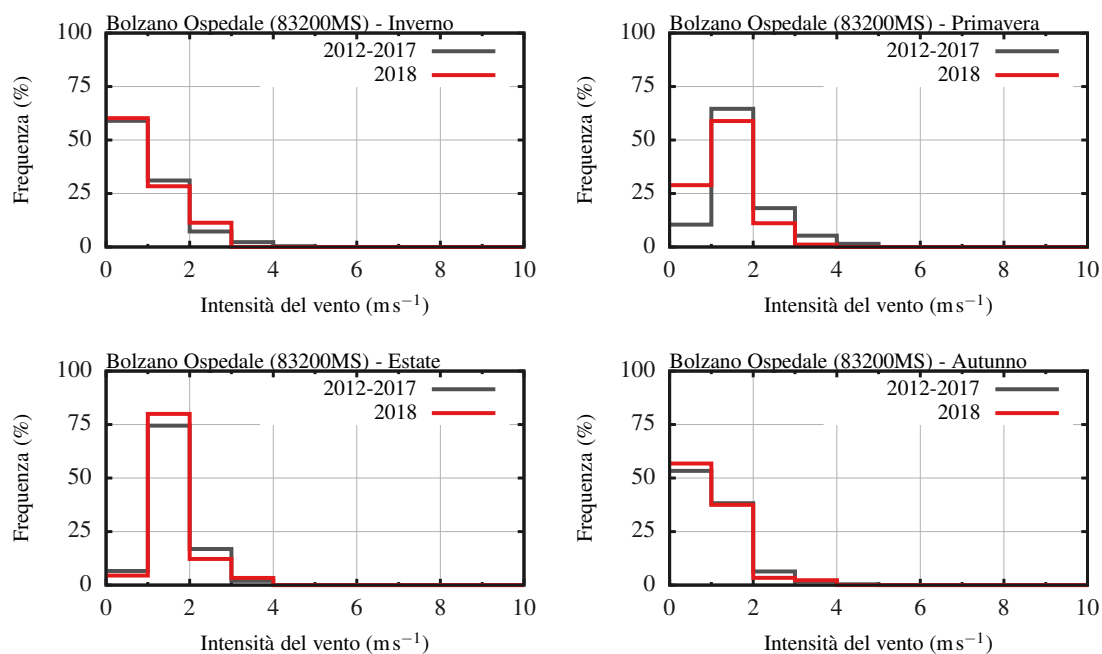


Figura 2.6: Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Bolzano Ospedale. Periodo di riferimento: 2012-2017, anno di controllo: 2018 (in rosso).

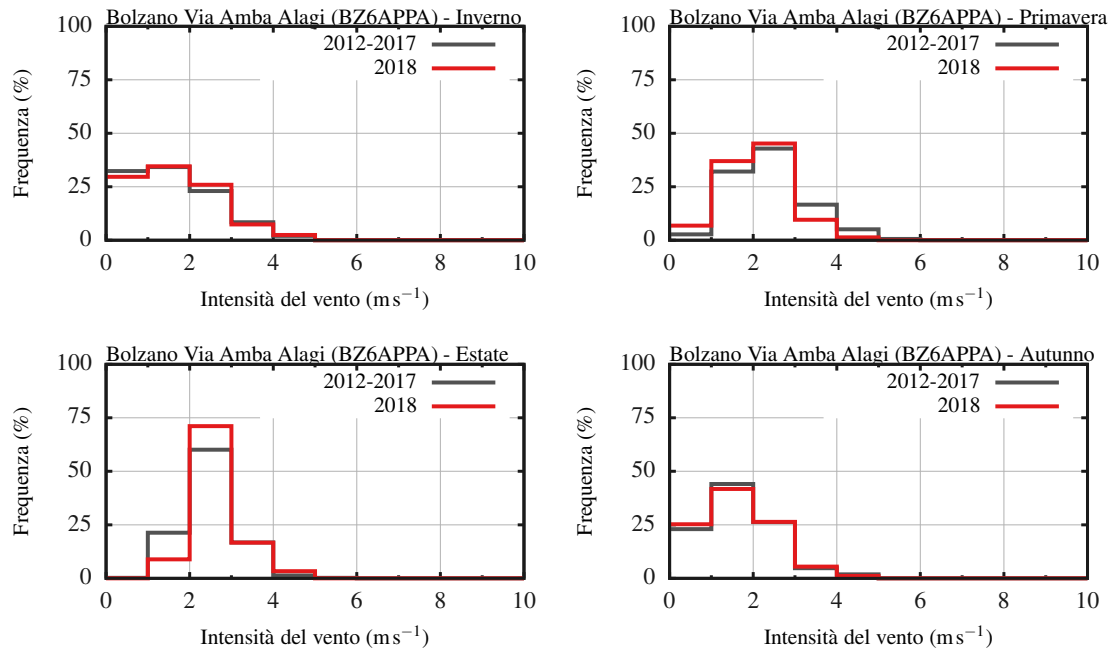


Figura 2.7: Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Bolzano Via Amba Alagi. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

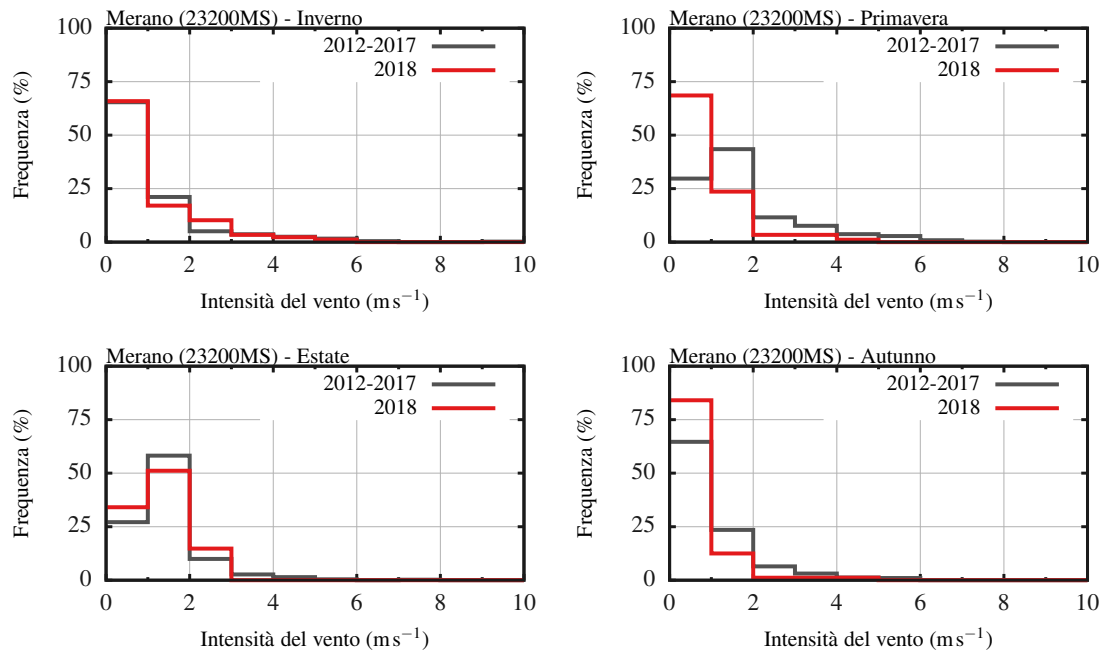


Figura 2.8: Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Merano. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

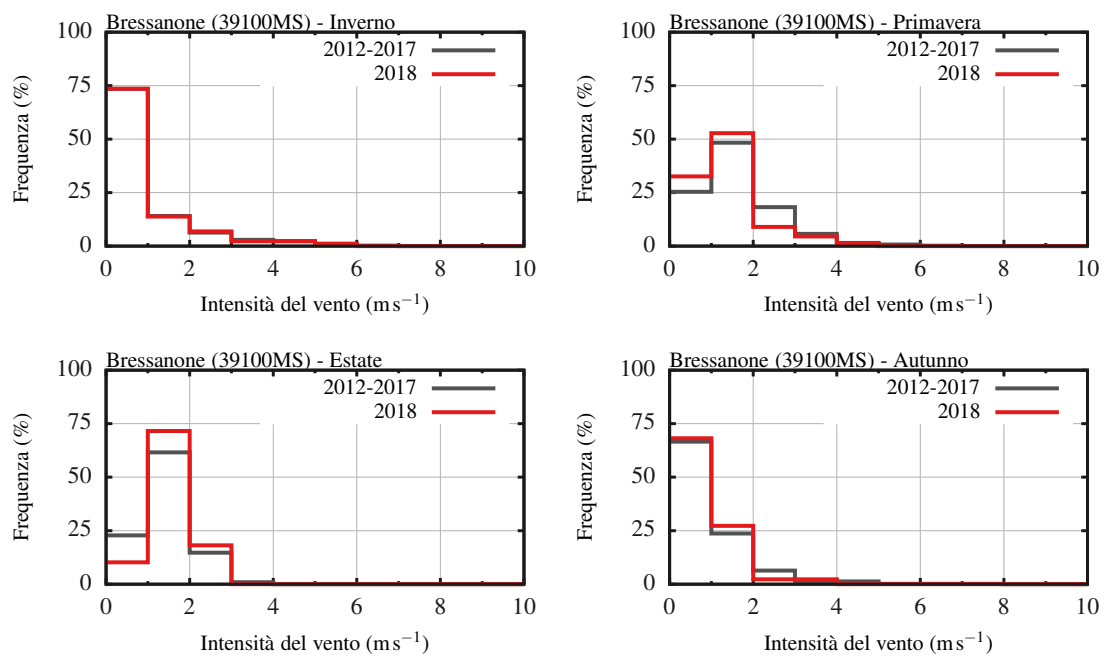


Figura 2.9: Distribuzioni in frequenza dell'intensità del vento a scala stagionale per la stazione di Bressanone. Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

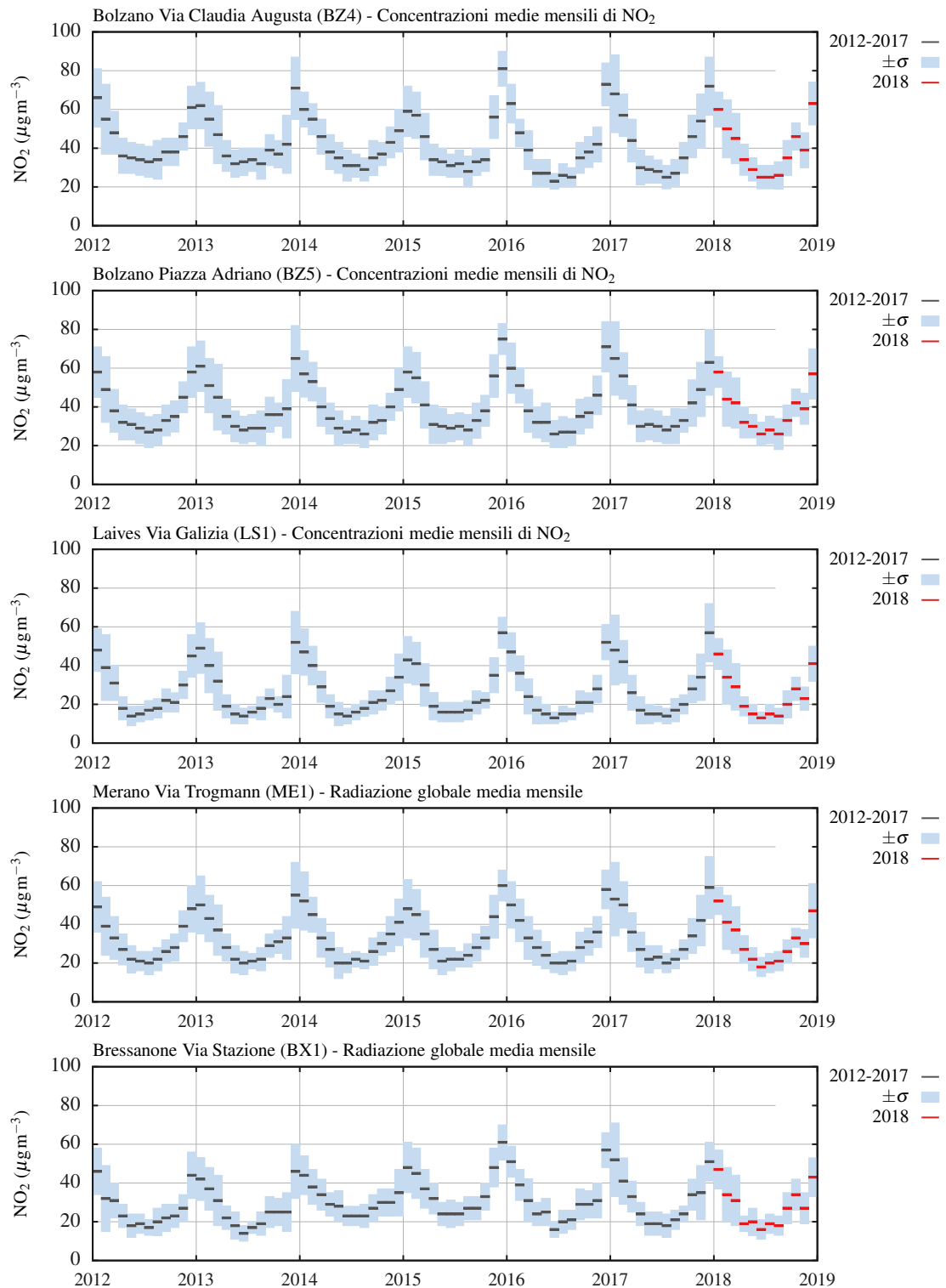


Figura 2.10: Cicli mensili di concentrazioni di NO₂ misurati dalle stazioni di Bolzano Via Claudia Augusta (BZ4), Bolzano Piazza Adriano (BZ5), Laives (LS1), Merano (ME1) e Bressanone (BX1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

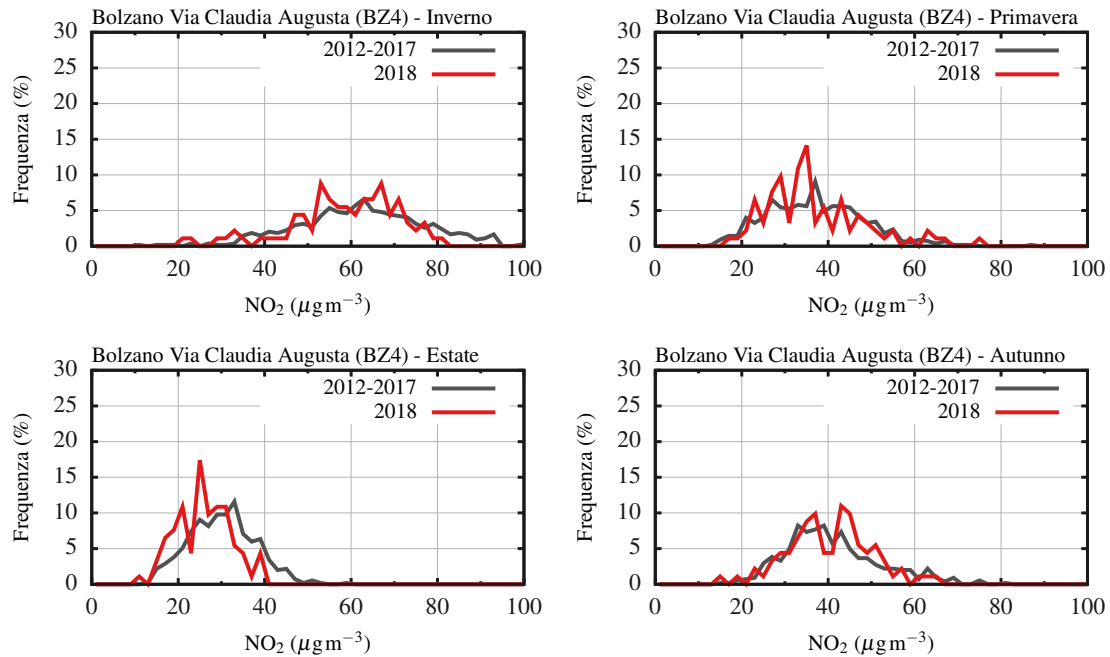


Figura 2.11: Distribuzioni in frequenza delle concentrazioni di NO₂ a scala stagionale per la stazione di Bolzano Via Claudia Augusta (BZ4). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

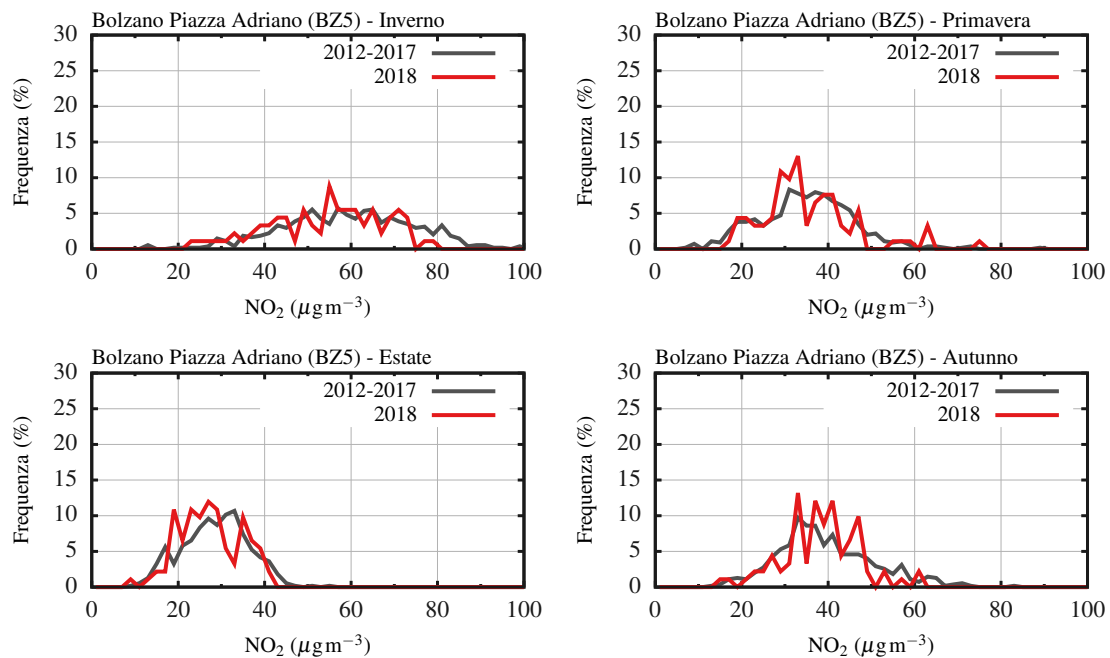


Figura 2.12: Distribuzioni in frequenza delle concentrazioni di NO₂ a scala stagionale per la stazione di Bolzano Piazza Adriano (BZ5). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

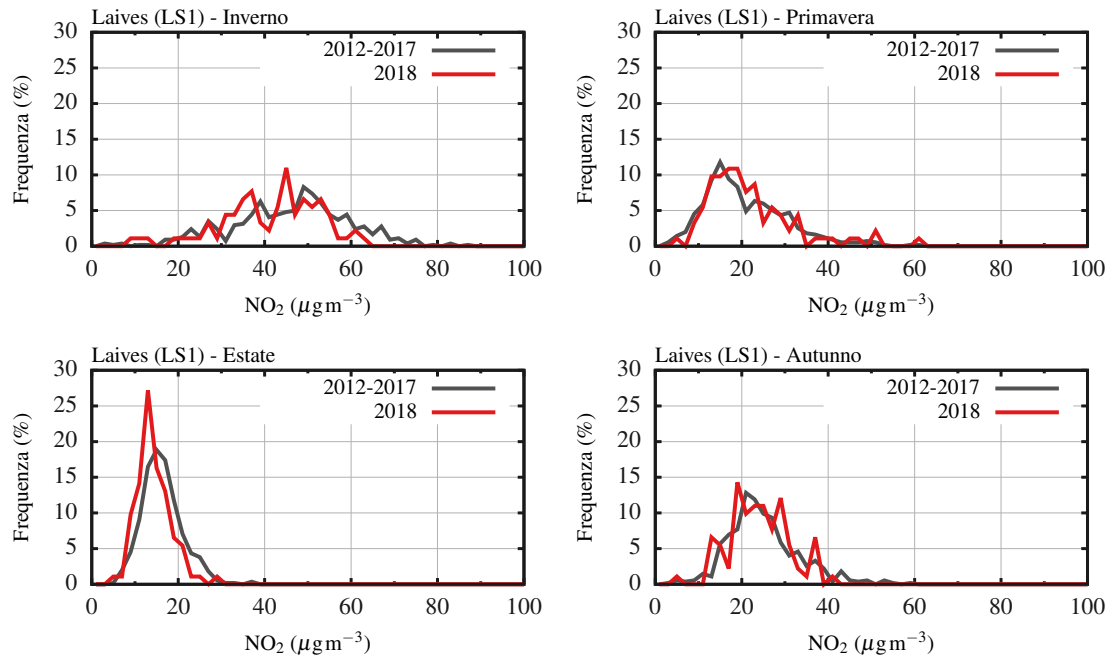


Figura 2.13: Distribuzioni in frequenza delle concentrazioni di NO₂ a scala stagionale per la stazione di Laives (LS1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

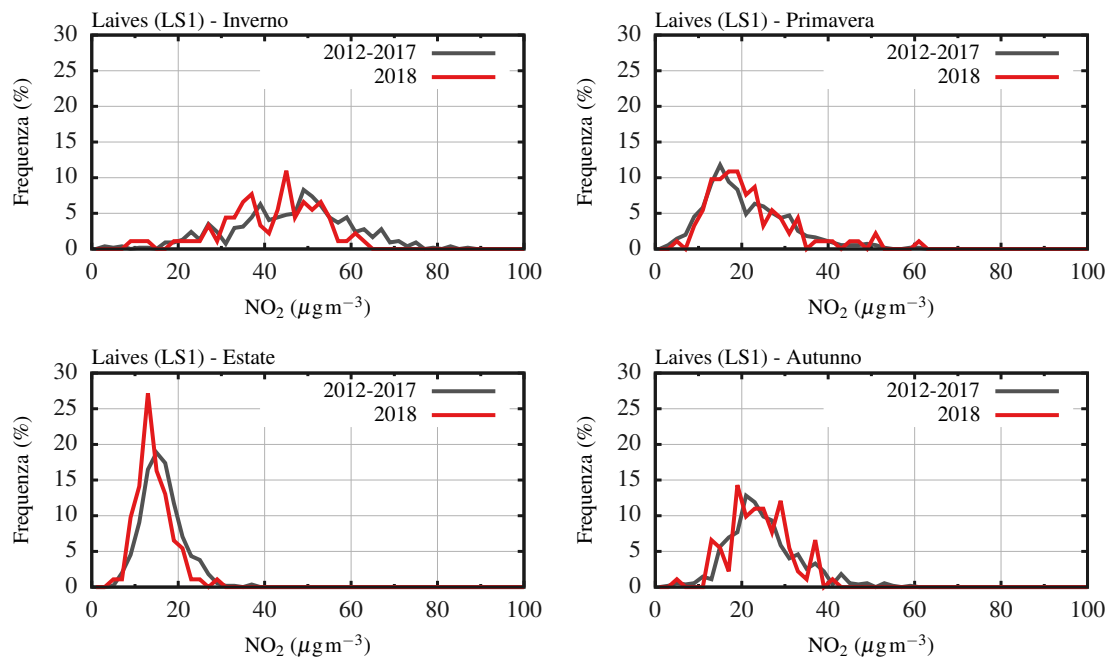


Figura 2.14: Distribuzioni in frequenza delle concentrazioni di NO₂ a scala stagionale per la stazione di Merano (ME1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

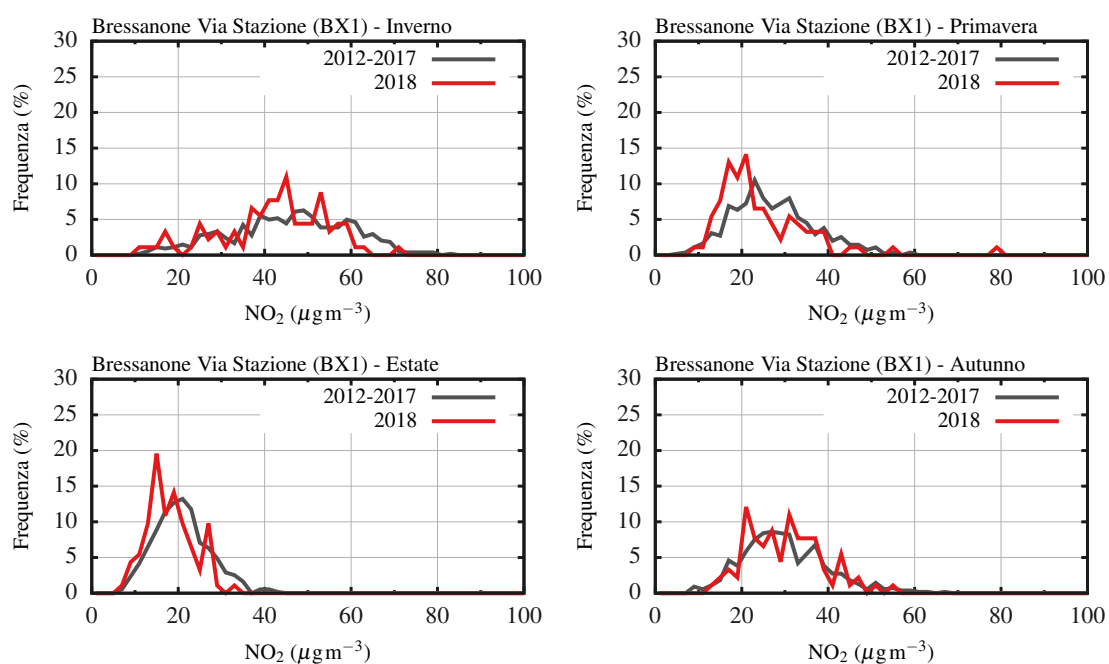


Figura 2.15: Distribuzioni in frequenza delle concentrazioni di NO₂ a scala stagionale per la stazione di Bressanone (BX1). Periodo di riferimento: 2012–2017, anno di controllo: 2018 (in rosso).

CAPITOLO 3

Implementazione dell'algoritmo di normalizzazione meteorologica

Coerentemente con quanto riportato nel Capitolo 1, in questo studio, la normalizzazione meteorologica (NM) delle serie di NO_2 è stata eseguita implementando un algoritmo di classificazione dei dati basato sulla tecnica di *machine learning* nota come *random forest* (nel seguito RM). In particolare, si è fatto utilizzo del pacchetto open-source **rmweather** (Grange, 2018b), sviluppato in ambiente **R** (R Core Team, 2017) per condurre la NM di dati di qualità dell'aria.

Nel presente Capitolo si dettaglia la procedura implementata e la struttura di cartelle e di files che deve essere garantita al fine di effettuare la NM per le stazioni di qualità dell'aria descritte al Cap. 2.

3.1 Organizzazione della cartella di lavoro

La cartella di lavoro contenente gli scripts, i dati e i risultati della NM è la cartella **NO2Normalization**. Da un punto di vista generale, si consideri una generica stazione di qualità dell'aria, denominata XX1. Al fine di effettuare la NM, è necessario disporre dei seguenti files e cartelle:

- cartella **db**: contiene i file di input necessari per effettuare la NM. La descrizione dei file che costituiscono il dataset osservativo disponibile e la formattazione dei dati, sono descritte in Sez. 3.2.1;
- cartella **scripts**: raccoglie, in diversi files scritti in linguaggio **R**, le istruzioni e gli algoritmi necessari per effettuare la NM. Per la descrizione dei files contenuti nella cartella si rimanda alla Sez. 3.3.2.
- cartella **aqsXX1**: raccoglie i file di output della NM per la stazione di qualità dell'aria (air quality station) XX1. Come descritto in Sez. 3.3.1 la procedura

di NM meteorologica implementata permette di effettuare l'analisi utilizzando il medesimo dataset osservativo, ma diversificando la porzione di dati su cui “far crescere” la RF e su cui validarne la capacità predittiva. In base a queste scelte dell'utilizzatore, l'algoritmo crea all'interno di questa cartella una sotto-cartella contenete gli output della NM. Quest'ultima cartella informa sul periodo utilizzato per effettuare il training delle RF e su quale periodo è effettuata la NM. As esempio, la cartella `train_2012_2017_norm_2018` indica che la RF è stata costruita sul periodo 2013–2017 (`train`), mentre la normalizzazione meteorologica (`norm`) è stata condotta sul solo anno 2018. Per la descrizione dei file di output si rimanda alla Sez. 3.4.

- `settings_XX1.r`: file che contiene tutti i parametri utili per effettuare la NM. Il nome di questo file contiene l'acronimo della stazione di qualità dell'aria di riferimento. Questo file può essere modificato, all'occorrenza, dall'utilizzatore sulla base di quanto descritto in Sez. 3.2.2.
- `N02Normalization_24h_XX1.r`: file che contiene le istruzioni e gli algoritmi in linguaggio **R**. Sulla base dei parametri forniti nel file di setting e l'informazione osservativa disponibile, la procedura implementa la RF ed effettua la NM così come descritto in Sez. 3.3.1. Il nome del file è composto da tre stringhe di caratteri che richiamano: il nome della cartella di lavoro, la risoluzione temporale dei dati di input (in questo caso 24 ore) e l'acronimo della stazione di qualità dell'aria di riferimento. Questo file non deve essere modificato, ma solo eseguito per avviare la NM con il comando:

```
Rscript N02Normalization_24h_XX1.r;
```

3.2 Files di input

3.2.1 Cartella: db

La cartella raccoglie di files contenenti i dati disponibili per effettuare la NM. Nel presente studio, come descritto al Cap. 2, si è fatto uso di dati medi giornalieri di: grandezze meteorologiche al suolo, concentrazioni di NO₂ e profili di temperatura. Tutti i files dei dati devono essere strutturati come segue:

- riga 1: intestazione.
- riga 2 e successive: ciascuna riga contiene nella prima colonna l'informazione temporale, mentre nelle colonne successive i valori delle variabili di interesse. In particolare, per i dati misurati dalle stazioni meteorologiche al suolo e per

i profili di temperatura è sufficiente fornire la data in formato YYYY-MM-DD, mentre per i dati di concentrazione di inquinante è necessario fornire anche l'ora con formato YYYY-MM-DD HH:MM. In quest'ultimo caso l'ora deve essere fornita come ora solare (UTC+1) e riferita alla mezzanotte del giorno di interesse.

- l'informazione temporale deve essere continua anche qualora il dato delle variabile di interesse sia assente. Ciò implica che in ciascun file sono presenti 365 righe di dati (366 con anni bisestili).
- l'assenza di dato è indicata sostituendo al dato mancante una stringa composta da tre trattini ---.
- le colonne devono essere tra loro separate da un `tab` (*tab separated file*).

Grandezze meteorologiche al suolo. Il nome del file deve essere fornito come segue:

[ID stazione]_[anno]_[risoluzione].dat

E.g. il file 23200MS_2013_av1d.dat contiene i dati medi giornalieri (av1d) misurati dalla stazione di Merano (23200MS) durante l'anno 2013.

L'intestazione della prima colonna (riferimento temporale) è `date`, mentre nelle restanti colonne si riporta l'intestazione delle grandezze meteorologiche nel formato:

[ID stazione][ID grandezza meteo]

con il seguente ordine: intensità del vento (WG), direzione del vento (WR), temperatura dell'aria (LT), umidità relativa (LF), pressione atmosferica (LD), radiazione globale (GS), soleggiamento (SD) e precipitazione (N).

Profili di temperatura. Il nome del file deve essere fornito come segue:

mtp5_[anno]_[risoluzione].dat

E.g. il file mtp5_2013_av1d.dat contiene i profili di temperatura medi giornalieri (av1d) misurati dal profilatore termico (mtp5) durante l'anno 2013.

L'intestazione della prima colonna (riferimento temporale) è `Date`, mentre nelle restanti colonne si riporta l'intestazione delle grandezze meteorologiche nel formato:

[T][quota da elevazione zero]

ossia T0 per le temperature misurate alla quota del sensore (250 m s.l.m.), T50 per le temperature misurate 50 m al di sopra del sensore (300 m s.l.m.) e così le seguenti fino a T1000, ossia temperature misurate 1000 m al di sopra del sensore (1250 m s.l.m.).

Concentrazioni di NO₂. Il nome del file deve essere fornito come segue:

`[no2]_[anno]_[risoluzione].dat`

E.g. il file `no2_2013_av1d.dat` contiene le concentrazioni medie giornaliere (`av1d`) di biossido di azoto (`no2`) misurate da tutte le stazioni di qualità dell'aria considerate nello studio durante l'anno 2013.

L'intestazione della prima colonna (riferimento temporale) è `DateTime`, mentre nelle restanti colonne si riporta l'acronimo delle stazioni di qualità dell'aria nel seguente ordine: Bolzano BZ4 e BZ5, Laives LS1, Merano ME1 e BressanoneBX1.

3.2.2 File: `settings_XX1.r`

Come anticipato, il file `settings_XX1.r` contiene i parametri che possono essere modificati dall'utente al fine di personalizzare l'esecuzione della NM. Al fine di rendere la procedura di più semplice utilizzo, tutti i passaggi ed i parametri sono anticipati da un commento che ne descrive il funzionamento. Nel seguito si descrive la struttura del file, i suoi contenuti ed i parametri che possono essere modificati dall'utilizzatore.

Study area L'utilizzatore deve definire l'acronimo da utilizzare per identificare l'area di studio e/o la stazione di qualità dell'aria di riferimento, che viene assegnato alla variabile `area`. Questo acronimo coincide con il nome della cartella in cui verranno riportati i file di output. Per la stazione di qualità dell'aria `XX1`, ad esempio il comando sarà:

```
area <- "aqsXX1"
```

Data set (ds) characteristics In questa sezione del file si definiscono le caratteristiche generali del dataset disponibile e come la procedura scritta possa raggiungere i file dei dati. I comandi, che non devono essere necessariamente modificati dall'utilizzatore sono:

- risoluzione temporale dei dati osservati in minuti, che nel presente studio è pari a un giorno:
`ds_dt <- 1440`
- stringa di testo che sostituisce dati assenti o non rappresentativi:
`ds_nan <- "---"`
- acronimo utilizzato nel nome del file di input per identificare la risoluzione temporale:
`ds_av <- "_av1d"`

- estensione dei file di input:

```
ds_ext <- ".dat"
```

Air quality stations (aqs) Si assegnano qui:

- l'acronimo della specie di inquinante da normalizzare che compare nel file di input:

```
aqs_ds <- "no2"
```

- i metadata relativi alla stazione di qualità dell'aria, forniti riempiendo i campi del vettore sotto riportato e rispettando il seguente ordine: a) ID stazione (*e.g.* XX1); b) nome della stazione; c) quota della stazione; d) ID dell'inquinante come riportato nell'intestazione del file di input; e) unità di misura; f) risoluzione temporale in minuti (nel caso di dati giornalieri 1440).

```
aqs_info <- c("XX1", "[b]", [c], "[d]", "[e]", "[f]")
```

L'utilizzatore deve qui modificare le informazioni riportate in base alla stazione di qualità dell'aria presa in considerazione.

Ground weather stations (gws) Si definiscono qui le stazioni meteorologiche al suolo che verranno utilizzate per effettuare la NM. Il programma consente di inserire più stazioni meteorologiche: la prima è quella di riferimento per la stazione di qualità dell'aria, le seguenti sono utilizzate per il calcolo del *temperature lapse-rate*. Si assegna qui:

- numero di stazioni meteorologiche al suolo (modificabile dall'utilizzatore):

```
gws_n <- 3
```

- matrice dei metadata (NON modificabile dall'utilizzatore):

```
gws_info <- matrix(1:(gws_n*3), nrow=gws_n, ncol=3)
```

- per ciascuna stazione meteorologica l'utilizzatore deve inserire una linea di codice in cui si specifica: a) ID stazione meteorologica, b) sito sperimentale, c) quota. Queste linee di codice costituiscono le righe della matrice dei metadata, pertanto l'indice di riga # deve essere aggiornato.

```
gws_info[,] <- c("[ID]", "[Sito]", [Quota])
```

- intestazione delle colonne per le variabili meteorologiche (al suolo e *temperature lapse-rate*) che comparirà nei file di output alla *random forest* (Sez. 3.4):

```
new_header <- c("WGaaa", "WRaaa", ..., "Gab", "Gac", "mtpG100")
```

Nel comando sopra riportato, ad esempio, si riportano i nomi delle variabili meteorologiche di interesse (WG: intensità del vento, WR: direzione del vento,

etc.) e un acronimo per identificare la stazione di appartenenza (*aaa*, *ab*, *ac*). Ciò può risultare utile qualora si utilizzino informazioni provenienti da diversi siti di misura o nel caso del *temperature lapse-rate* per identificarne la provenienza (*Gab*, *mtpG100*).

Le istruzioni che seguono, invece, NON devono essere modificate dall'utilizzatore, in quanto definiscono la struttura del file di input ed in particolare contengono l'ordine in cui si presentano le varie grandezze meteorologiche.

- numero di grandezze meteorologiche:

```
meteo_n <- 8
```

- matrice dei metadati:

```
meteo_info <- matrix(1:(meteo_n*5),nrow=meteo_n,ncol=5)
```

- grandezze meteorologiche elencate coerentemente alla loro posizione nei file di input (*WG*, *WR*, *LT*, *LF*, *LD*, *GS*, *SD*, *N*). Gli ultimi due campi sono la risoluzione temporale e il numero di righe nell'intestazione, così come figurano nei file forniti dall'Ufficio Idrografico della Provincia Autonoma di Bolzano. In questa versione del programma non sono utilizzati in quanto i file di input sono già predisposti nel formato richiesto, dopo un pre-processing. Istruzione:
- ```
meteo_info[#,] <- c("[ID]", "[Nome esteso]", "[Unità]", [...], [...])
```

**MTP-5HE temperature profiler (*mtp5*)** La procedura prevede di poter utilizzare i profili di temperatura misurati dal profilatore termico presso l'aeroporto di Bolzano per quantificare il *temperature lapse-rate* nei primi 100 m e 500 m di atmosfera al di sopra del suolo. A tal fine l'utilizzatore deve specificare ("*yes*" or "*no*") se desidera utilizzare questa informazione nell'esecuzione della NM:

```
temperature.profiles <- "no"
```

Le seguenti istruzioni, invece, NON devono essere modificate dall'utilizzatore, in quanto assegnano:

- la stringa di caratteri che identifica il file di input:

```
mtp5_ds <- 'mtp5'
```

- il numero di livelli verticali dei profili di temperatura:

```
mtp5_n <- 21
```

**Data set for random forest (*rf*): training** Le seguenti istruzioni sono necessarie per definire l'estensione temporale del dataset utilizzato dalla NM per far crescere la RF:

- elenco degli anni di dati da utilizzare per il training fornito dall'utilizzatore. L'istruzione riportata prevede di effettuare il training della RF utilizzando gli anni dal 2013 al 2017:

```
rf_training_years <- c('2013','2014','2015','2016','2017')
```

- numero di anni contenuti in `rf_training_years`. NON modificabile dall'utilizzatore:

```
rf_training_nyears <- length(rf_training_years)
```

- estensione temporale del dataset, deve essere modificato dall'utilizzatore, coerentemente con quanto definito in `rf_training_years`. Queste date permettono di calcolare l'estensione temporale del dataset:

```
rf_training_dttm0 <- "2013-01-01 00:00:00"
```

```
rf_training_dttm1 <- "2018-01-01 00:00:00"
```

**Data set for random forest (rf): normalization** Le seguenti istruzioni sono necessarie per definire l'estensione temporale del dataset su cui effettuare la NM:

- elenco degli anni di dati da normalizzare, fornito dall'utilizzatore. L'istruzione riportata prevede di effettuare il calcolo per il solo anno 2018:

```
rf_normalization_years <- c('2018')
```

- numero di anni contenuti in `rf_normalization_years`. NON modificabile dall'utilizzatore:

```
rf_normalization_nyears <- length(rf_normalization_years)
```

- estensione temporale del dataset, deve essere modificato dall'utilizzatore, coerentemente con quanto definito in `rf_normalization_years`. Queste date permettono di calcolare la scansione temporale del dataset:

```
rf_normalization_dttm0 <- "2018-01-01 00:00:00"
```

```
rf_normalization_dttm1 <- "2019-01-01 00:00:00"
```

**Data set for random forest (rf): meteorological quantities** Le istruzioni fornite in questa sezione del file sono necessarie per la costruzione dei due dataset (training e normalization). Ovviamente i due dataset differiscono solo per l'estensione temporale, mentre la struttura dei dati deve essere la medesima. Il dataset, oltre alla serie storica di concentrazioni di inquinanti, conterrà anche i dati meteorologici misurati dalla stazione di riferimento (`gws`) ed i *temperature lapse-rate*, calcolati dalle stazioni al suolo e/o dai profili di temperatura misurati dal profilatore (`mtp`).

- Numero di stazioni meteorologiche al suolo da utilizzare nel dataset per la RF (modificabile dall'utilizzatore):

```
rf_gws_n <- 3
```

- Matrice dei metadati, NON modificabile dall'utilizzatore:

```
rf_gws <- matrix(1:(rf_gws_n*5),nrow=rf_gws_n,ncol=10)
```

- Le seguenti istruzioni devono essere fornite dall'utilizzatore e compilano le righe della matrice dei metadati, quindi il numero di righe deve coincidere con il numero di stazioni al suolo utilizzate. La prima istruzione è relativa alla stazione meteorologica di riferimento, mentre le righe successive alle stazioni meteorologiche utilizzate per il calcolo del *temperature lapse-rate*. Di conseguenza il numero di riga # deve essere aggiornato opportunamente. L'istruzione deve contenere: a) ID della stazione, b) numero delle grandezze meteorologiche da considerare, c) ID delle grandezze meteorologiche. Quindi i comandi:

```
rf_gws[1,] <- c("AAAAAAA",8,"WG","WR","LT","LF","LD","GS","SD","N")
```

```
rf_gws[2,] <- c("BBBBBBB",1,"LT","-","-","-","-","-","-","-")
```

informano che il dataset conterrà le 8 grandezze meteorologiche misurate dalla stazione con ID AAAAAAA e la sola serie "LT" misurata dalla stazione con ID BBBBBBB.

- L'utilizzatore deve specificare ("yes" or "no") se desidera calcolare il *temperature lapse-rate* a partire dai dati delle stazioni al suolo. In caso affermativo, la procedura, in fase di costruzione del dataset, sostituirà la colonna delle temperature misurate dalla stazione al suolo sul versante con il corrispondente valore di  $\Gamma$ :

```
rf_lr_gws <- "yes"
```

- L'utilizzatore deve specificare ("yes" or "no") se desidera calcolare il *temperature lapse-rate* a partire dai profili di temperatura misurati dal profilatore termico (mtp). In caso affermativo, il dataset conterrà una colonna aggiuntiva contenente il valore calcolato di  $\Gamma$ :

```
rf_lr_mtp <- "yes"
```

**Random forest (rf): parameters** L'utilizzatore deve specificare i parametri con cui realizzare la RF, ossia:

- il numero di alberi

```
rf_trees <- 2000
```



- il numero di campionamenti  
`rf_samples <- 3000`

**Make the output directory or check if it exists** Le istruzioni che compaiono nel file di setting (e non riportate nel seguito) creano le cartelle contenenti l'output della NM, specificando l'estensione dei periodi di training e del periodo utilizzato per la normalizzazione. Questi comandi non devono essere modificati dall'utilizzatore.

## 3.3 Codice principale

Si procede ora ad illustrare la struttura del file `N02Normalization_24h_XX1.r` che implementa la NM.

### 3.3.1 File: `N02Normalization_24h_XX1.r`

**1. Load packages and libraries** Sono elencate le librerie di R necessarie alla RF e alla NM.

**2. Random forest functions** Si richiamano le porzioni di codice, contenute nella cartella `scripts`, che effettuano operazioni specifiche durante l'esecuzione della procedura (si veda Sez. 3.3.2).

**Identify the current directory** Percorso in cui si trova la cartella di lavoro.

**3. Import settings listed in the configuration file** Carica il file `settings_XX1.r`. L'utilizzatore deve aggiornare il nome del file.

**4. Manage the metadata** La procedura modifica e ristruttura i metadata (`md`) forniti nel file di settings per la stazione di qualità dell'aria (`aqsm`), per le stazioni meteorologiche al suolo (`gwsmd`) e per le grandezze meteorologiche (`meteom`).

**5. Random forest: make the data sets for training and normalization** La procedura crea i set di dati necessari per il training della RF e su cui effettuare la NM. Queste azioni sono condotte richiamando i file:

- `make_rf_dataset_training.r`
- `make_rf_dataset_normalization.r`

contenuti nella cartella `scripts`. I dataset creati sono salvati nella corrispondente cartella di output come files:

- `XX1_av1d_training.dat`
- `XX1_av1d_normalization.dat`

**6. Random forest: training of the random forest** La procedura elabora il dataset per il training della RF e salva nella corrispondente cartella di output il file: `XX1_av1d_training_prepared.dat`.

Il file, oltre alle grandezze meteorologiche fornite, contiene anche altre variabili che sono utilizzate dal modello per individuare i trend. In particolare, il codice calcola:

- la data unix (secondi dal 01-01-1970) rappresentativa del trend di lungo periodo;
- il giorno giuliano (numero del giorno da inizio anno);
- il giorno della settimana (numero del giorno della settimana): lunedì =1, martedì =2, ..., domenica=7;
- l'ora del giorno.

In questo studio l'ora del giorno non è rilevante in quanto si è fatto uso di dati medi giornalieri.

Sulla base delle informazioni disponibili fornite, si procede a far crescere il modello di RM, che viene allocato in:

`rf_trained_model`. Le azioni sopra descritte sono effettuate mediante le funzioni:

- `rmw_prepare_data`
- `rmw_train_model`

contenute nel pacchetto `rmweather`, a cui si rimanda per ulteriori dettagli, e personalizzate per gli scopi dello studio.

**7. Random forest: normalization** La procedura elabora il dataset per la NM e salva nella corrispondente cartella di output il file:

`XX1_av1d_normalization_prepared.dat`.

Analogamente al punto precedente, il codice elabora i dati delle grandezze meteorologiche fornite e calcola:

- la data unix (secondi dal 01-01-1970) rappresentativa del trend di lungo periodo;
- il giorno giuliano (numero del giorno da inizio anno);

- il giorno della settimana (numero del giorno della settimana): lunedì =1, martedì =2, ..., domenica=7;
- l'ora del giorno.

In questo studio l'ora del giorno non è rilevante in quanto si è fatto uso di dati medi giornalieri. Quindi il codice effettua la NM per il corrispondente dataset sulla base del modello di RF realizzato al passo precedente mediante le funzioni:

- `rmw_prepare_data`
- `rmw_do_normalization`

contenute nel pacchetto `rmweather`, a cui si rimanda per ulteriori dettagli, e personalizzate per gli scopi dello studio. In particolare, la NM è eseguita escludendo il trend di lungo periodo, *i.e.* la data unix, così come suggerito da Grange et al. (2018); Grange & Carslaw (2019). In questo studio la NM è quindi effettuata sulle seguenti variabili:

- `"day_julian", "weekday"`
- `"WG", "WR", "LT", "LF", "LD", "GS", "SD", "N"`
- `"Gamma1", "Gamma2"` sono i valori di *temperature lapse-rate* calcolati utilizzando le stazioni al suolo e/o i dati del profilatore di temperatura.

La serie storica normalizzata è infine salvata nella corrispondente cartella di output nel file:

`XX1_normalized_av1d.dat`

**8. Analisi degli output** Le righe di codice che concludono il file consentono di valutare la performance della RF per la NM. In particolare, il codice valuta i seguenti parametri:

- *variable importances*: valuta e classifica il peso delle variabili esplicative utilizzate nel descrivere la variabilità delle concentrazioni di inquinante;
- *partial dependences*: il codice calcola la dipendenza delle concentrazioni di inquinante in funzione delle variabilità di ciascuna variabile esplicativa, considerata singolarmente e indipendentemente dalle altre.

**NOTA:** Qualora l'intervallo temporale utilizzato per la “crescita” della *random forest* coincida con il periodo utilizzato per la normalizzazione meteorologica, nel file di settings i vettori:

`rf_training_years`

`rf_normalization_years`

contengono gli stessi elementi. In questo caso, la procedura è implementata come segue:

1. si crea il data set per la sola normalizzazione, in quanto coincidente con quello di training, attraverso il comando:  
`make_rf_dataset_normalization.r`
2. la *random forest* è fatta crescere utilizzando l'80% delle osservazioni per il training ed il restante dataset per la validazione.

In questo modo si evita che la *random forest* sia validata con dati utilizzati per calibrarla.

### 3.3.2 Cartella: scripts

Scripts per la rielaborazione dei dati di input al fine di creare un dataset da fornire in ingresso alla procedura:

- `make_rf_dataset_training.r`
- `make_rf_dataset_normalization.r`

Script per creare il data set da utilizzare per costruire la RF e su cui eseguire la NM:

- `rmw_prepare_data.r`

Script per costruire la RM:

- `rmw_train_model.r`

Scripts per eseguire la NM:

- `rmw_do_normalization.r`

Scripts per valutare l'esito della NM in termini di overfitting, stima dell'importanza e delle dipendenze parziali e produrre i grafici:

- `rmweather_plotting_functions.r`
- `rmw_find_breakpoints.r`

- `rmw_model_statistics.r`
- `rmw_partial_dependencies.r`
- `rmw_predict.r`
- `rmw_predict_the_test_set.r`

Scripts a supporto della procedura:

- `rmweather_helpers.r`

## 3.4 Files di output

Come anticipato, l'esito di ciascuna NM è salvato all'interno della cartella avente come percorso:

`./NO2Normalization/aqsXX1/`

e nome contenente i periodi utilizzati per il training della RF e su cui è condotta la NM. Durante il computo, la procedura implementata salva in questa cartella:

- i data sets rielaborati a partire dai dati forniti in input nella cartella `db`:  
`XX1_av1d_training.dat`  
`XX1_av1d_normalization.dat`  
 Questi files hanno una linea di intestazione e sono strutturati come descritto in Sez. 3.2.1.
- i data sets predisposti per la crescita della RF e per la normalizzazione:  
`XX1_av1d_prepared_training.dat`  
`XX1_av1d_prepared_normalization.dat`  
 La procedura rimuove dai data sets precedentemente creati i periodi in cui non sono disponibili i dati di concentrazione dell'inquinante ( $\text{NO}_2$ ), aggiunge il termine di trend di lungo periodo (data unix), il giorno giuliano, il giorno della settimana e l'ora del giorno. Inoltre, aggiunge una nuova colonna (prima della data) contenente una stringa di caratteri `testing` o `training` che identifica se i dati contenuti in quella riga sono stati utilizzati per effettuare il training della RF o per stimare la serie storica normalizzata. Per come è strutturata la procedura implementata, il codice utilizza il 90% dei dati in `XX1_av1d_training.dat` per far crescere la RM, mentre utilizza il 90% dei dati contenuti in `XX1_av1d_normalization.dat` per la NM.
- le serie storiche di inquinanti normalizzate con i rispettivi valori di deviazione standard, output del processo di NM sono invece salvate in:

**XX1\_normalized\_av1d.dat**

il file contiene nella prima colonna l'informazione temporale, nella seconda colonna la concentrazione normalizzata e nella terza la deviazione standard. Nel caso di studio, il file riporta per ciascun giorno investigato la concentrazione media giornaliera normalizzata di  $\text{NO}_2$ . Si noti che la deviazione non va intesa come l'errore associato alla stima della concentrazione normalizzata. Infatti, la concentrazione normalizzata è la media (d'insieme) delle concentrazioni predette da ciascun albero della *random forest* quando istruito con le variabili esplicative associate al ciascun istante temporale. La deviazione standard è quindi da intendersi come deviazione standard d'insieme ed indica quanto la *random forest* creata e le variabili esplicative fornite sono in grado di restituire previsioni tra loro coerenti. Ad esempio, una deviazione standard nulla sarebbe indice del fatto che tutti gli alberi della foresta, indipendentemente dalle variabili esplicative fornite come input, sono in grado di prevedere lo stesso valore di concentrazione.

- il peso delle variabili esplicative selezionate (*variable importances*) per descrivere la variabilità delle concentrazioni di inquinante è salvata in:

**XX1\_importance\_av1d.dat**

nel file le variabili esplicative sono ordinate in ordine decrescente di importanza (**rank**). Il valore numerico riportato in terza colonna è il valore dell'importanza valutato dalla RM. La distribuzione ottenuta permette quindi di identificare quali sono le variabili selezionate che sono maggiormente responsabili o che possono essere utilizzate per valutare e modellare la dipendenza delle concentrazioni osservate dalle condizioni meteorologiche.

- la relazione tra la singola variabile esplicativa e le concentrazioni di inquinante è invece descritta in termini di *partial dependences*, salvate nel file:

**XX1\_partialdependences\_av1d.dat**

Le dipendenze parziali sono ottenute valutando il valore delle concentrazioni medie di inquinante in funzione della variabile esplicativa. Il file di output contiene quindi: la variabile esplicativa considerata (prima colonna), i valori della variabile esplicativa (seconda colonna) e i corrispondenti valori di concentrazione di inquinante (terza colonna). Nelle distribuzioni ottenute, tanto più è elevato il peso della variabile esplicativa, definito dalla *variable importances*, tanto è ampio l'intervallo di concentrazioni di inquinante esplorato. Ciò implica che, un valore nullo o costante di concentrazione a fronte di una variazione significativa della variabile esplicativa è indice di una debole

o nulla dipendenza della variabilità dell'inquinante dalla variabile esplicativa considerata.

- la capacità predittiva della RF è valutata su di un data set ricampionato di osservazioni ed in particolare sulle osservazioni etichettate con `testing` nel file `XX1_partialdependences_av1d.dat`. Questa stima è effettuata valutando i coefficienti di correlazione lineare tra la concentrazione media misurata e la concentrazione media giornaliera stimata per il medesimo istante temporale a partire dalle variabili esplicative disponibili. Le serie storiche di concentrazioni (osservate e stimate) sono riportate nel file:

`XX1_predictions_av1d.dat`

che contiene il dataset presente in `XX1_partialdependences_av1d.dat` per i soli istanti temporali etichettati come `testing`, mentre nell'ultima colonna sono salvate le concentrazioni stimate.

- il file `XX1_summary_YYYY.pdf` o `XX1_summary_YYYY_YYYY.pdf` contenente i grafici relativi a *variable importances*, concentrazioni osservate *vs* concentrazioni stimate, serie normalizzata e *partial dependences*.

## CAPITOLO 4

---

### Analisi delle concentrazioni normalizzate di NO<sub>2</sub>

---

La normalizzazione delle serie di NO<sub>2</sub>, misurate dalle stazioni di qualità dell'aria elencate in Tab. 2.1, è stata condotta in questo studio generando una RM di 500 alberi (`rf_trees=500`), mentre i valori normalizzati sono stati stimati mediante 3000 ricampionamenti (`rf_samples=3000`). Per ciascuna stazione di qualità dell'aria sono state eseguite tre diverse NM con:

1. training dal 2012-01-01 al 2017-12-31 e periodo normalizzato dal 2018-01-01 al 2018-12-31;
2. training coincidente con il periodo da normalizzare, dal 2013-01-01 al 2018-12-31 (come previsto dalla prima convenzione).
3. training coincidente con il periodo da normalizzare, dal 2012-01-01 al 2018-12-31 (sulla base degli approfondimenti richiesti).

La prima normalizzazione, coerentemente con quanto richiesto nello studio tecnico, permette di valutare la capacità predittiva di una RM calibrata su di un periodo temporale molto esteso e poi valutata su di un periodo diverso. La seconda e la terza, invece, hanno lo scopo di valutare l'evoluzione temporale delle concentrazioni di NO<sub>2</sub> sul periodo 2012–2018 e 2013–2018 al netto della forzante meteorologica. Il confronto si è reso necessario in quanto, come osservato dagli sviluppatori del pacchetto **rmweather** (Grange, 2018b), la RM può restituire dei valori anomali agli estremi del periodo investigato. Per questo motivo, l'estensione del periodo 2013–2018 al 2012–2018, permette di valutare se fluttuazioni nelle serie normalizzate per l'anno 2013 sono effettivamente legate a variazioni del regime delle emissioni o se, invece, dipendono da un non corretto funzionamento dell'algoritmo di normalizzazione. Gli sviluppatori del pacchetto **rmweather** rendono nota questa possibile anomalia dell'algoritmo ma non riportano le cause che la determinano.

In Tabella 4.1 si riassumono, per ciascuna stazione di qualità dell'aria, le grandezze meteorologiche utilizzate per effettuare la normalizzazione meteorologica,



Tabella 4.1: Utilizzo del dataset meteorologico per la normalizzazione delle serie di NO<sub>2</sub>.

| Stazione                               | ID      | WG | WR | LT | LF | LD | GS | N | F |
|----------------------------------------|---------|----|----|----|----|----|----|---|---|
| <b>BZ4</b>                             |         |    |    |    |    |    |    |   |   |
| 1. Bolzano Via Amba Alagi              | BZ6APPA | X  | X  |    |    |    | X  | X |   |
| 2. Bolzano Ospedale                    | 83200MS | X  | X  | X  | X  | X  |    |   |   |
| 3. Bronzolo                            | 85700MS | X  | X  |    |    |    |    |   |   |
| Profilatore ( $\Gamma_{100\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| Profilatore ( $\Gamma_{500\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| <b>BZ5</b>                             |         |    |    |    |    |    |    |   |   |
| 1. Bolzano Via Amba Alagi              | BZ6APPA | X  | X  |    |    |    | X  | X |   |
| 2. Bolzano Ospedale                    | 83200MS |    |    | X  | X  | X  |    |   |   |
| Profilatore ( $\Gamma_{100\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| Profilatore ( $\Gamma_{500\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| <b>LS1</b>                             |         |    |    |    |    |    |    |   |   |
| 1. Bronzolo                            | 85700MS | X  | X  | X  | X  | X  | X  | X |   |
| Profilatore ( $\Gamma_{100\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| Profilatore ( $\Gamma_{500\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| <b>ME1</b>                             |         |    |    |    |    |    |    |   |   |
| 1. Merano                              | 23200MS | X  | X  | X  | X  | X  | X  | X |   |
| 2. Lauregno                            | 90000SF |    |    |    |    |    |    |   | X |
| Profilatore ( $\Gamma_{100\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| Profilatore ( $\Gamma_{500\text{m}}$ ) | MTP     |    |    |    |    |    |    |   | X |
| <b>BX1</b>                             |         |    |    |    |    |    |    |   |   |
| 1. Bressanone                          | 39100MS | X  | X  | X  | X  | X  | X  | X |   |
| 2. Fiè allo Sciliar                    | 75600MS |    |    |    |    |    |    |   | X |
| 3. Plose                               | 69900MS |    |    |    |    |    |    |   | X |

mentre in Fig. 4.1 si confrontano le serie normalizzate di NO<sub>2</sub> ottenute sull'intero periodo 2012–2018. Infine, nelle Figure 4.2, 4.3, 4.4, 4.5 e 4.6 si rappresentano le serie storiche osservate e normalizzate. In particolare:

- Fig. 4.2a–4.6a riportano le serie misurate di NO<sub>2</sub> e rappresentate in termini di medie giornaliere e di medie annue.
- Fig. 4.2b–4.6b riportano la serie normalizzata di NO<sub>2</sub> per l'anno 2018 e ottenute utilizzando il periodo 2012–2017 per il training della *random forest*. Oltre alla serie delle concentrazioni medie giornaliere normalizzate (punti blu) si riportano le deviazioni standard associate alla media d'insieme (banda azzurra) e, a titolo di confronto, le concentrazioni misurate medie annue.
- analogamente al caso precedente, nelle Fig. 4.2c–4.6c e Fig. 4.2d–4.6d si riportano, invece, le serie normalizzate di NO<sub>2</sub> calcolate rispettivamente sui periodi 2013–2018 e 2012–2018.

Le concentrazioni medie giornaliere, osservate da tutte le stazioni di qualità dell'aria (Fig. 4.2a–4.6a), mostrano un ben marcato ciclo annuale, con un massimo

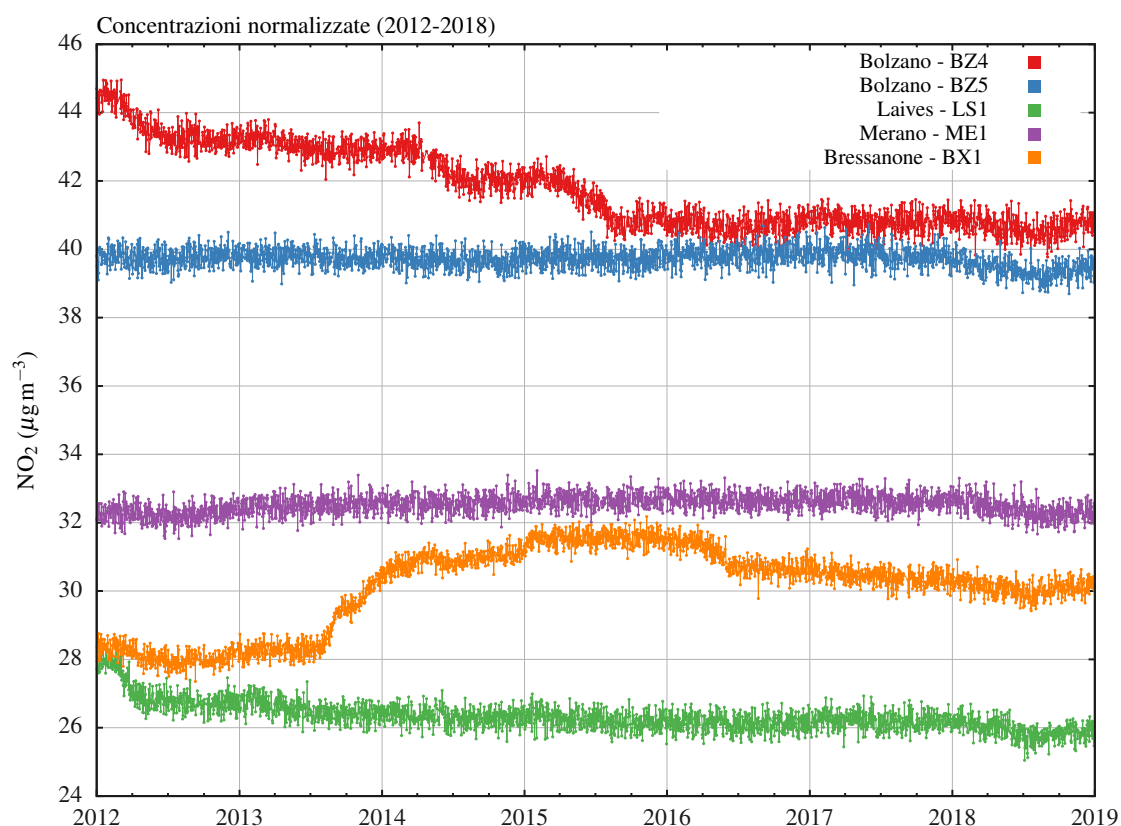


Figura 4.1: Concentrazioni  $\text{NO}_2$  normalizzate sul periodo 2012–2018.

in inverno ed un minimo in estate. Le serie normalizzate, invece, proprio perché indipendenti dalla variabilità meteorologica, non mostrano fluttuazioni legate alla ciclicità annuale. Scostamenti tra la media annua misurata e la serie normalizzata possono essere interpretati come segue: concentrazioni normalizzate superiori alla media annua misurata sono indice di anni meteorologicamente favorevoli in termini di dispersione degli inquinanti, mentre, viceversa, si osservano concentrazioni normalizzate inferiori alla media annua misurata in presenza di anni in cui la forzante meteorologica ha inibito la diluizione degli inquinanti. Infatti, si osserva che l'estate del 2014 è stata particolarmente perturbata. Ciò ha favorito la dispersione di  $\text{NO}_2$  come confermato dalle concentrazioni misurate, che per questo anno appaiono inferiori alla serie normalizzata. Viceversa, l'estate molto calda del 2015 ha giocato un ruolo sfavorevole in termini di concentrazioni di  $\text{NO}_2$ , la cui media annua osservata è superiore al trend normalizzato. Questo effetto è ben visibile in tutte le stazioni di qualità dell'aria investigate, ad eccezione di Bressanone, che invece mostra un comportamento anomalo. Infine, come introdotto, trend e fluttuazioni delle concentrazioni normalizzate possono essere ricondotte a:

1. variazioni dei regimi emissivi associati all'attuazione di politiche per il conte-

nimento delle sostanze inquinanti, *e.g.* zone a traffico limitato, modifiche del traffico;

2. variazioni dei regimi emissivi locali associati ad interventi urbanistici in prossimità del punto di misura, *e.g.* realizzazione di piste ciclabili, apertura di cantieri, demolizione/costruzione di edifici in grado di alterare localmente la ventilazione dell'area;
3. rinnovo del parco veicolare ed in particolare il rinnovo del parco mezzi per il trasporto pubblico;
4. interventi sui sensori utilizzati per il monitoraggio (*e.g.* calibrazioni).

L'analisi delle serie normalizzate permette di osservare che:

1. In tutte le stazioni si intravede una debole flessione delle concentrazioni normalizzate per l'anno 2018. Ciò sembrerebbe indicare che durante questo anno vi sia stata un'effettiva riduzione nel regime delle emissioni. Tuttavia, questa flessione è veramente contenuta e, visti i possibili malfunzionamenti dell'algoritmo di normalizzazione in prossimità dei bordi del periodo investigato, è necessario attendere ulteriori dati del 2019 per poter confermare questo comportamento.
2. Le serie normalizzate delle stazioni di Bolzano Piazza Adriano (BZ5) e Laives (LS1) non mostrano trend significativi, ma le concentrazioni normalizzate assumono valori pressoché costanti e pari a  $40 \mu\text{g m}^{-3}$  per BZ5 e  $26 \mu\text{g m}^{-3}$  per LS1. La significativa differenza dei valori medi tra le due stazioni è attribuibile alla diversa collocazione dei punti di misura nell'area di studio.
3. La brusca riduzione degli  $\text{NO}_2$  normalizzati all'inizio della serie storica di LS1 non è da ricercare in variazioni del regime emissivo, bensì è dovuto al problema sopra indicato della RF nel restituire valori attendibili ai bordi del periodo investigato. Analogamente, anche la brusca riduzione degli  $\text{NO}_2$  osservata presso Bolzano Via Claudia Augusta (BZ4) ad inizio 2012 va attribuita alle medesime cause.
4. La serie di Merano (ME1), invece, mostra un debole incremento delle concentrazioni normalizzate tra il 2014 e il 2015, seguito da una diminuzione nel biennio 2017–2018.
5. La serie di Bolzano Via Claudia Augusta (BZ4) mostra una chiara e progressiva diminuzione delle concentrazioni normalizzate che da metà 2012 prosegue

sino a metà 2015. Da questo istante in poi gli  $\text{NO}_2$  normalizzati restano pressoché costanti e tendono ad allinearsi con i valori normalizzati osservati a BZ5 (Fig. 4.1 serie rossa e blu). In questo caso, la causa di questa riduzione può essere ricercata in interventi urbanistici attorno al punto di misura.

6. Come anticipato, la serie di  $\text{NO}_2$  (osservata e normalizzata) presso Bressanone (BX1) mostra un comportamento “anomalo” rispetto alle serie investigate negli altri siti. Infatti, già nelle serie misurate, si osserva che le concentrazioni medie annue di  $\text{NO}_2$  aumentano nel 2014, in controtendenza con quanto osservato per le altre serie, mentre riprendono a diminuire nel 2016. La serie normalizzata (Fig. 4.1 serie arancio) permette di apprezzare ancor meglio questo comportamento. Dal 2012 a metà 2013, le concentrazioni normalizzate assumono valori costanti attorno a  $28 \mu\text{g m}^{-3}$ , per poi crescere fino a metà 2015, raggiungendo valori prossimi a  $32 \mu\text{g m}^{-3}$ . Le concentrazioni normalizzate tornano a diminuire ad inizio 2016, raggiungendo i  $30 \mu\text{g m}^{-3}$  nel 2018. Anche in questo caso, come per BZ4, le cause che possono aver prodotto questo comportamento vanno ricercate in interventi effettuati nelle immediate vicinanze del punto di misura o valutando eventuali modifiche del parco veicolare circolante.

Come discusso nelle Sezioni 1.2 e 3.4, oltre alle serie normalizzate, l'algoritmo *random forest* restituisce in output informazioni relative al peso delle singole variabili esplicative nel caratterizzare la variabilità delle concentrazioni osservate. In Figura 4.7 si rappresentano i pesi che le singole variabili esplicative hanno per ciascuna serie di concentrazioni disponibile (file di output `XX1_importance_av1d.dat`) considerando il periodo 2012–2018, in quanto più esteso temporalmente. Si osserva così che per tutte le stazioni di qualità dell'aria la temperatura dell'aria è la variabile esplicativa che maggiormente spiega la variabilità delle concentrazioni di  $\text{NO}_2$ . Ciò è sensato in quanto questa variabile contiene anche i termini di trend annuale e stagionale, similmente agli  $\text{NO}_2$ . Viceversa, le variabili esplicative che hanno meno importanza sono pressione atmosferica, precipitazione cumulata giornaliera e direzione del vento. In questo caso, la direzione del vento non gioca un ruolo chiave in quanto l'analisi è stata effettuata su scala giornaliera. Infatti, in terreno complesso, il regime dei venti è molto variabile durante la giornata e quindi il dato di direzione media giornaliera non è sufficientemente rappresentativo. Potrebbe invece diventare una variabile importante qualora la normalizzazione meteorologica venisse applicata a scala oraria.

Un altro output della *random forest* è costituito dalle dipendenze parziali (files `XX1_partialdependences_av1d.dat`), che mostrano il legame tra le concentrazioni osservate di  $\text{NO}_2$  e le singole variabili esplicative. I grafici riportati nelle Figure 4.8–

4.12 identificano le distribuzioni associate alle prime 6 variabili esplicative che, per ciascuna stazione di qualità dell'aria (Fig. 4.7), contribuiscono significativamente nello "spiegare" la variabilità degli  $\text{NO}_2$ . In questi grafici, l'ampiezza dell'intervallo di concentrazioni esplorate (asse delle ordinate) è una misura di quanto la variabile esplicativa influenzi il dato di concentrazione. In altre parole, tanto più è ampio l'intervallo di concentrazioni esplorate, tanto più la variabile esplicativa è in grado di apprezzare la variabilità delle concentrazioni. Tutte le distribuzioni associate ad una grandezza meteorologica, inoltre, mostrano un pattern in cui è presente un intervallo nel quale il valore della concentrazione è indipendente dai valori assunti dalla variabile esplicativa. Ciò implica che le variabili esplicative, prese singolarmente in quell'intervallo, non sono in grado di spiegare la variabilità delle concentrazioni osservate.

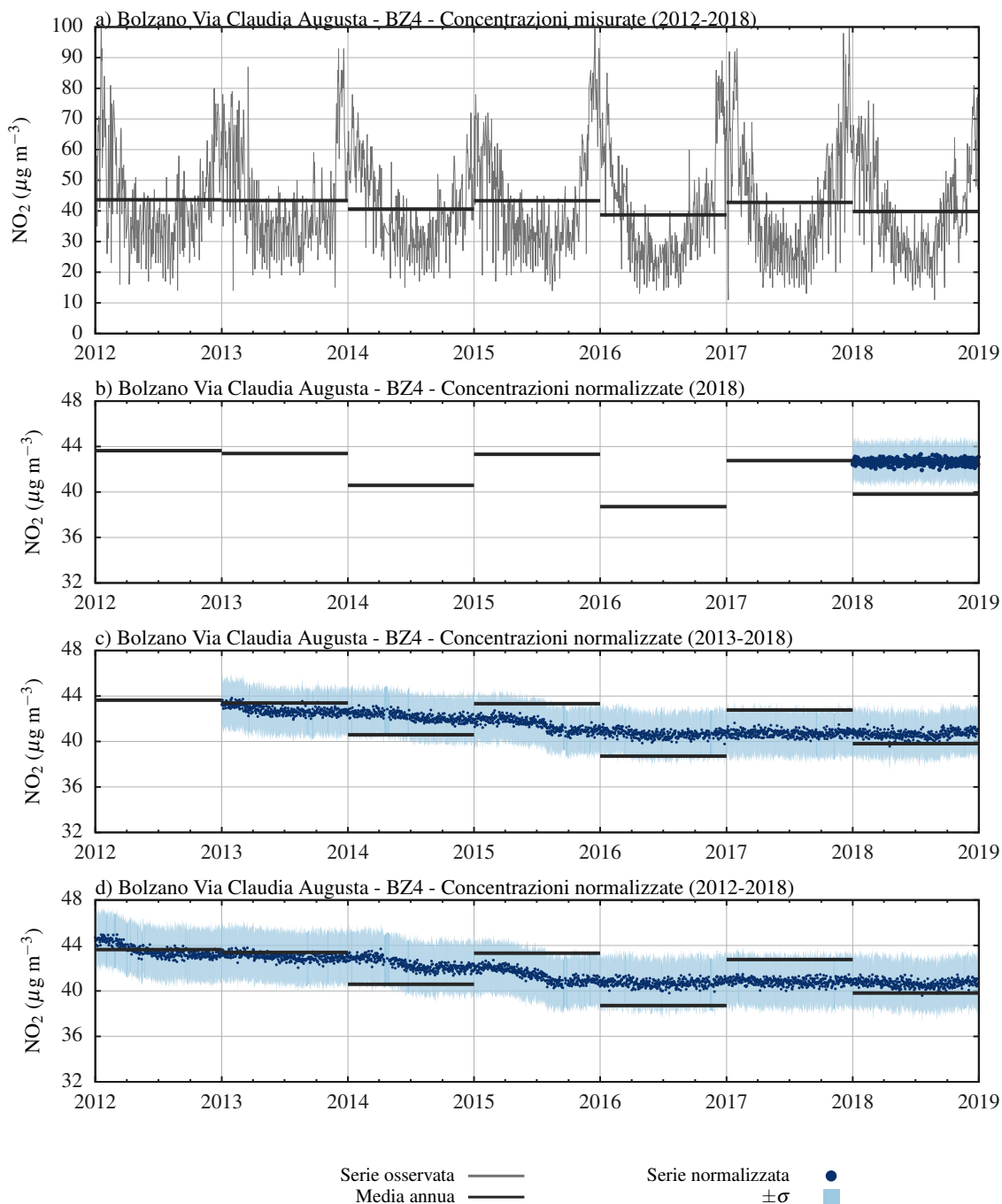


Figura 4.2: Concentrazioni presso la stazione di Bolzano Via Claudia Augusta (BZ4). Concentrazioni di NO<sub>2</sub> misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO<sub>2</sub> normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della *random forest* e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.

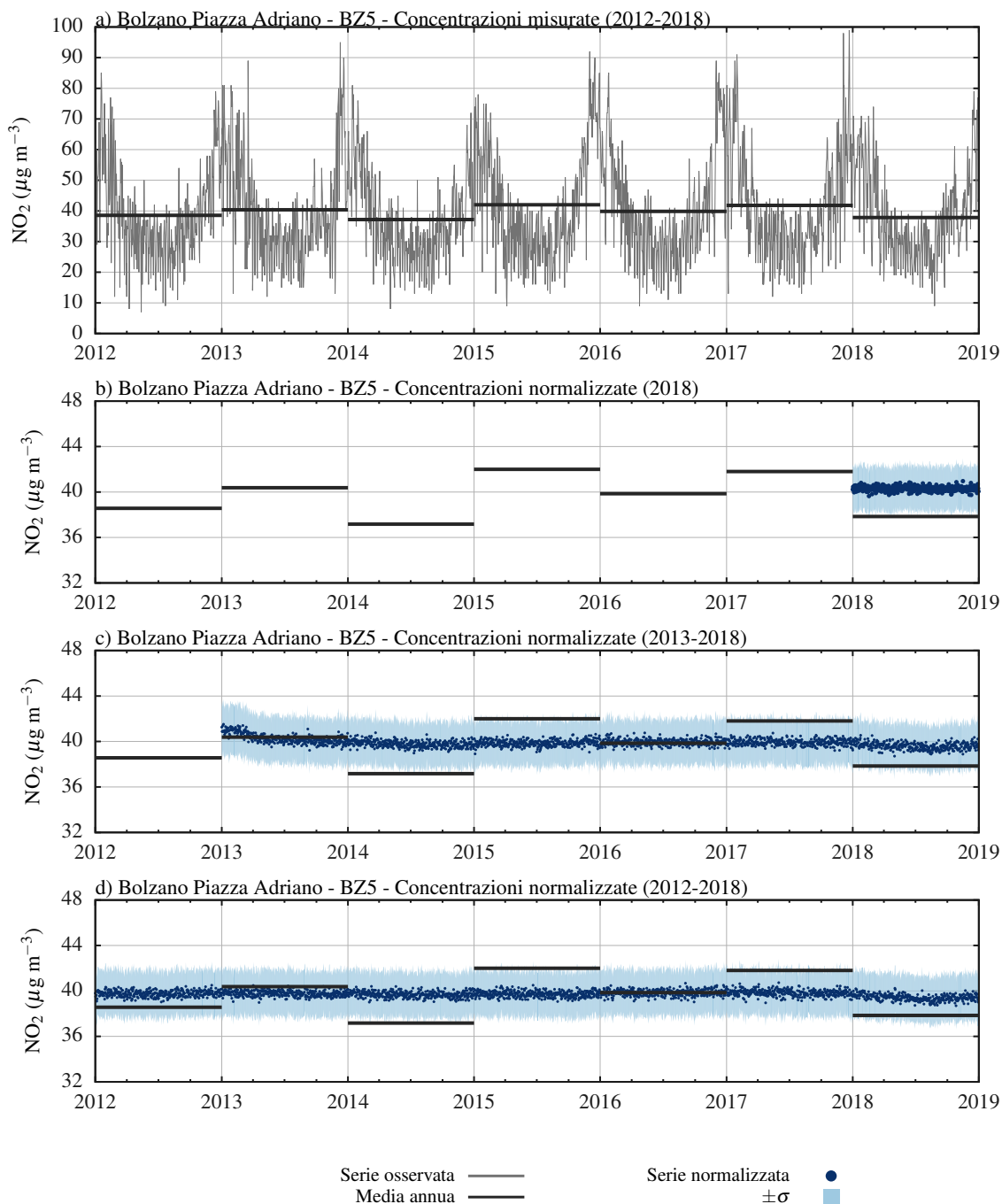


Figura 4.3: Concentrazioni presso la stazione di Bolzano Piazza Adriano (BZ5). Concentrazioni di NO<sub>2</sub> misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO<sub>2</sub> normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della *random forest* e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.

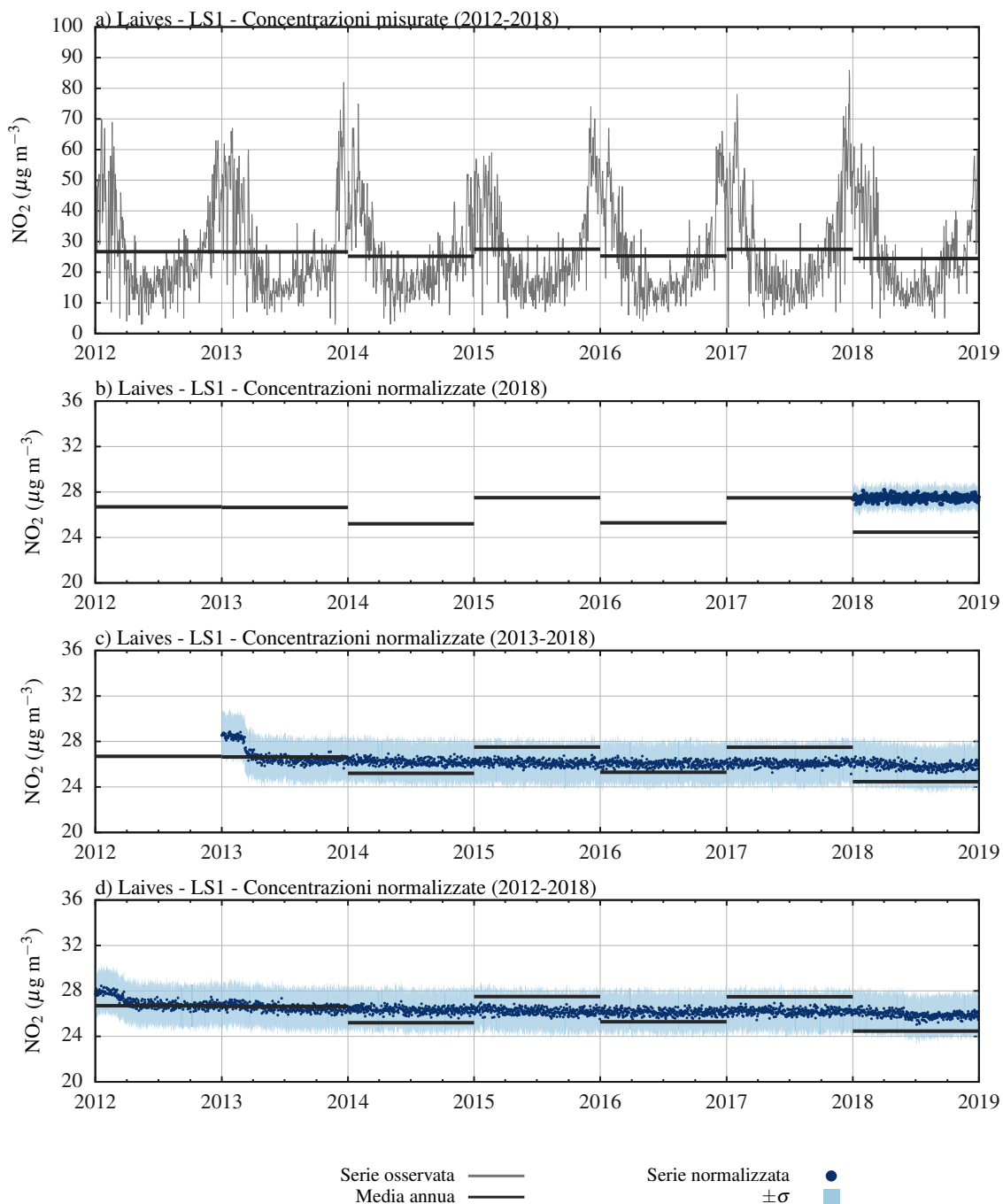


Figura 4.4: Concentrazioni presso la stazione di Laives (LS1). Concentrazioni di NO<sub>2</sub> misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO<sub>2</sub> normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della *random forest* e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.



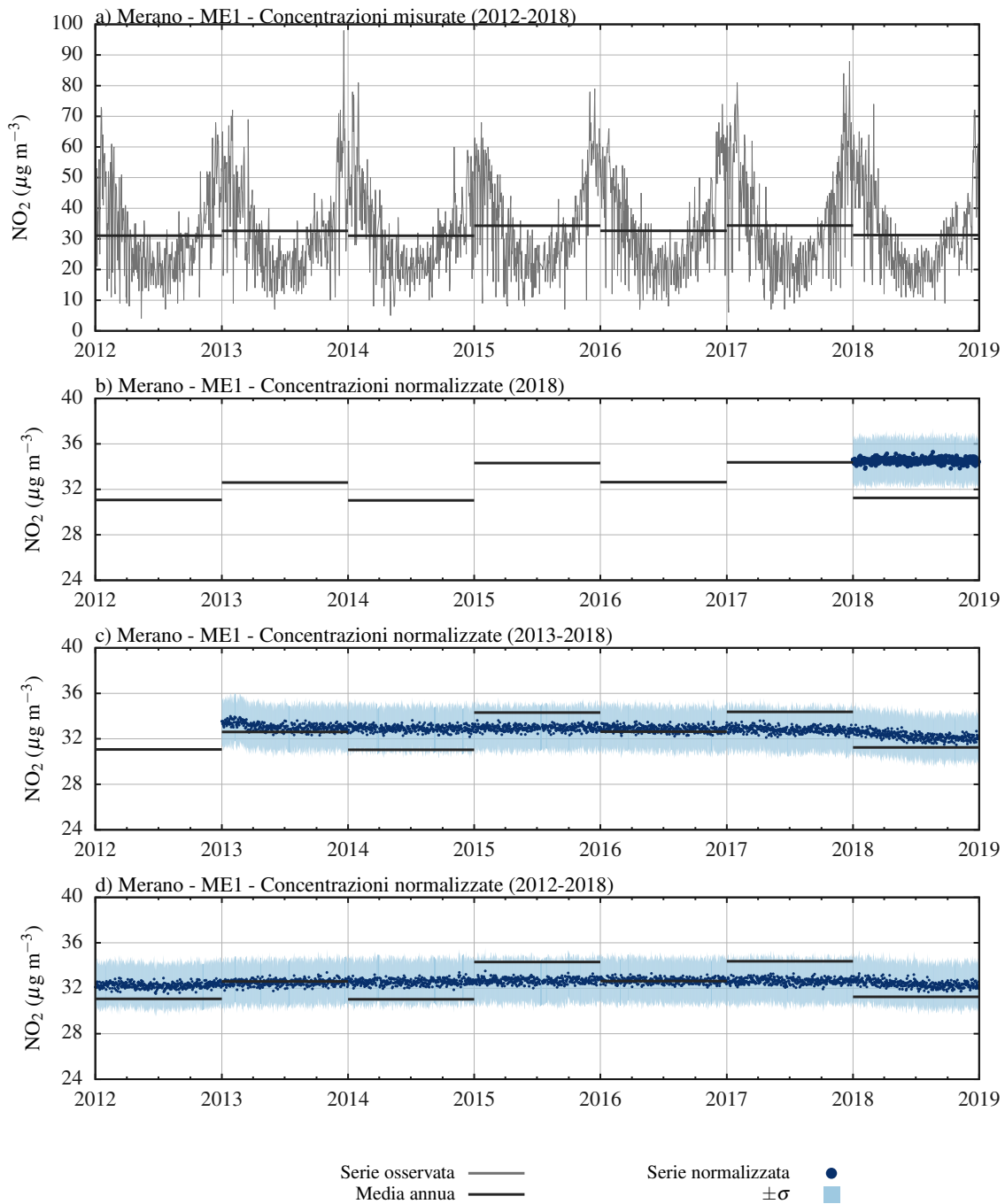


Figura 4.5: Concentrazioni presso la stazione di Merano (ME1). Concentrazioni di NO<sub>2</sub> misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO<sub>2</sub> normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della *random forest* e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.

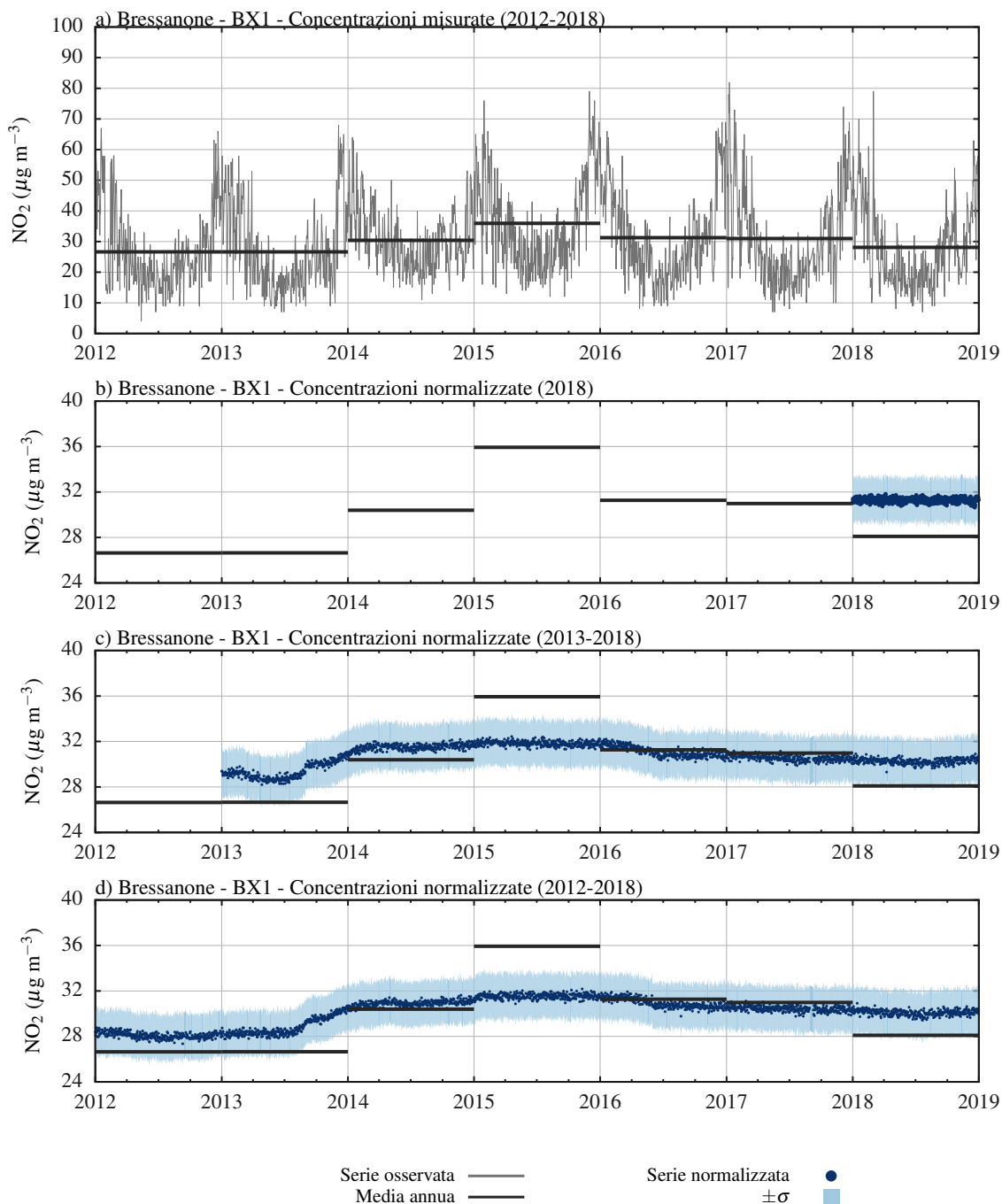


Figura 4.6: Concentrazioni presso la stazione di Bressanone (BX1). Concentrazioni di NO<sub>2</sub> misurate e rappresentate come medie giornaliere e medie annue (a). Concentrazioni NO<sub>2</sub> normalizzate, utilizzando: (a) il periodo 2012–2017 per la calibrazione della *random forest* e il 2018 per la validazione; (b) il periodo 2013–2018 e (c) il periodo 2012–2018. In (a), (b), (c) i tratti orizzontali rappresentano le medie annue osservate, le distribuzioni in blu le concentrazioni normalizzate medie giornaliere, le bande in azzurro le deviazioni standard associate alla media d'insieme.

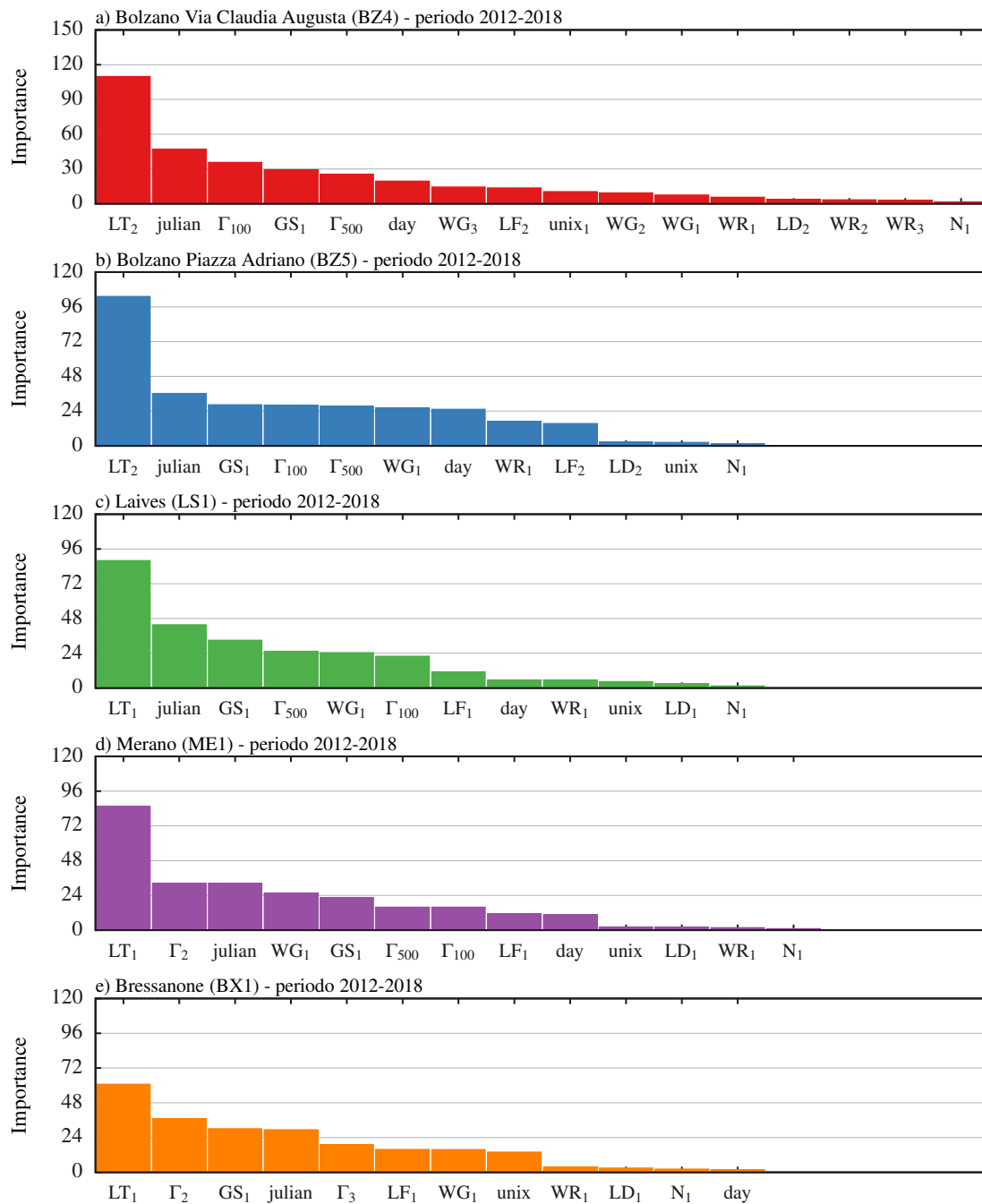


Figura 4.7: Peso delle variabili esplicative nella *random forest* realizzata per analizzare le concentrazioni di NO<sub>2</sub> sul periodo 2012–2018. Gli acronimi delle grandezze esplicative sono: WG intensità del vento, WR: direzione del vento, LT: temperatura dell’aria, LF umidità relativa, LD pressione atmosferica, GS radiazione globale, N precipitazione,  $\Gamma$  *temperature lapse rate*, “unix” data unix, “day” giorno della settimana, “julian” giorno giuliano. I numeri al pedice corrispondono alla numerazione delle stazioni al suolo elencate in Tab. 4.1 e utilizzate per ciascuna serie di NO<sub>2</sub>, mentre si indicano con  $\Gamma_{100}$  e  $\Gamma_{500}$  i *temperature lapse rate* ricavati dal profilatore termico per uno strato di atmosfera con spessore rispettivamente di 100 e 500 m.

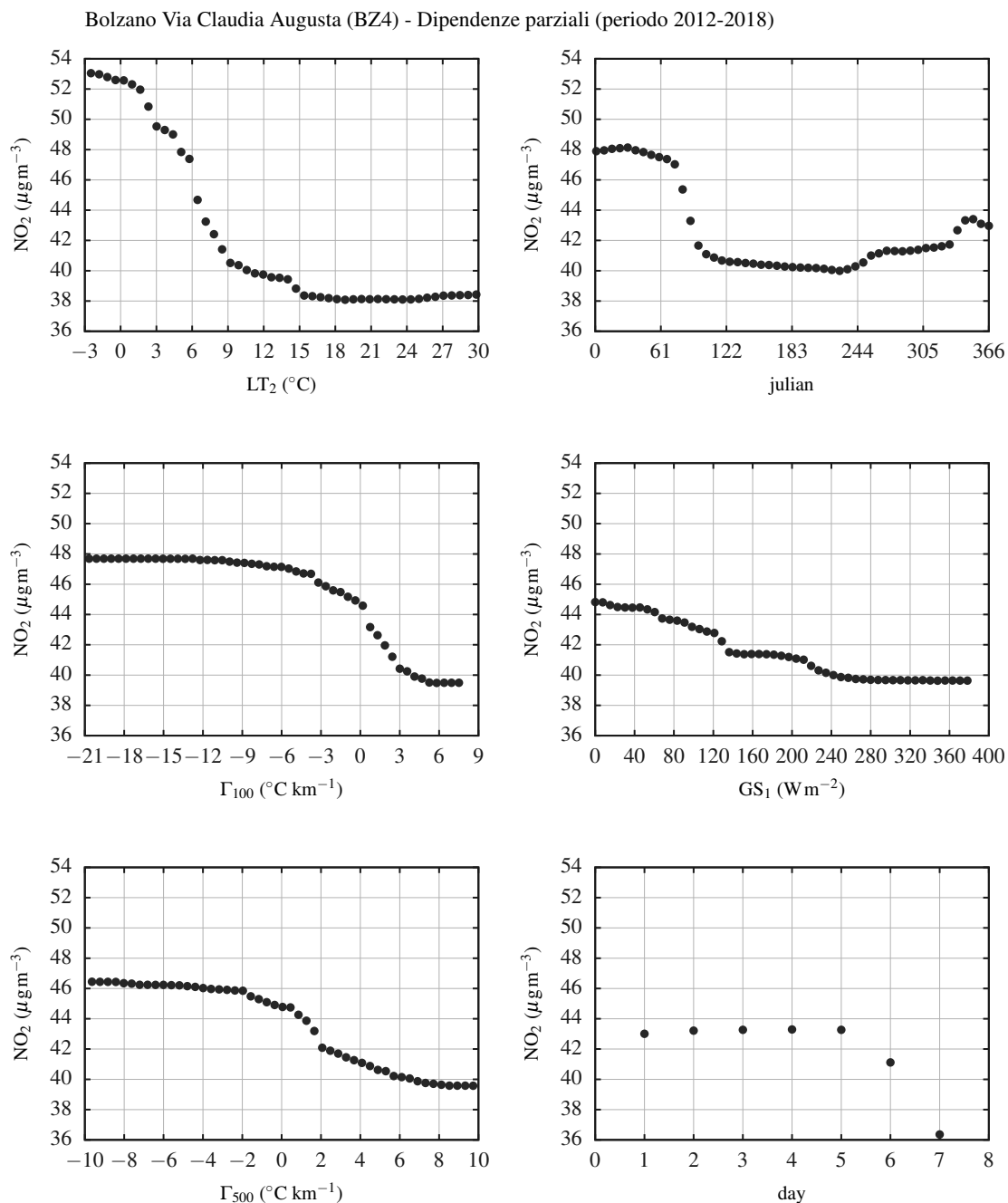


Figura 4.8: Dipendenza delle concentrazioni osservate di  $\text{NO}_2$  da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bolzano Via Claudia Augusta (BZ4). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.

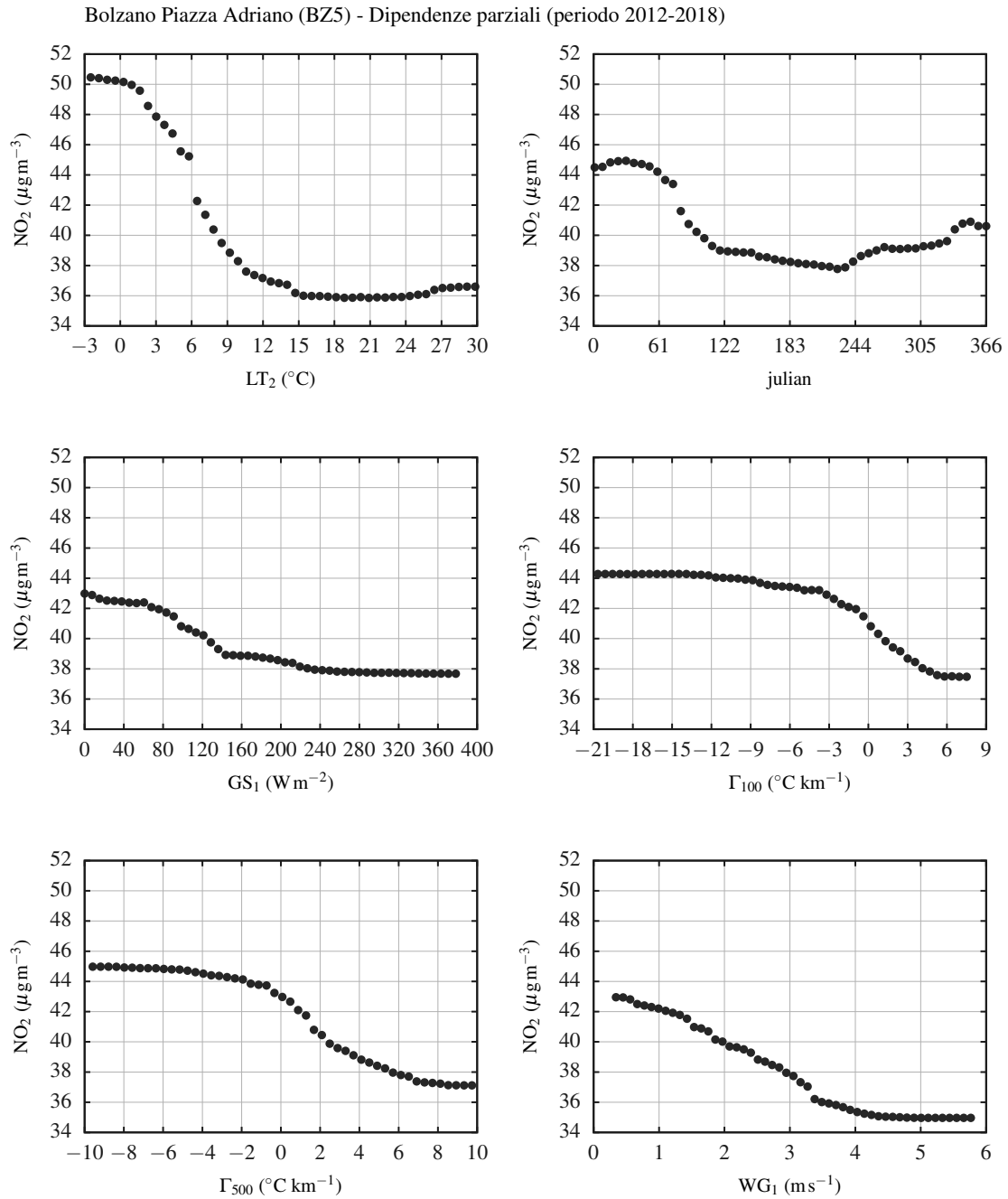


Figura 4.9: Dipendenza delle concentrazioni osservate di  $\text{NO}_2$  da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bolzano Piazza Adriano (BZ5). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.

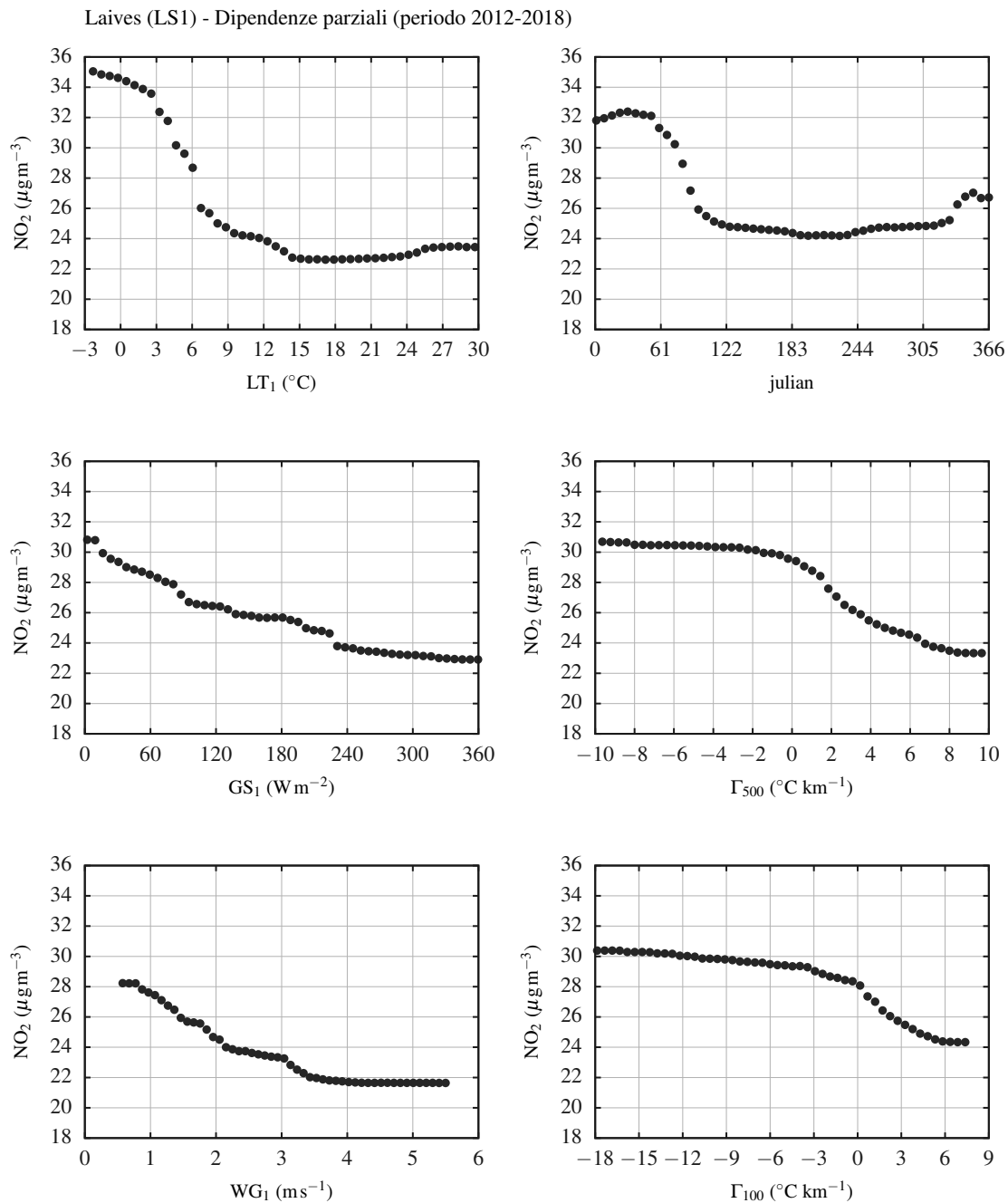


Figura 4.10: Dipendenza delle concentrazioni osservate di  $\text{NO}_2$  da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Laives (LS1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.

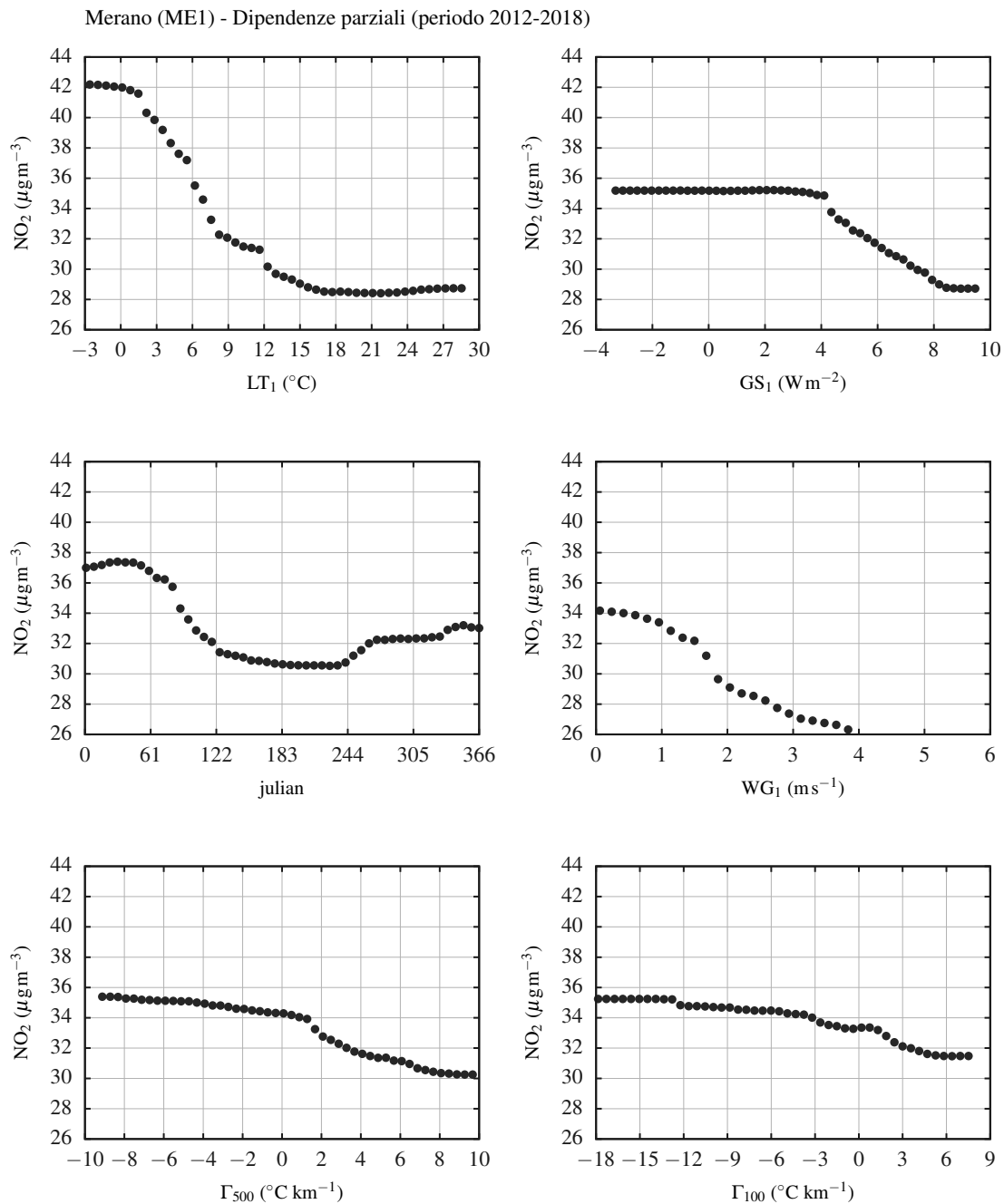


Figura 4.11: Dipendenza delle concentrazioni osservate di  $\text{NO}_2$  da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Merano (ME1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.

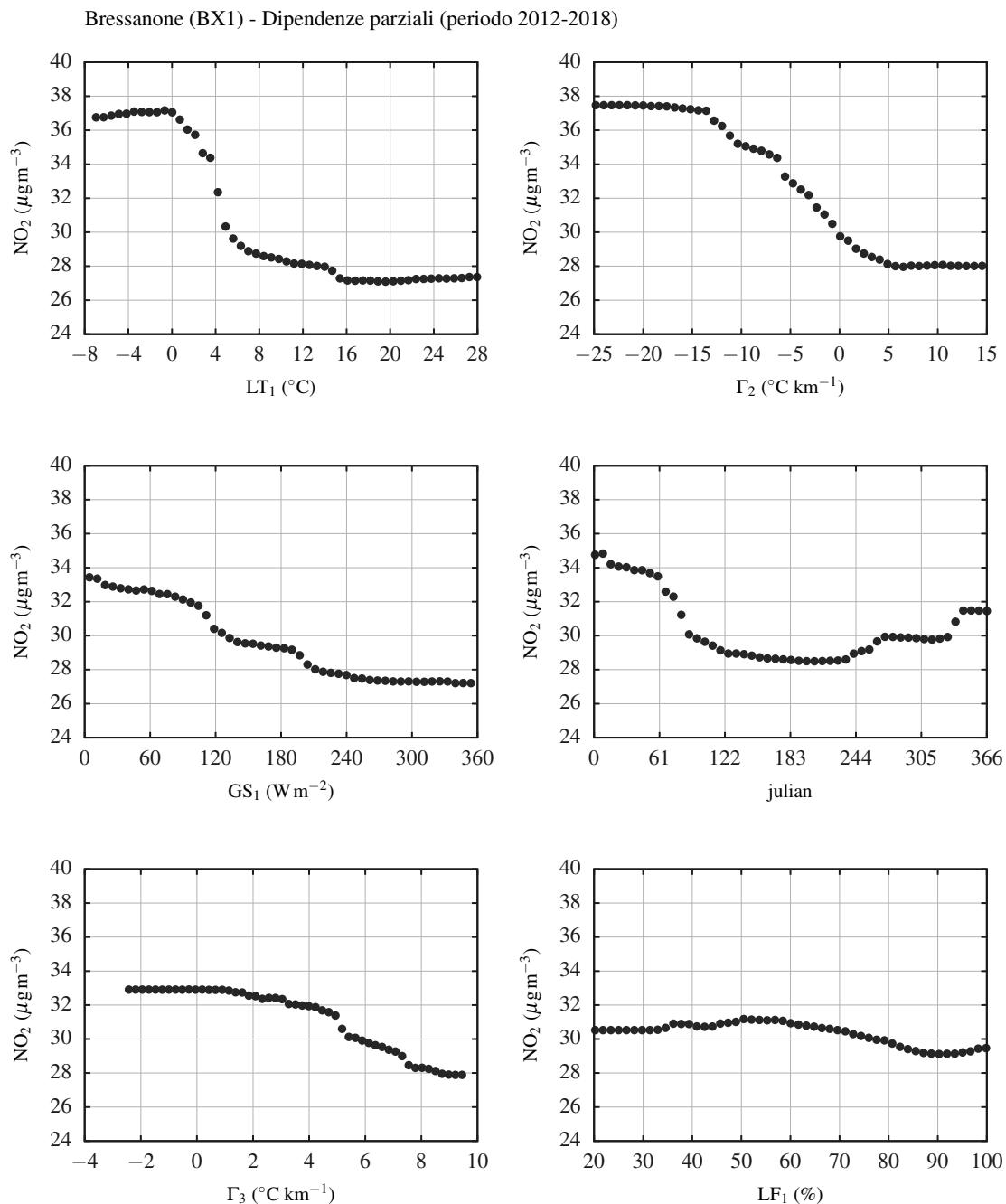


Figura 4.12: Dipendenza delle concentrazioni osservate di  $\text{NO}_2$  da ciascuna delle variabili esplicative utilizzate per normalizzare i dati disponibili presso la stazione di qualità dell'aria di Bressanone (BX1). Si riportano i legami per le prime sei variabili esplicative che maggiormente descrivono la variabilità del dataset.



---

## Conclusioni

---

Sulla base dei risultati conseguiti con le azioni descritte nei Capitoli precedenti si possono trarre alcune considerazioni in merito ai quesiti formulati nell'incarico conferito dall'Agenzia provinciale per l'ambiente e la tutela del clima della Provincia Autonoma di Bolzano circa lo sviluppo e la verifica di una metodologia per la normalizzazione delle concentrazioni medie annue di biossido di azoto ( $\text{NO}_2$ ) rispetto alle condizioni meteorologiche per la Provincia di Bolzano.

1. L'utilizzo di un algoritmo per la normalizzazione meteorologica basato sulla tecnica di *machine learning* nota come *random forest* si è dimostrato efficace per la valutazione dell'effetto delle condizioni meteorologiche sulle serie di biossido di azoto anche in un ambiente a terreno complesso come la Provincia Autonoma di Bolzano.
2. La normalizzazione meteorologica è stata effettuata implementando ed adattando ai diversi casi studio il pacchetto **rmweather** (Grange, 2018b), sviluppato in ambiente R. Questo pacchetto è già stato recentemente applicato con successo per valutare trend di  $\text{PM}_{10}$  in Svizzera (Grange et al., 2018) e per valutare l'effetto di politiche atte al controllo di  $\text{SO}_2$ ,  $\text{NO}_x$  ed  $\text{NO}_2$  a Londra (Grange & Carslaw, 2019). In vista di un utilizzo routinario della normalizzazione meteorologica, durante l'implementazione della metodologia si è prestata particolare attenzione alla complessità di utilizzo dell'algoritmo. In particolare, i parametri su cui un eventuale utilizzatore potrà poi intervenire sono stati ridotti al minimo, automatizzando invece, laddove possibile, il processo (Cap. 3).
3. La normalizzazione meteorologica è stata condotta per le stazioni di qualità dell'aria di Bolzano Via Claudia Augusta (BZ4), Bolzano Piazza Adriano (BZ5), Laives (LS1), Merano (ME1) e Bressanone (BX1), utilizzando due diverse strategie: (1) periodo 2012–2017 per la calibrazione della *random forest* e stima delle concentrazioni normalizzate riferite al solo 2018; (2) normalizzazione meteorologica dell'intero periodo 2012–2018 per la determinazione di trend e fluttuazioni nel regime emissivo.

4. Per ciascuna stazione di qualità dell'aria è stata condotta un'analisi meteo-climatologica e delle concentrazioni osservate di biossido di azoto finalizzata alla contestualizzazione delle diverse aree studio (Cap. 2). In particolare, l'analisi è stata effettuata utilizzando il periodo 2012–2017 come periodo di riferimento da confrontare con l'anno 2018. Da questa analisi è emerso come il 2018 sia stato caratterizzato da condizioni meteorologiche confrontabili a quelle osservate nel periodo 2012–2018, mentre si osservano concentrazioni di  $\text{NO}_2$  inferiori.
5. L'applicazione della metodologia finalizzata alla stima delle concentrazioni normalizzate per il solo anno 2018 (Figure 4.2b–4.6b) mostra che per tutte le stazioni investigate, le concentrazioni medie annue sono inferiori alle concentrazioni normalizzate. Ciò può essere legato principalmente a due effetti, ossia: (1) la presenza di condizioni meteorologiche favorevoli alla dispersione degli inquinanti e (2) la diminuzione delle emissioni di  $\text{NO}_2$ . Al fine quindi di poter valutare la presenza di trend nel regime delle emissioni, sono state stimate le serie normalizzate di  $\text{NO}_2$  per l'intero periodo 2012–2018.
6. Tutte le serie normalizzate (Fig. 4.1) mostrano una debole diminuzione delle concentrazioni di  $\text{NO}_2$  soprattutto nella prima metà del 2018. Sulla base del dataset osservativo, ad ora disponibile, non è ancora possibile definire se il segnale osservato sia effettivamente legato ad una variazione del regime delle emissioni di  $\text{NO}_2$  in tutte le stazioni investigate. Solo estendendo l'analisi all'anno 2019 sarà possibile confermare o smentire questa flessione nelle serie normalizzate.
7. Dal confronto delle serie normalizzate (Fig. 4.1), risulta invece evidente come le concentrazioni di  $\text{NO}_2$  presso la stazione BZ4 siano progressivamente diminuite dal 2012 al 2016. Al contrario, dal 2016 al 2018 le concentrazioni normalizzate di BZ4 sono pressoché costanti nel tempo ed assumono valori prossimi a quelli calcolati per BZ5 (situata in centro alla città di Bolzano). La serie normalizzata per la stazione di Bressanone (BX1), invece, mostra un significativo aumento delle concentrazioni a partire dalla seconda metà del 2013 fino ad inizio 2016. Dopodiché le concentrazioni normalizzate tendono a diminuire, anche se debolmente.

Queste fluttuazioni, in quanto indipendenti dalla forzante meteorologica, vanno ricercate in interventi che localmente possono aver alterato il regime emissivo.

8. L'algoritmo implementato, oltre a restituire le serie normalizzate, permette anche di valutare quali sono le variabili che maggiormente influenzano le concentrazioni e che quindi meglio descrivono la variabilità dell' $\text{NO}_2$ . Come mostrato in Fig. 4.7, la temperatura dell'aria è la variabile che maggiormente descrive il comportamento delle concentrazioni di  $\text{NO}_2$ , in quanto contiene informazioni relative al ciclo stagionale e annuale. Tuttavia, questa variabile non è sufficiente per spiegare l'intera variabilità dei biossidi di azoto, soprattutto quando la temperatura è superiore a circa  $18^\circ\text{C}$ . Infatti, come mostrato nelle Figure 4.8–4.12, le concentrazioni di  $\text{NO}_2$  sono indipendenti dalla temperatura quanto essa è superiore a circa  $18^\circ\text{C}$ . Solo utilizzando altre variabili e grandezze meteorologiche è possibile “spiegare” l'evoluzione del biossido di azoto.

Al fine di poter valutare l'esito di politiche atte a contenere le emissioni di  $\text{NO}_2$  o valutare fluttuazioni del regime emissivo, si consiglia di effettuare la normalizzazione meteorologica a scala annuale. L'esecuzione della normalizzazione meteorologica a scala semestrale, invece, potrebbe risultare utile come metodo per un controllo intermedio.

---

## Bibliografia

---

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Ezau, I. N., Wolf, T., Miller, E. A., Repina, I. A., Troitskaya, Y. I., & Zilitinkevich, S. S. (2013). The analysis of results of remote sensing monitoring of the temperature profile in lower atmosphere in Bergen (Norway). *Russian Meteorology and Hydrology*, 38(10), 715–722.
- Falocchi, M., Giovannini, L., Tomasi, E., Antonacci, G., & Zardi, D. (2018). Climatological characteristics of the temperature field within the Bolzano basin (Italy). In *Primo Congresso Nazionale della Associazione Italiana Scienze dell’Atmosfera e Meteorologia (CN1-AISAM)*. Bologna, 10–13 settembre 2018.
- Gaikovich, K. P., Kadygrov, E. N., Kosov, A. S., & Troitskii, A. V. (1992). Thermal sounding of the atmospheric boundary layer at the center of the oxygen absorption line. *Radiofizika*, 35, 130—136.
- Giovannini, L., Antonacci, G., Zardi, D., Laiti, L., & Panziera, L. (2014b). Sensitivity of simulated wind speed to spatial resolution over complex terrain. *Energy Procedia*, 59, 323–329.
- Giovannini, L., Laiti, L., Serafin, S., & Zardi, D. (2017). The thermally driven diurnal wind system of the Adige Valley in the Italian Alps. *Quarterly Journal of the Royal Meteorological Society*.
- Giovannini, L., Zardi, D., de Franceschi, M., & Chen, F. (2014a). Numerical simulations of boundary-layer processes and urban-induced alterations in an Alpine valley. *International Journal of Climatology*, 34(4), 1111–1131.
- Grange, S. K. (2018b). **rmweather**: *Tools to conduct meteorological normalization on air quality data*.  
URL <https://github.com/skgrange/rmweather>
- Grange, S. K., & Carslaw, D. C. (2019). Using meteorological normalisation to detect interventions in air quality time series. *Science of The Total Environment*, 653, 578–588.  
URL <http://www.sciencedirect.com/science/article/pii/S004896971834244X>

- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for swiss pm<sub>10</sub> trend analysis. *Atmospheric Chemistry and Physics*, 18(9), 6223–6239.  
URL <https://www.atmos-chem-phys.net/18/6223/2018/>
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.  
URL <https://www.R-project.org/>
- Scanzani, F. (2010). Ground based passive microwave radiometry and temperature profiles. In A. Pena, & C. B. Hasager (Eds.) *Remote Sensing for Wind Energy*, (pp. 260–275). DTU Risoe Campus, Roskilde, Denmark: Technical Univ. of Denmark, DTU Wind Energy.
- Tomasi, E., Giovannini, L., Falocchi, M., Zardi, D., Antonacci, G., Ferrero, E., Bisignano, A., Alessandrini, S., & Mortarini, L. (2018). Dispersion Modeling Over Complex Terrain in the Bolzano Basin (IT): Preliminary Results from a WRF-CALPUFF Modeling System. In C. Mensink, & G. Kallos (Eds.) *Air Pollution Modeling and its Application XXV*, (pp. 157–161). Cham: Springer International Publishing.
- Tong, W., Hong, H., Fang, H., Xie, Q., & Perkins, R. (2003). Decision forest: combining the predictions of multiple independent decision tree models. *Journal of Chemical Information and Computer Sciences*, 43(2), 525–531.
- Troitskii, A. V., Gaikovich, K. P., Gromov, V. D., Kadygrov, E. N., & Kosov, A. S. (1993). Thermal sounding of the atmospheric boundary layer in the oxygen absorption band center at 60 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, 31, 116–120.
- Westwater, E. R., Han, Y., Irisov, V. G., Leuskiy, V., Kadygrov, E. N., & Viazankin, S. A. (1999). Remote Sensing of Boundary Layer Temperature Profiles by a Scanning 5-mm Microwave Radiometer and RASS: Comparison Experiments. *Journal of Atmospheric and Oceanic Technology*, 16(7), 805–818.
- Wolf, T., Esau, I., & Reuder, J. (2014). Analysis of the vertical temperature structure in the Bergen valley, Norway, and its connection to pollution episodes. *Journal of Geophysical Research: Atmospheres*, 119(18).
- Zeldin, M., & Meisel, W. (1978). Use of meteorological data in air quality trend analysis. *Technology Service Corp.*, (PB-283448).