



Esperienze di utilizzo di modelli machine learning per la mappatura della qualità dell'aria

Camillo Silibello, Nicola Pepe, Francesco Uboldi*

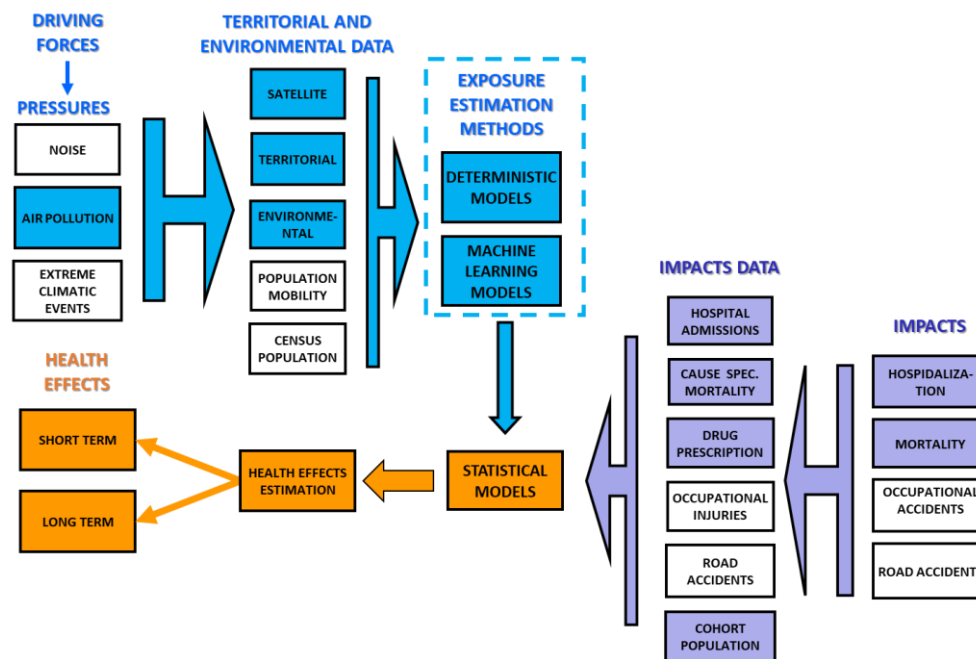
**attualmente Fondazione CIMA*



**X giornata sulla modellistica in aria(net)
Milano, 29 marzo 2023**

- ✓ Il Progetto BEEP (2017-2019)
 - Modelli ML: schema di applicazione
 - Valutazione della Qualità dell'aria sulla Regione Piemonte
 - Campi di concentrazione sulla Regione Lombardia (EPIMED)
 - EuroCC Italy: QualeAria Local
 - Valutazione della Qualità dell'aria sulla Regione Sardegna
 - Progetto «Air Break»

BEEP - Uso di **B**ig data in **E**pidemiologia ambi**E**ntale ed occu**P**azionale Finanziato da **Inail**, Piano di attività di ricerca (Bric ID 04) Periodo: 2017 - 2019



VII^a giornata sulla Modellistica in ARIA(NET)

Milano – 30 Gennaio 2020 – dalle 9:45 alle 17:00

Programma

09.45	Introduzione - Suite modellistiche 2019: novità principali	ARIANET
10.30	Rassegna delle attività svolte all'interno del progetto BEEP a scala nazionale, locale e urbana	ARIANET
10.45	Applicazione di modelli di <i>machine learning</i>	SIMULABIA

- Il Progetto BEEP (2017-2019)
- ✓ **Modelli ML: schema di applicazione**
- Valutazione della Qualità dell'aria sulla Regione Piemonte
- Campi di concentrazione sulla Regione Lombardia (EPIMED)
- EuroCC Italy: QualeAria Local
- Valutazione della Qualità dell'aria sulla Regione Sardegna
- Progetto «Air Break»

METODI

Applicazione dell'algoritmo ML **Random Forest** in due fasi:

- 1. addestramento**, ovvero alla sua capacità di riprodurre le concentrazioni osservate nei siti di monitoraggio (**fase di training**) sulla base di un insieme di predittori (**training data**: misure di qualità dell'aria e predittori)
- 2. generalizzazione**, è volta a stimare le concentrazioni nelle celle della griglia dove non sono disponibili osservazioni (10-fold Cross-Validation "by monitor")

PREDITTORI

• SPAZIALI (STATICI):

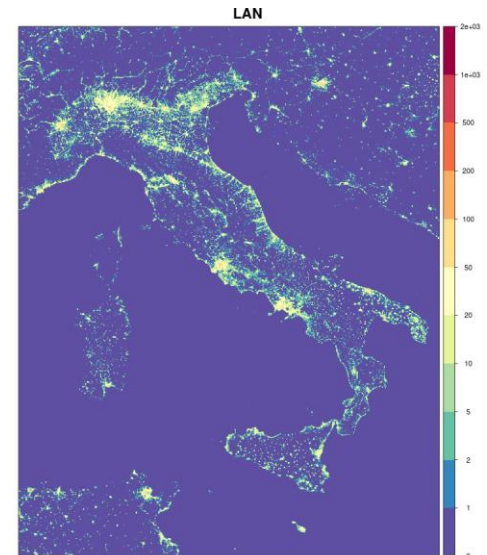
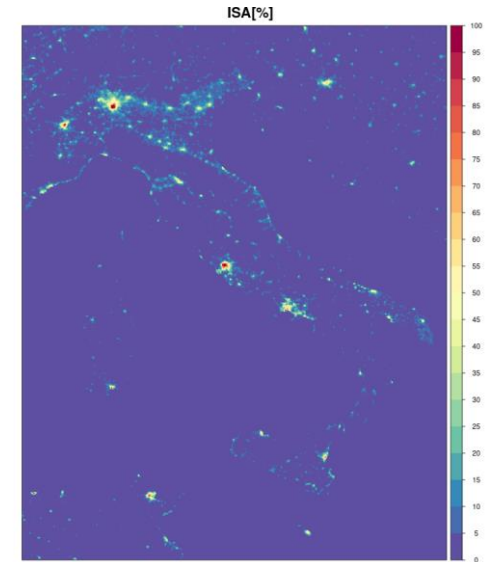
- ✓ Corine Land Cover
- ✓ superficie impermeabile costruita (Impervious Surface Area)
- ✓ immagini satellitari della luce notturna (Light At Night)
- ✓ statistiche di vario relative alle tipologie di strade (distanza, lunghezza, volume di traffico, ...)
- ✓ distanza dal mare;
- ✓ elevazione media;

• SPAZIO-TEMPORALI:

- ✓ Leaf Area Index (mensile)
- ✓ Campi prodotti da CTM FARM (orari e giornalieri – PM-)

• TEMPORALI:

- ✓ giorni Feriali/festivi, seno e il coseno del giorno giuliano (moltiplicato per $2\pi/365$, per tener conto della ciclicità)



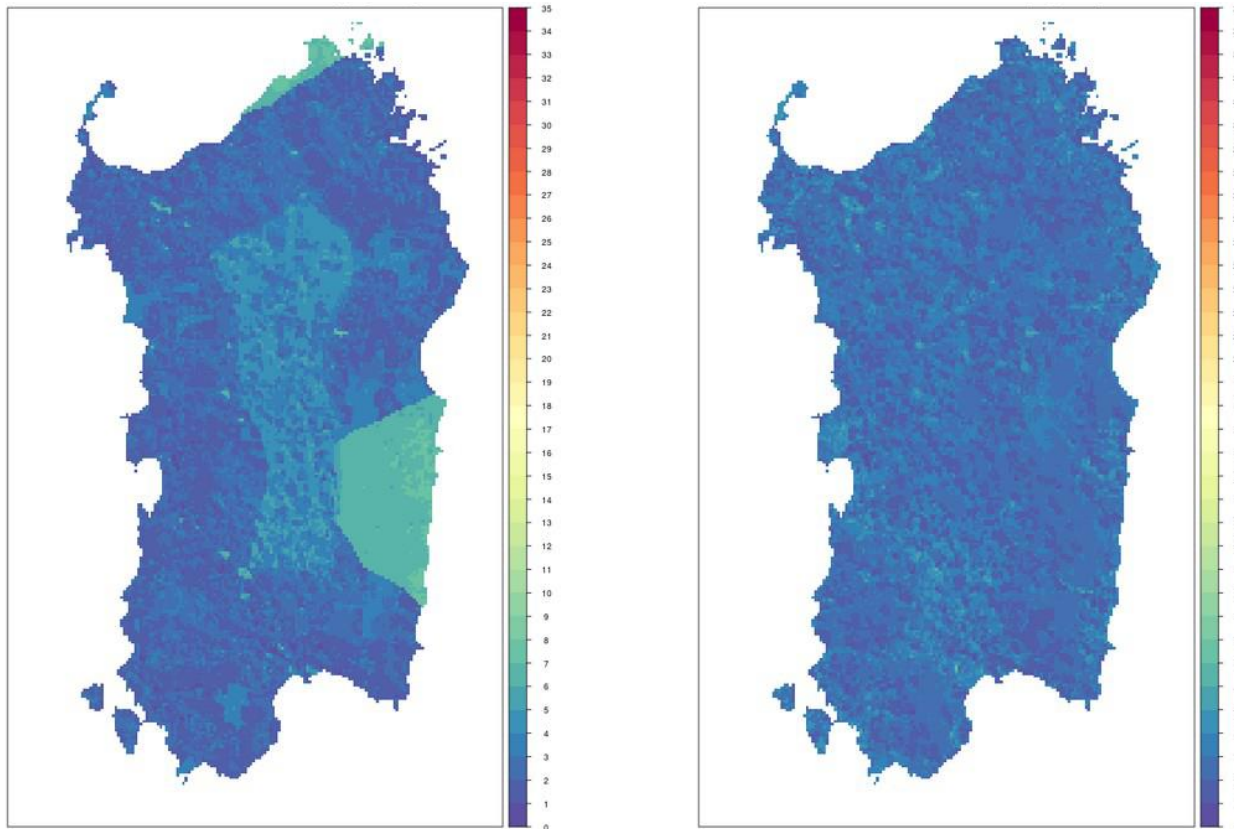
Osservazioni

- <https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm>
- serie temporali di altre stazioni usate per validazione (Piemonte, Lombardia) o per training (Sardegna)

Casi studio

- ❑ Regione Piemonte, Training: anno 2019
 - *risoluzione 1000 m, nx= 221, ny= 285*
 - Campi FARM a 4 km risoluzione
 - PM_{10} , $PM_{2.5}$ (medie giornaliere); NO_2 , O_3 (medie orarie)
 - **testato anche algoritmo XGBoost**
- ❑ Regione Lombardia (EPIMED), Training: anni 2019, 2020, 2021
 - *risoluzione 1000 m, nx= 241, ny= 233*
 - Campi di analisi da ARPA Lombardia anni 2019 (4 km ris.) e 2020 e 2021 (1 km ris.)
 - PM_{10} , $PM_{2.5}$, NO_2 , O_3 (medie mensili)
- ❑ EuroCC – Italia, Training: anno 2021
 - *risoluzione 1000 m, nx=1050, ny= 1320*
 - Campi FARM da QualeAria 12km di ris.
 - PM_{10} , $PM_{2.5}$ (medie giornaliere); NO_2 , O_3 , CO , C_6H_6 , SO_2 (medie orarie)
- ❑ Regione Sardegna: campi estratti da EuroCC – Italia
 - *risoluzione 1000 m, nx=157, ny=287*
- ❑ AIR Break, città di Ferrara, Training: Gennaio, Aprile, Luglio, Ottobre 2019
 - *risoluzione 200 m, nx= 1196, ny= 1046*
 - Campi FARM su dominio regionale 1 km ris.
 - PM_{10} , $PM_{2.5}$ (medie giornaliere); NO_2 , O_3 (medie orarie)
 - **testato l'utilizzo dei campi meteo WRF come predittori**

ESEMPIO di CRITICITA'



Mappa ottenuta
escludendo i
seguenti predittori:
DSEA, DHWY,
DXU, DLO
(distanza del centro
della gridbox dal
mare,
dall'**autostrada**,
dalla **strada**
extraurbana e
locale più vicina)

- Il Progetto BEEP (2017-2019)
- Modelli ML: schema di applicazione
- ✓ Valutazione della Qualità dell'aria sulla Regione Piemonte
- Campi di concentrazione sulla Regione Lombardia (EPIMED)
- EuroCC Italy: QualeAria Local
- Valutazione della Qualità dell'aria sulla Regione Sardegna
- Progetto «Air Break»

CONTESTO

Applicazione di due algoritmi ML: RF (**Random Forest**) e XGB (**XGBoost**) ai campi modellistici prodotti da ARPA Piemonte sul territorio regionale al fine di produrre campi di concentrazione ad 1 km di risoluzione spaziale relativi ai seguenti inquinanti: PM₁₀, PM_{2.5}, NO₂ ed O₃.

SCORES

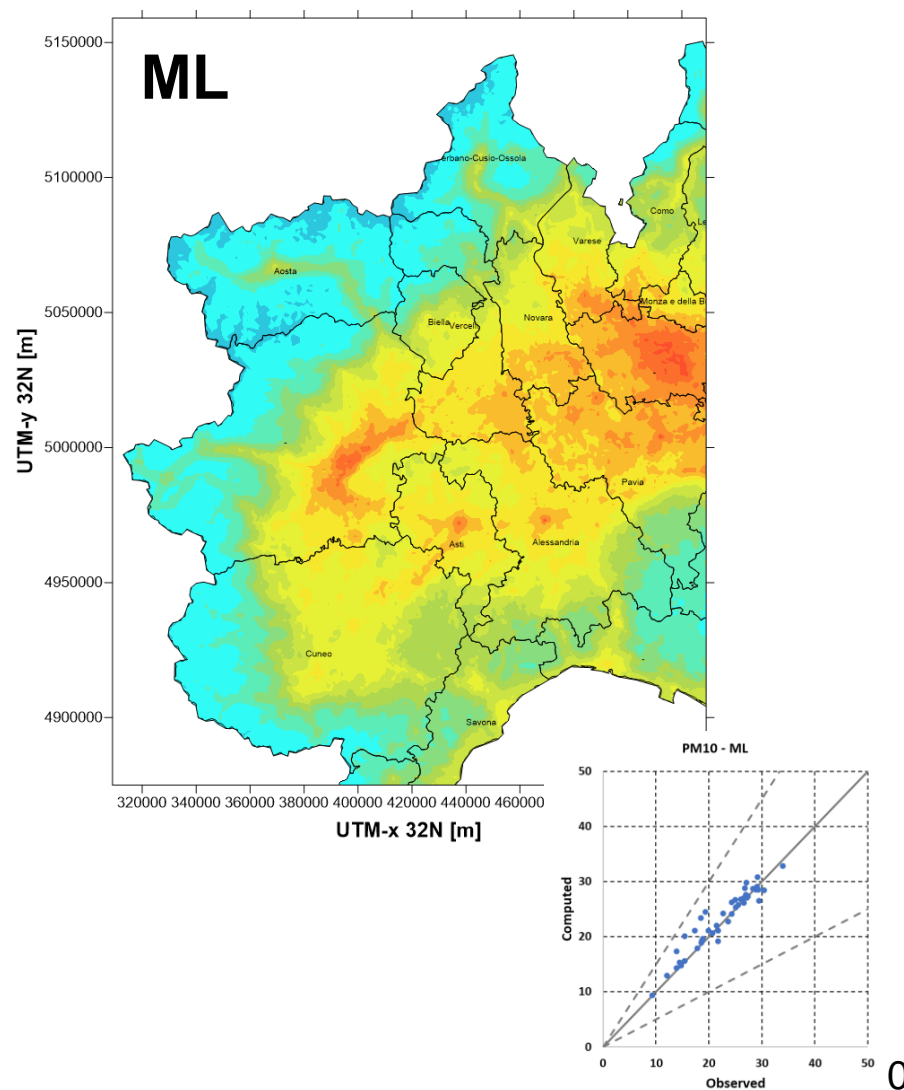
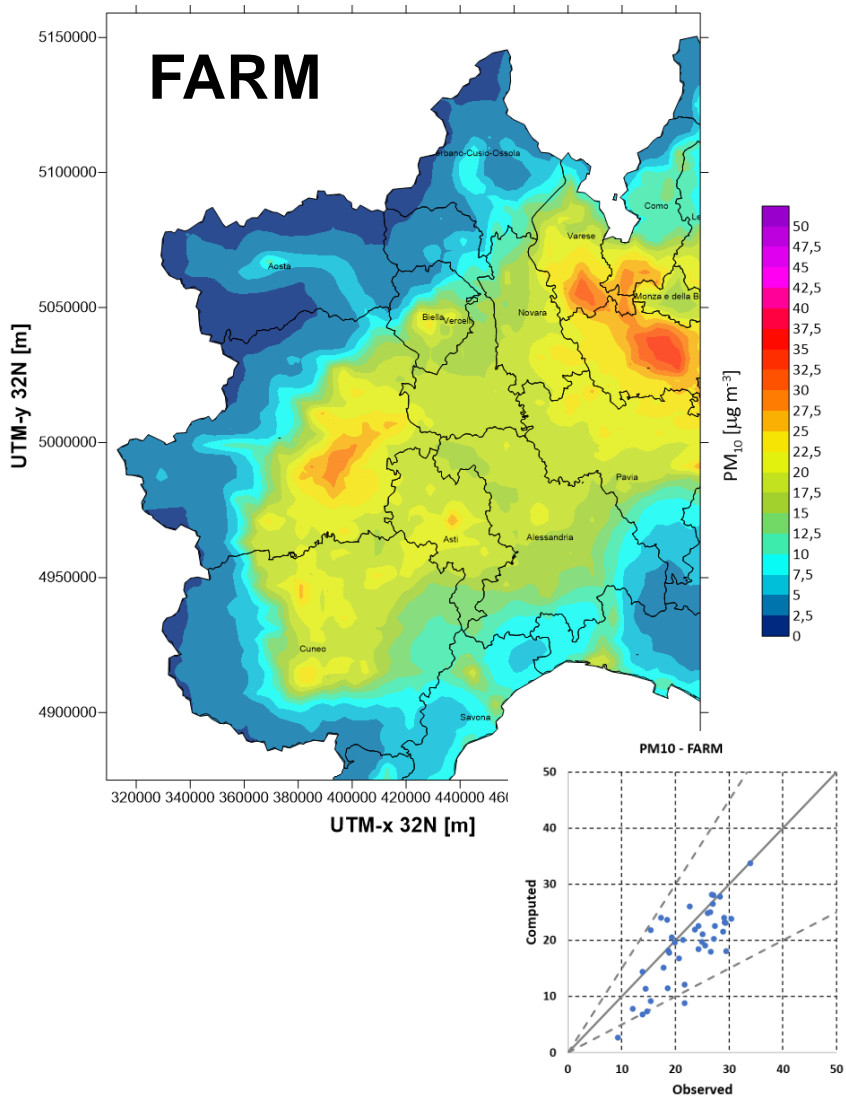
Per valutare il successo di tali algoritmi nell'avvicinare la stima ai dati è stato usato il Root-Mean-Square Error (RMSE), ovvero la radice quadrata dello scarto quadratico medio tra stima e osservazioni.

Nella tabella seguente sono riportati il RMSE di FARM e quelli ottenuti dai due algoritmi in fase di ottimizzazione e come Cross Validation (CV) score al 10 % (10-fold Cross-Validation "by monitor").

MODELLO	NO ₂	O ₃	PM ₁₀	PM _{2.5}
FARM	17.87 (1.00)	28.98 (1.00)	11.22 (1.00)	8.34 (1.00)
RF – addestramento	7.80 (0.44)	9.76 (0.34)	6.44 (0.57)	5.19 (0.62)
RF – CV score 10%	14.59 (0.82)	17.22 (0.59)	7.98 (0.71)	6.62 (0.79)
XGB – addestramento	10.43 (0.58)	14.56 (0.50)	6.28 (0.56)	4.88 (0.59)
XGB– CV score 10%	14.88 (0.83)	18.97 (0.65)	7.92 (0.71)	6.42 (0.77)

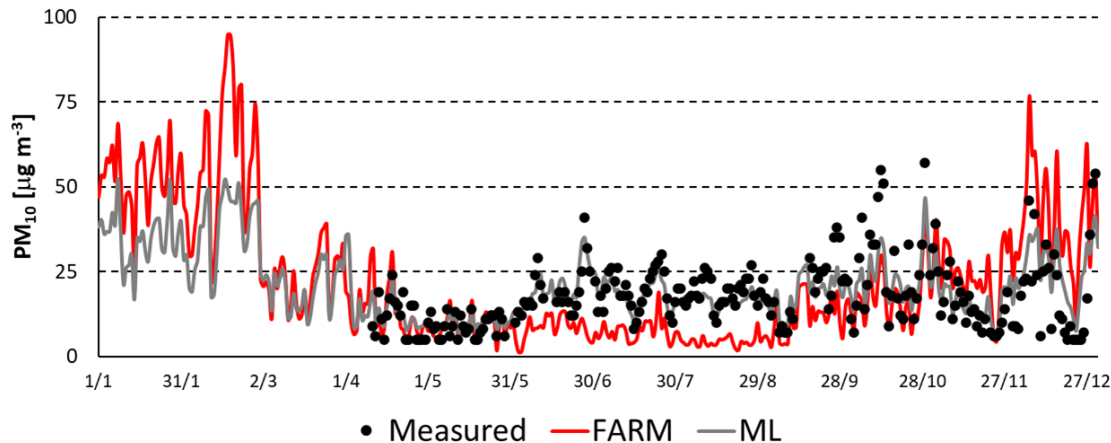
Valutazione della Qualità dell'aria sulla Regione Piemonte

PM₁₀



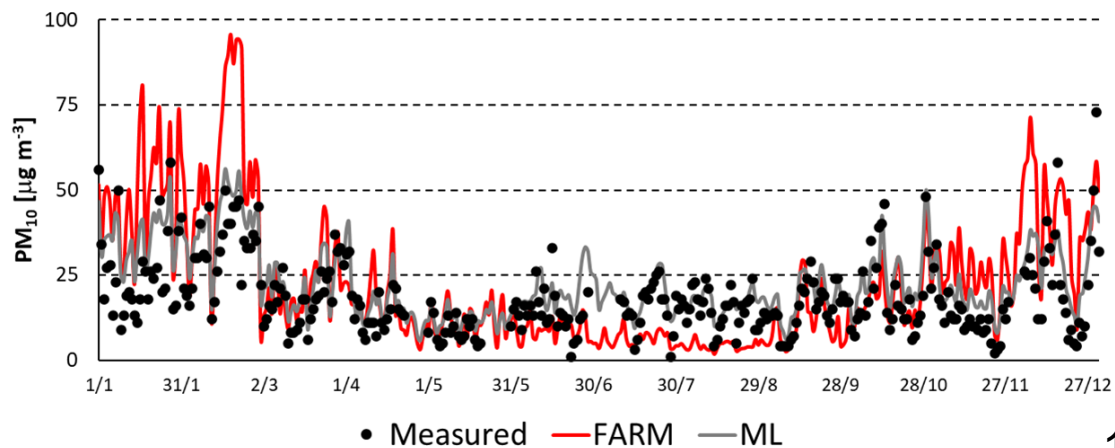
PM₁₀

BALDISSERO TORINESE (ACEA) - PARCO - (PIEMONTE, F-R)



Confronti con osservazioni non usate nella calibrazione del modello ML

Pinerolo - (PIEMONTE, F-U)



- Il Progetto BEEP (2017-2019)
- Modelli ML: schema di applicazione
- Valutazione della Qualità dell'aria sulla Regione Piemonte
- ✓ Campi di concentrazione sulla Regione Lombardia (EPIMED)
- EuroCC Italy: QualeAria Local
- Valutazione della Qualità dell'aria sulla Regione Sardegna
- Progetto «Air Break»

CONTESTO

Il Centro Ricerche in Epidemiologia e Medicina Preventiva (**EPIMED**), afferente al Dipartimento di Medicina e Chirurgia dell'Università degli Studi di Varese, ha attivato una Convenzione con Regione Lombardia dal titolo "Associazione tra inquinamento ambientale e COVID-19 in Provincia di Varese».

The screenshot shows the website of the University of Insubria. The header includes navigation links: FUTURO STUDENTE, STUDENTE, LAUREATO, ENTI E AZIENDE, PERSONALE, INTERNATIONAL, and LINK VELOCI. The main content area features the university logo and the title of a news article: "L'INQUINAMENTO ATMOSFERICO AUMENTA IL RISCHIO DI INFEZIONE DA COVID-19: STUDIO EPIMED SULLA CITTÀ DI VARESE". The article is dated 11 gennaio 2022. The text of the article states: "L'esposizione a lungo termine all'inquinamento atmosferico può aumentare il rischio di infezione da SARS-CoV-2: lo suggerisce una ricerca condotta da Epimed, il Centro di Epidemiologia e medicina preventiva dell'Università dell'Insubria i cui risultati sono pubblicati oggi online sulla rivista Occupational & Environmental Medicine, del gruppo editoriale Bmj. Lo studio, relativo alla popolazione adulta della città di Varese (62.848 persone), seguita nel tempo da inizio pandemia a marzo 2021, segnala un aumento del 5 per cento nel tasso di infezione per incremento di 1 microgrammo/metrocubo di PM2.5, 294 casi in più ogni centomila persone/anno." A small photo of a person at a computer is visible on the right side of the article.

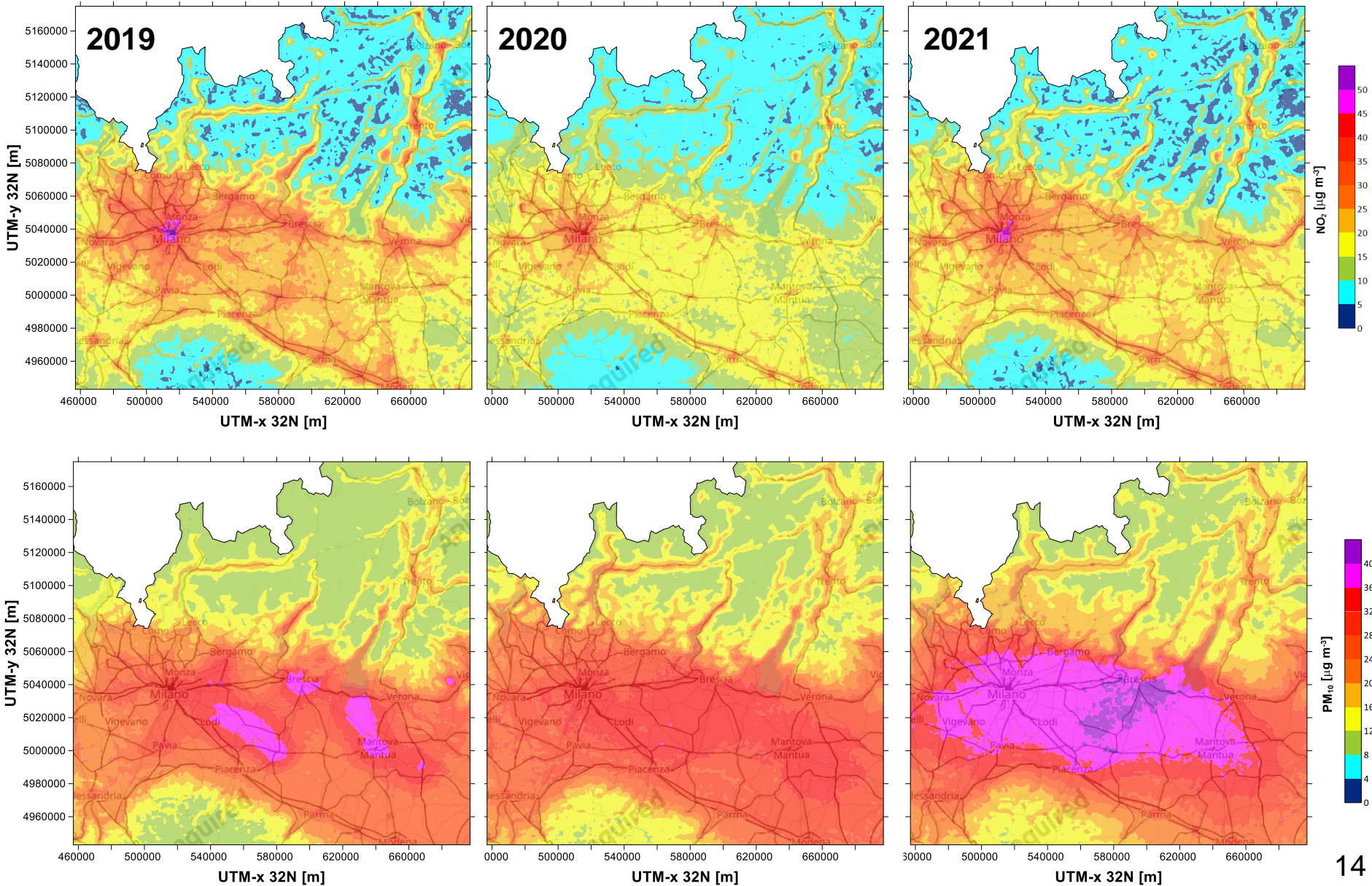
Per eseguire le attività previste da tale Convenzione è necessario disporre del dato relativo all'esposizione ambientale di lungo-periodo dei residenti in Provincia di Varese ai seguenti inquinanti $PM_{2.5}$, PM_{10} , NO_2 ed O_3 .

A partire dai campi modellistici medi mensili prodotti da ARPA Lombardia, relativi all'intera regione ed agli anni 2019, 2020 e 2021, ARIANET ha utilizzato algoritmi ML allo scopo di produrre campi corrispondenti alla risoluzione di 1 km sull'intero territorio regionale e quindi sulla Provincia di Varese.

Campi medi mensili (marzo) sulla Regione Lombardia (EPIMED)



UNIVERSITÀ DEGLI STUDI
DELL'INSUBRIA



- Il Progetto BEEP (2017-2019)
- Modelli ML: schema di applicazione
- Valutazione della Qualità dell'aria sulla Regione Piemonte
- Campi di concentrazione sulla Regione Lombardia (EPIMED)
- ✓ EuroCC Italy: QualeAria Local
- Valutazione della Qualità dell'aria sulla Regione Sardegna
- Progetto «Air Break»

Cos'è EuroCC Italy?

EuroCC Italy è il centro di competenza italiano, all'interno del network EuroHPC costituito da trentatre centri di competenza, che si occupa di trasferimento tecnologico verso il mondo industriale.

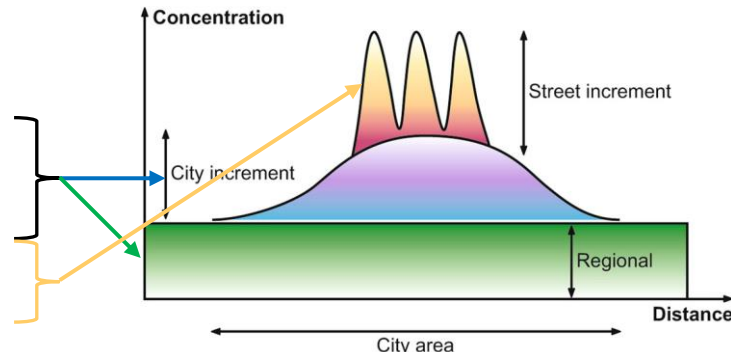
Obiettivi principali del progetto

- supportare le le PMI nei progetti di innovazione attraverso le tecnologie di supercalcolo (**HPC**), analisi dei big data (**HPDA**) e intelligenza artificiale (**IA**)
- creare consapevolezza circa i vantaggi produttivi di tali tecnologie

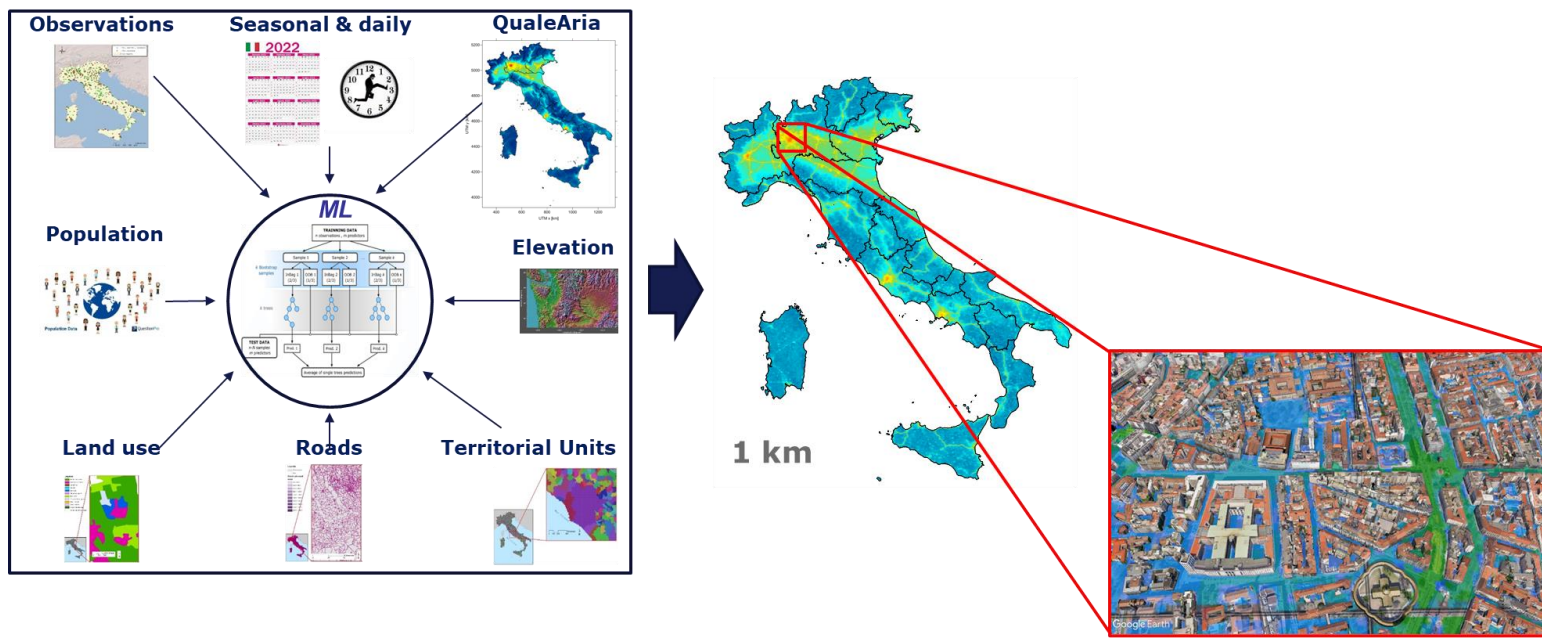
THE SOLUTION

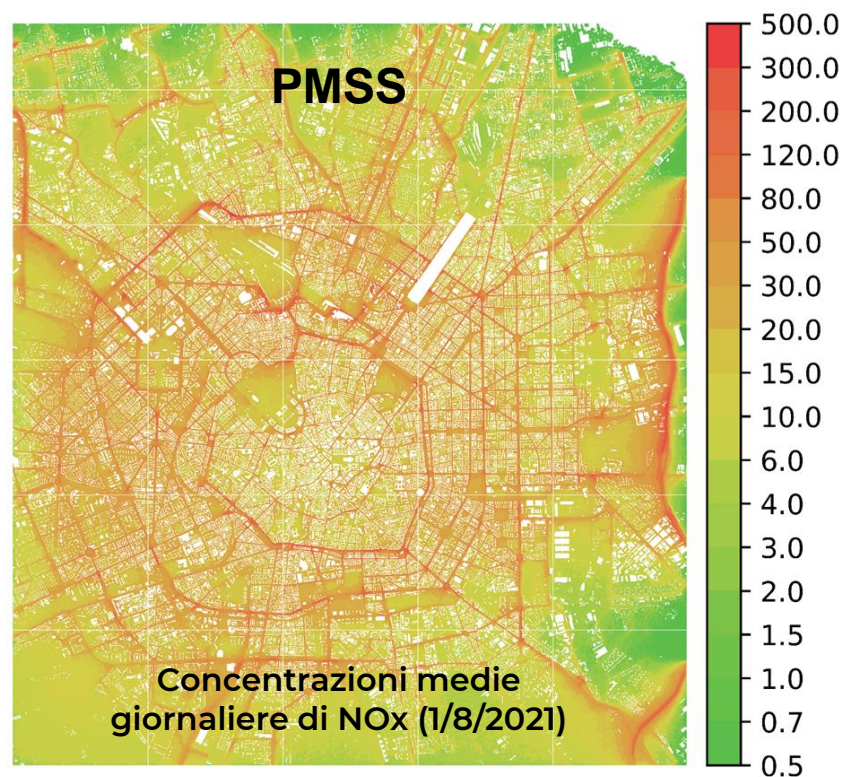
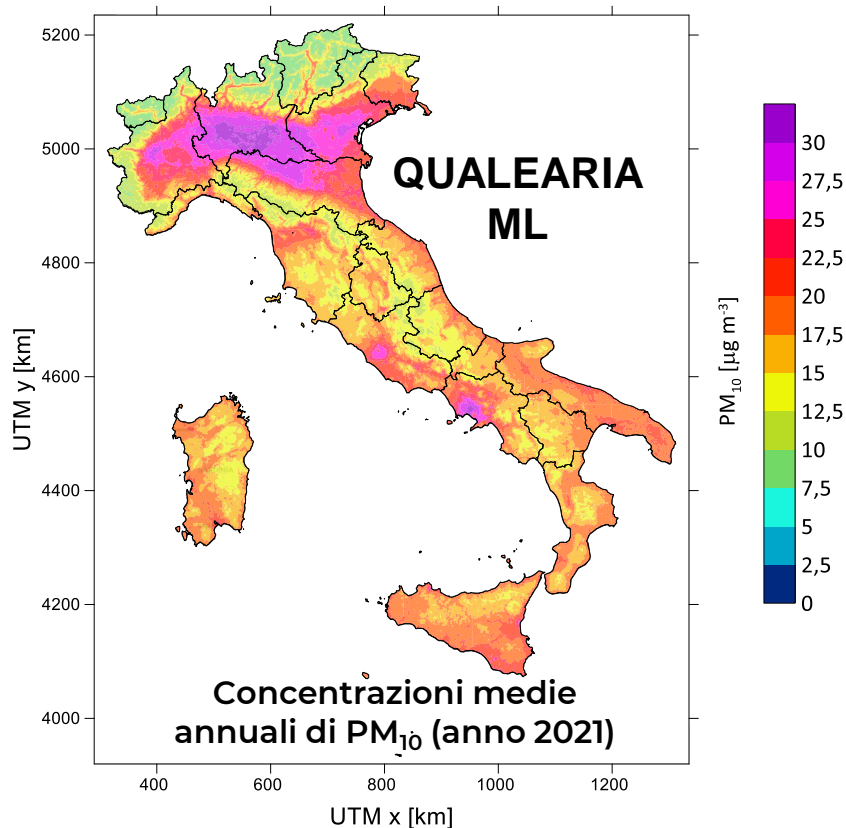
QualeAria-Local consists of three elements:

1. QualeAria Air Quality Forecast System (AQFS)
2. Random Forest Machine Learning (ML-RF) algorithm
3. μ -scale Parallel-Micro-Swift-Spray (PMSS) modelling suite.



The PM_{10} , $PM_{2.5}$, NO_2 and O_3 concentration fields produced by Qualearia for a test period (year 2021), together with a set of spatial-temporal predictors and available air quality observations were processed by the ML-RF algorithm to obtain corresponding higher resolution (1 km) concentration over the national territory. The resulting concentration fields provide the regional and urban background fields to the PMSS suite which is used to produce forecasts of atmospheric particulate matter and nitrogen oxides at very high spatial resolution (4 m) over a Milan urban area (<https://www.euroccitaly.it/qualearia-local-2/>).





Mediante modelli ML è possibile produrre previsioni di qualità dell'aria ad elevate risoluzioni spaziali con un onere computazionale inferiore rispetto a quello necessario per simulazioni con CTM. Particolarmente delicata è l'identificazione di un insieme coerente di predittori spazio-temporali, che si suppone siano correlati con la qualità dell'aria. I campi di concentrazione prodotti dal CTM e gli indicatori stagionali sono risultati i più importanti predittori spazio-temporali considerati dall' algoritmo ML utilizzato.

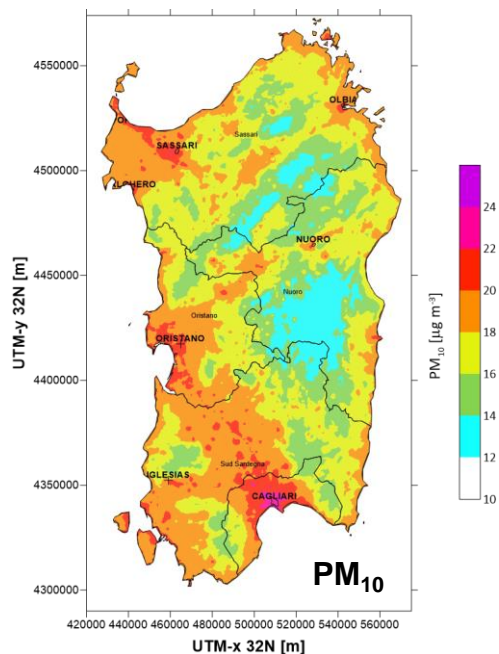
Il progetto ha anche confermato la necessità di risorse HPC per aumentare ulteriormente la risoluzione spaziale, fino ad alcuni metri, delle previsioni sulla qualità dell'aria nella e aree urbane. Ad esempio, una simulazione di 26 ore, effettuata mediante la suite modellistica **PMSS**, su un dominio di dimensioni pari a **10x11 km²**, che include l'area urbana milanese, alla risoluzione di **4 metri (2500x2750 punti griglia)** ha richiesto un tempo di calcolo pari a circa **90 minuti** utilizzando **391 core** sull'infrastruttura di calcolo **Galileo100**.

- Il Progetto BEEP (2017-2019)
- Modelli ML: schema di applicazione
- Valutazione della Qualità dell'aria sulla Regione Piemonte
- Campi di concentrazione sulla Regione Lombardia (EPIMED)
- EuroCC Italy: QualeAria Local
- ✓ Valutazione della Qualità dell'aria sulla Regione Sardegna
- Progetto «Air Break»

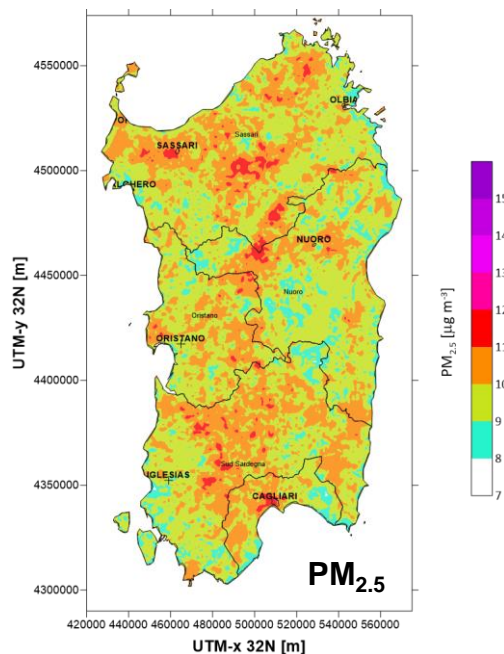
CONTESTO

Applicazione dell'algoritmo ML RF (**Random Forest**) per la produzione di campi di concentrazione di PM_{10} , $PM_{2.5}$, NO_2 , O_3 , benzene, CO e SO_2 spazializzate sul territorio regionale, alla risoluzione di 1 km, per l'anno 2021. In questa applicazione sono stati utilizzati i campi di concentrazione prodotti sul territorio nazionale dal sistema di previsione **QualeAria**.

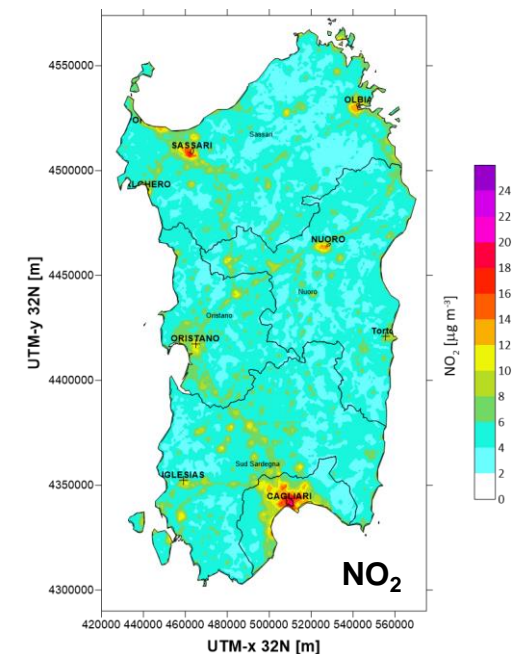
Media annuale – limite $40 \mu g m^{-3}$



Media annuale – limite $25 \mu g m^{-3}$



Media annuale – limite $40 \mu g m^{-3}$



Valutazione della Qualità dell'aria sulla Regione Sardegna



SardegnaArpa



IL SISTEMA AMBIENTALE DELLA SARDEGNA

- Arpas
 - Attività
 - Documentazione
 - SardegnaAmbiente
- Chi siamo Organigramma Posta elettronica certificata Albo pretorio Bandi e gare ...

sardegnaarpa > arpas > notizie > valutazioni modellistiche della qualità dell'aria

ARPAS

- Chi siamo
- Organigramma
- Posta elettronica certificata
- Albo pretorio
- Bandi e gare
- Concorsi e selezioni
- Cessioni gratuite di beni
- Comitato Unico di Garanzia
- > Notizie
 - Emergenza Covid-19
- Amministrazione Trasparente

Valutazioni modellistiche della qualità dell'aria



<https://www.sardegnaambiente.it/index.php?xsl=612&s=446293&v=2&c=4581&idsito=21>



REGIONE AUTONOMA DI SARDEGNA
REGIONE AUTONOMA DELLA SARDEGNA
AGENZIA REGIONALE PRO S'AMPARU DE S'AMBIENTE DE SARDIGNA
AGENZIA REGIONALE PER LA PROTEZIONE DELL'AMBIENTE DELLA SARDEGNA
ARPAS

Direzione Tecnico-Scientifica
Servizio Controlli, Monitoraggi e Valutazione Ambientale
Linea di Attività Monitoraggio Qualità dell'Aria

Valutazione modellistica dello stato di Qualità dell'aria della Sardegna

Anno 2021

Gennaio 2023

https://www.sardegnaambiente.it/documenti/21_393_20230207093545.pdf

- Il Progetto BEEP (2017-2019)
- Modelli ML: schema di applicazione
- Valutazione della Qualità dell'aria sulla Regione Piemonte
- Campi di concentrazione sulla Regione Lombardia (EPIMED)
- EuroCC Italy: QualeAria Local
- Valutazione della Qualità dell'aria sulla Regione Sardegna
- ✓ Progetto «Air Break»

CONTESTO

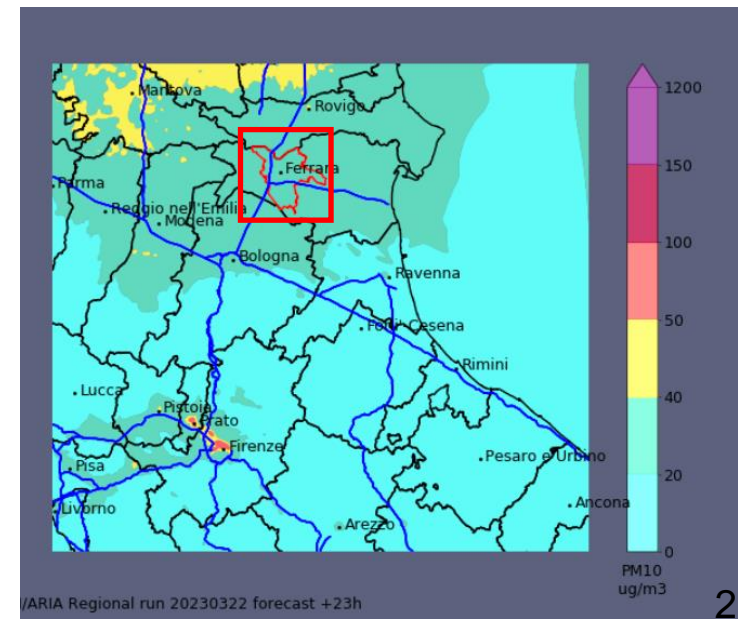
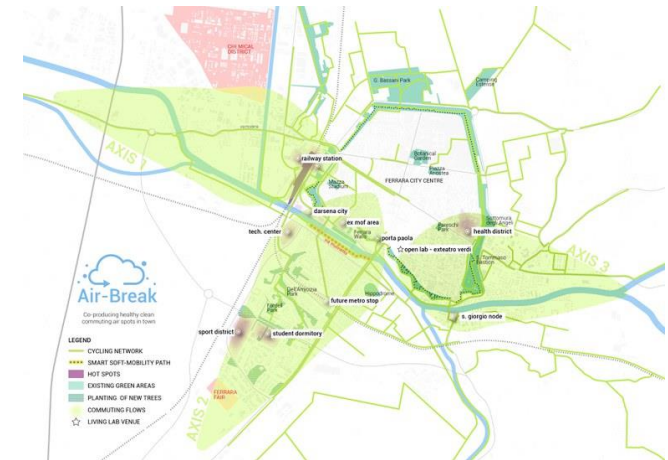
Il Progetto "AIR BREAK - Co-producing healthy clean commuting air spots in town" è stato presentato da **Hera** assieme al **Comune di Ferrara** nell'ambito di un bando emesso dal programma **Urban Innovative Actions (UIA)** con fondi provenienti dalla Comunità Europea.

Tale progetto mira a **ridurre l'inquinamento atmosferico del 25%** in aree selezionate della città di **Ferrara** e ad affrontare questioni ad esso correlate.

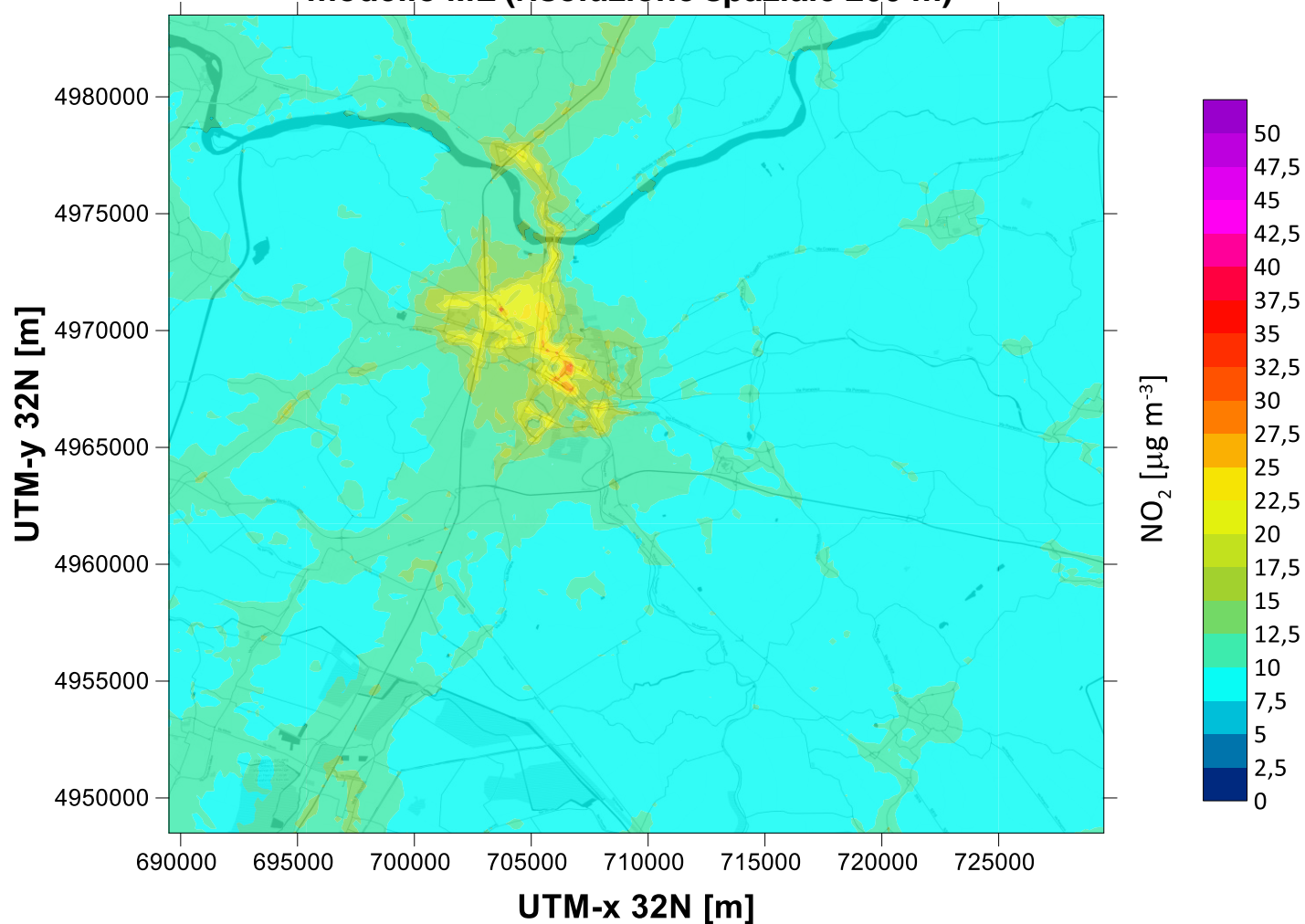
Uno dei principali deliverable del progetto AIR BREAK è costituito dalla realizzazione di una **dashboard, web-based**, che include le informazioni fornite:

- dal monitoraggio effettuato in continuo da una **rete di microsensori** di misura delle concentrazioni di inquinanti atmosferici
- da un **sistema modellistico atmosferico (SMA)** in grado di produrre **previsioni di qualità dell'aria a scala regionale**. **Le previsioni sull'area urbana di Ferrara a 200 m di risoluzione vengono prodotte mediante l'applicazione di un modello "Machine learning"**.

Le aree urbane di intervento previste dal progetto di ricerca Air-Break



Esempio di mappa di concentrazione di NO₂ prodotta sulla città di Ferrara ottenuta mediante l'applicazione di un modello ML (risoluzione spaziale 200 m)





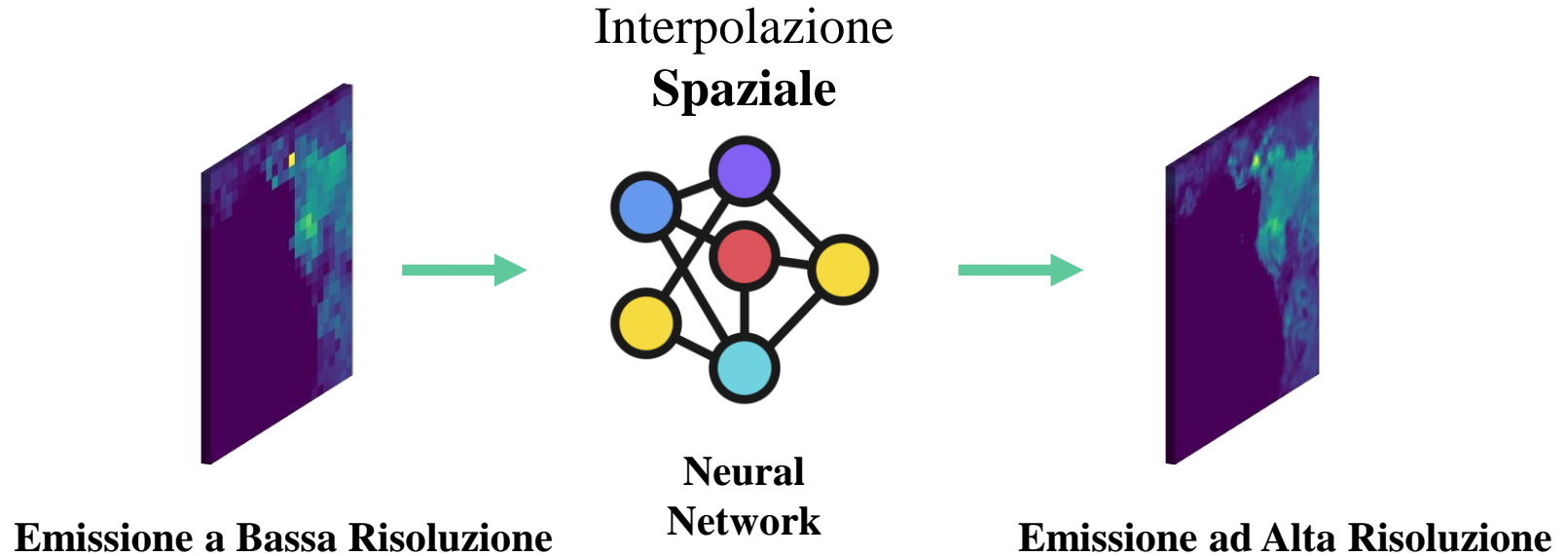
Esperienze di utilizzo di modelli AI per migliorare la risoluzione spaziale di mappe di emissione di BVOC

Antonio Giganti, Paolo Bestagini, Marco Marcon



**X giornata sulla modellistica in aria(net)
Milano, 29 marzo 2023**

PROBLEMA



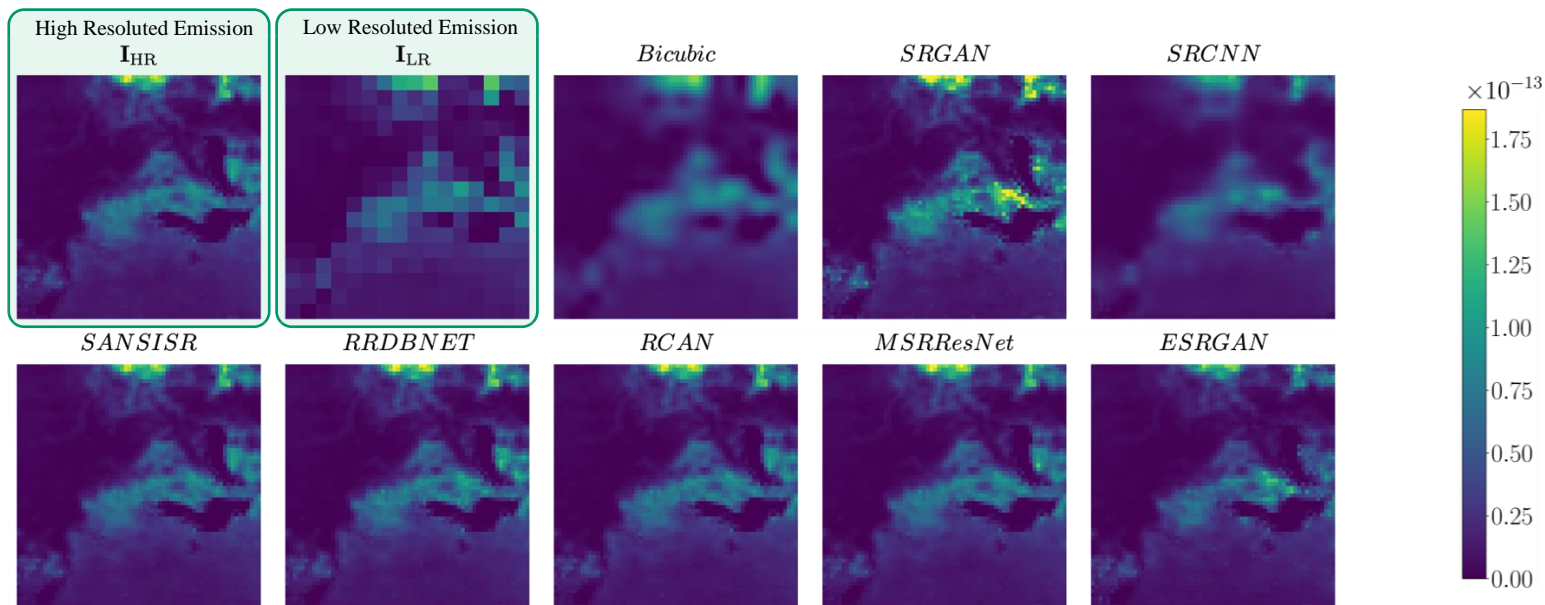
DATI

CAMS-GLOB-BIO [1] :

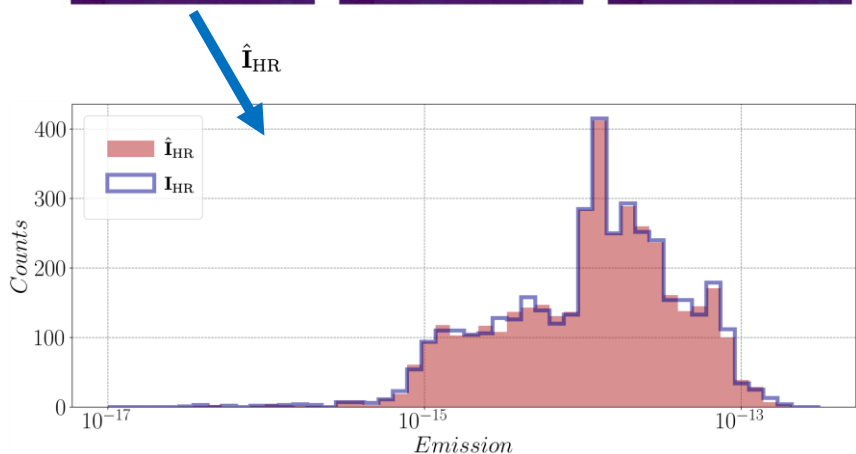
- Biogenic Parameters: 29 compounds
- Model: *MEGANv.2.1* on *ERA5* meteorology (reanalysis of ECWMF)
- **Spatial coverage:** Global [lat.min:-90° lat.max:90° lon.min:-180° lon.max:180°]
- **Spatial resolution:** 0.25°x0.25°
- **Temporal coverage:** from 2000 to 2020
- Monthly mean fluxes [kg / m^2s] and monthly averaged daily (each hour) profile

RISULTATI PRELIMINARI

Studio comparativo dell'utilizzo di reti neurali per migliorare la risoluzione spaziale di mappe di emissione di BVOC.

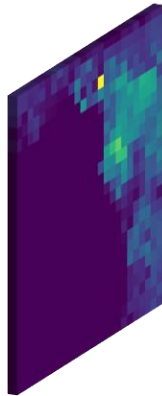


Mappe di emissione ad alta risoluzione stimate (\hat{I}_{HR}) da differenti reti neurali per super-risoluzione di immagini; Input I_{LR}

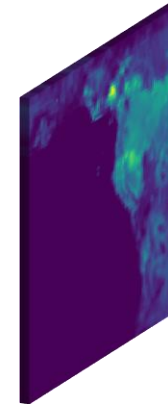


A.Giganti, S.Mandelli, P. Bestagini *et al.*,
 “Super-Resolution of BVOC Maps by
 Adapting Deep Learning Methods,”
arXiv preprint arXiv:2302.07570v2,
 2023.

Interpolazione Spaziale

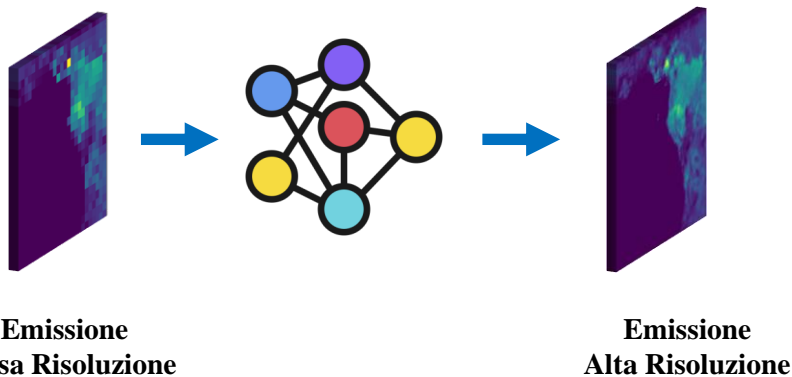


**Emissione
Bassa Risoluzione**

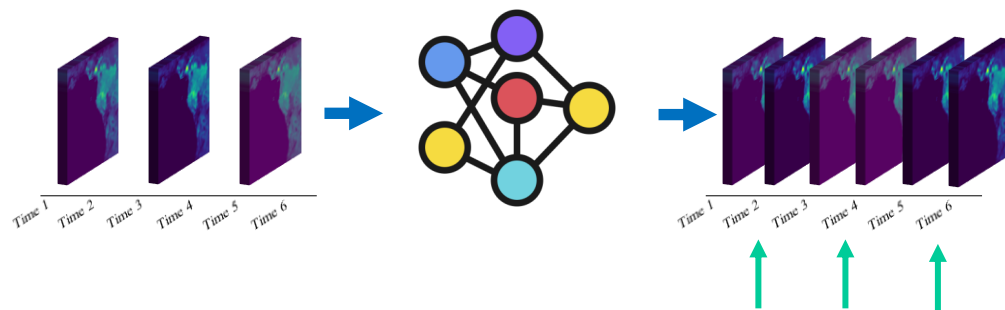


**Emissione
Alta Risoluzione**

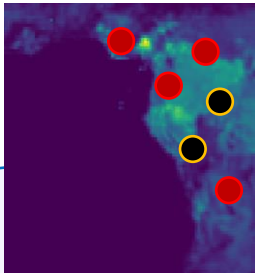
Interpolazione Spaziale



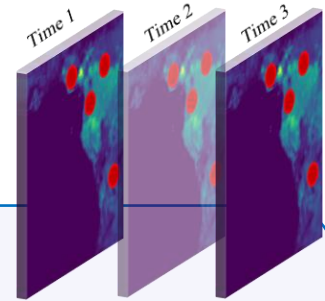
Interpolazione Temporale



combinare le 2 interpolazioni e
sfruttare entrambe le **informazioni**
SPAZIALI e **TEMPORALI**
nel processo di interpolazione



METODI AI-based



Interpolazione Spaziale

Graph Neural Network (GNN) [1]
rete neurale che lavora sui grafi.

Vantaggi:

- **Gestione dei dati irregolari:** grazie alla struttura dei grafi, si possono rappresentare le relazioni arbitrarie tra i dati, considerando diverse densità di campionamento (sparsity) in diverse aree geografiche;
- **Capacità di considerare la topologia della mappa,** quindi le connessioni spaziali tra i punti campionati;
- **Aggiornamento iterativo dei dati;**
- **Capacità di apprendere relazioni** complesse, come caratteristiche geologiche, altimetriche, etc.

Interpolazione Temporale

Transformer [2]

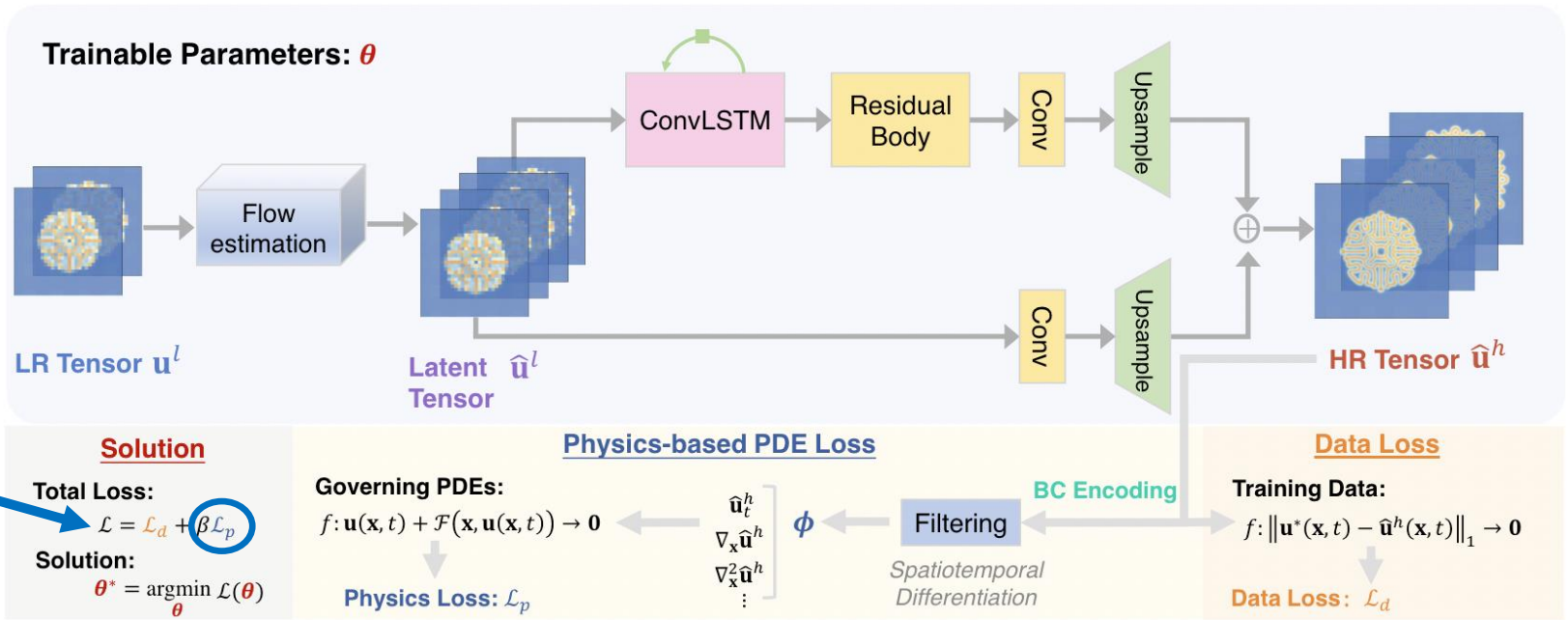
rete neurale per elaborare sequenze di dati

Vantaggi:

- **Gestione della dipendenza temporale,** grazie alla sua struttura interna;
- **Utilizzo di attenzione multi-head:** la rete è in grado di assegnare un peso differente a ciascun punto della sequenza in base alla sua importanza. Utile per l'interpolazione su serie temporali, poiché i punti vicini al punto che si vuole interpolare sono generalmente più importanti di quelli lontani;
- **Scalabilità:** la rete è altamente scalabile, il che significa che può gestire facilmente serie temporali di diverse lunghezze.

Oltre l'approccio data-driven nell'interpolazione spazio-temporale di mappe di emissione

Le reti neurali sono state spesso criticate per la loro mancanza di interpretabilità e la loro tendenza a essere delle "black-box". Negli ultimi anni si è cercato di incorporare le conoscenze fisiche all'interno del modello al fine di garantire trasparenza nei risultati ottenuti ed incrementare le performance.



[1] R.Pu et al. "Physics-informed Deep Super-resolution for Spatiotemporal Data", arXiv preprint, arXiv:2208.01462v1, 2022

Le **Physics-Informed Neural Network (PINN)** combinano le *conoscenze fisiche* con *l'apprendimento automatico*, al fine di risolvere problemi complessi che coinvolgono fenomeni fisici, come ad esempio le *emissioni di gas*, ma possono anche essere utilizzate come metodo per migliorare la precisione delle previsioni dei modelli fisici esistenti, e/o ridurre i tempi di calcolo delle simulazioni.



POLITECNICO
MILANO 1863

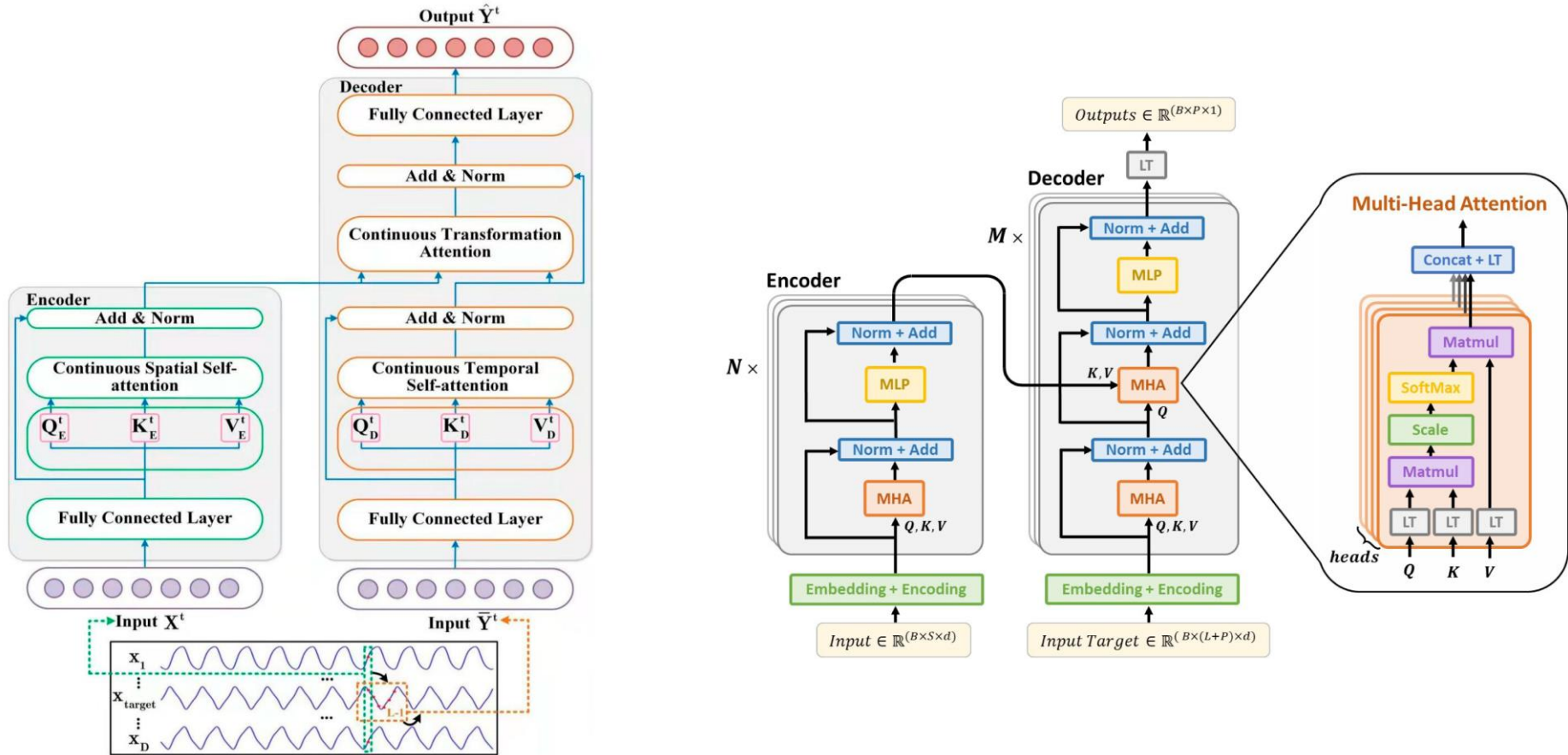


*Grazie per
l'attenzione*



Appendix

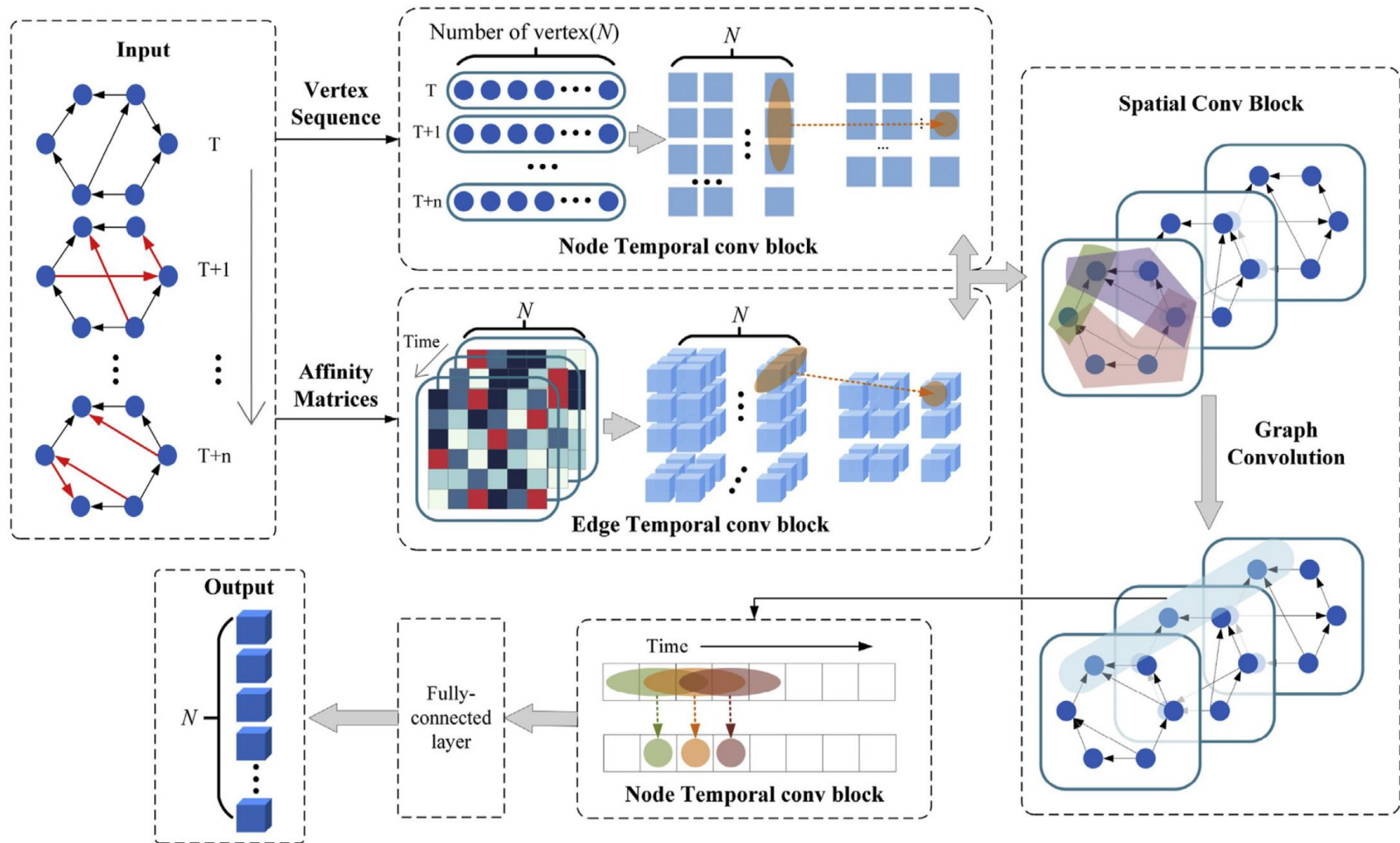
Transformer-based time series forecasting



You, Y., Zhang, L., Tao, P., Liu, S., & Chen, L. (2022). Spatiotemporal Transformer Neural Network for Time-Series Forecasting. *Entropy*, 24(11), Article 11. <https://doi.org/10.3390/e24111651>

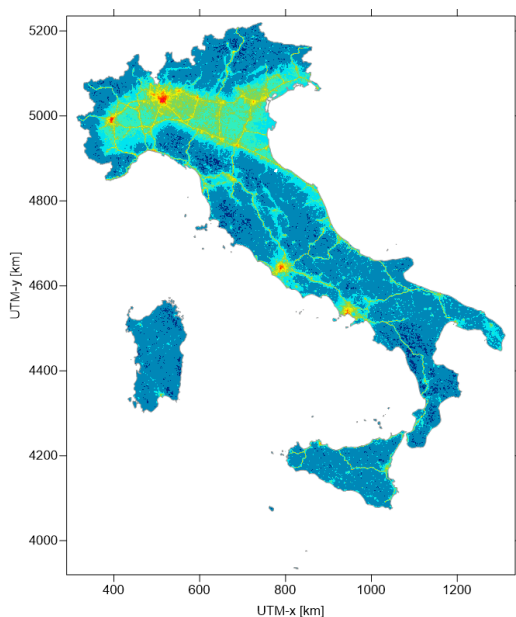
Bentsen, L. Ø., Warakagoda, N. D., Stenbro, R., & Engelstad, P. (2023). Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures. *Applied Energy*, 333, 120565. <https://doi.org/10.1016/j.apenergy.2022.120565>

GNN-based spatio-temporal forecasting



METODI

1. Utilizzo di un **sistema modellistico** per l'effettuazione di simulazioni triennali (**2013-2015**) meteorologiche e di qualità dell'aria sul **territorio nazionale** a **5 km** di risoluzione spaziale
2. Produzione di mappe ambientali (**NO₂** ed **O₃**) a più elevata risoluzione spaziale (**1km**) mediante l'utilizzo di modelli **Machine Learning**



METODI

1. Utilizzo dello stesso **sistema modellistico** per l'effettuazione di simulazioni triennali (**2013-2015**) meteorologiche e di qualità dell'aria su **6 aree urbane (Torino, Milano, Bologna, Roma, Bari e Palermo)** a **1 km** di risoluzione spaziale
2. Produzione di mappe ambientali (**PM₁₀**, **PM_{2.5}**, **NO₂** ed **O₃**) a più elevata risoluzione spaziale (**200 m**) mediante l'utilizzo di modelli **Machine Learning**

