



**Politecnico  
di Torino**



**CENTRALE  
LYON**



**Politecnico di Torino & École Centrale de Lyon**

Laurea magistrale in Ingegneria Matematica

Luglio, 2025

# **Development of a Data-Driven Surrogate Model for Street-Scale Prediction of NO<sub>x</sub> Passive Scalar Concentrations in the San Salvario District of Turin**

**Riccardo Kiefer**

**Relatori:**

Prof. L. Ridolfi

Dr.ssa S. Fellini

Dr. C. V. Nguyen (École Centrale)

Prof. P. Salizzoni (École Centrale)

# Inquinamento dell'aria

- Rischio ambientale
- Rischio **sanitario**



**Anziani e bambini** soggetti più a rischio

## **Impatto salute fisica e mentale:**

- 4,2 milioni: morti premature a livello globale (WHO)
- 240.000 solo in Europa (EEA)
- Riduzione dell'aspettativa di vita

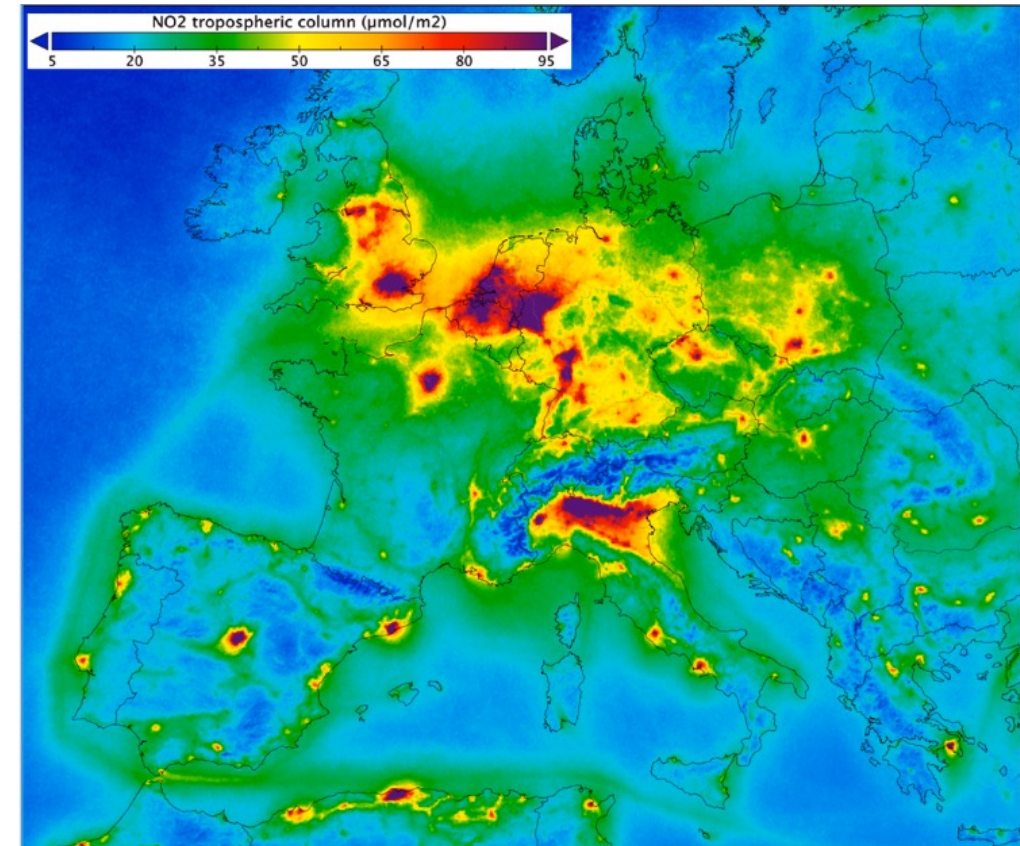
## **Impatto economico.**

- Da €268 a €428 miliardi di euro ogni anno (EEA)
- 2% pil europeo (2021)

# Inquinamento dell'aria - Torino

**Barriere naturali (Alpi, Appennini)**  
**Zona riccamente industrializzata**

- Città più inquinata d'Italia e d'Europa in molti periodi dell'anno
- Livelli inquinanti eccedono soglie europee per la tutela della salute pubblica

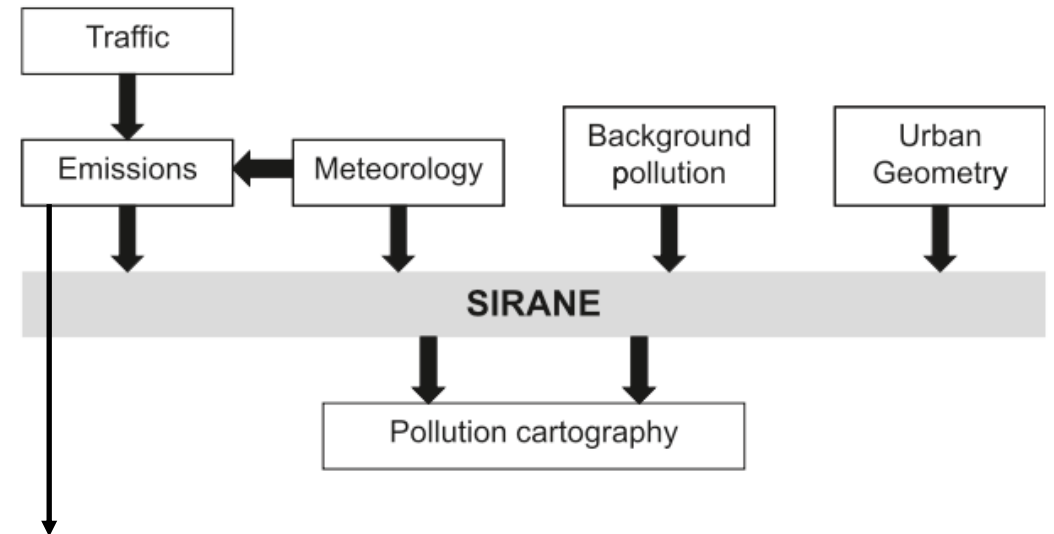
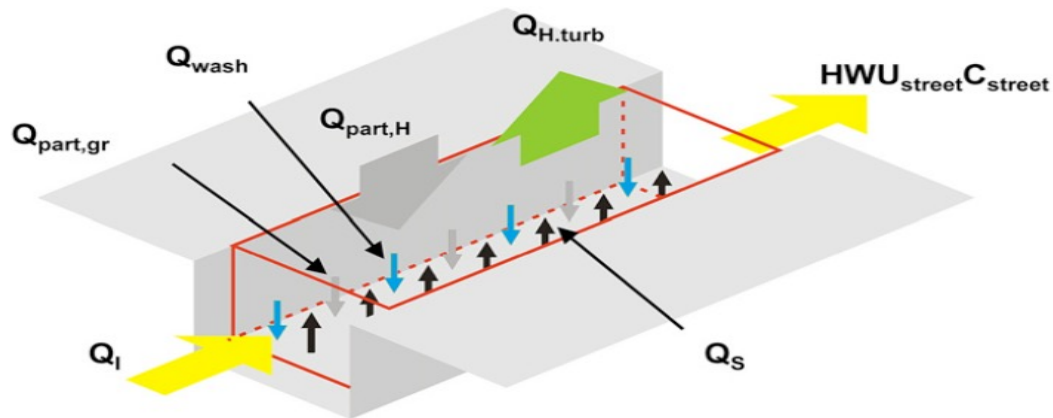


## Modello preesistente: SIRANE

Software operativo deterministico per simulare **dispersione aerea di inquinanti** su scala urbana ( $L \sim 10$  m)

Sfrutta:

- **Modello 3D** canyon urbano costruito a partire da mappa 2D.
- **Bilancio di massa** per modellare diffusione e trasporto a livello stradale.
- **Scambio turbolento** tra bassa atmosfera e strade



- **Lineari** (traffico) – modulazione oraria
- **Superficiali** (riscaldamento, industrie)
- **Puntuali** (ciminiere, sottopassaggi)

## Scopo della tesi

Sviluppo di un **modello** statistico **semplificato** per **stimare** la **concentrazione** a livello **stradale** di  $\text{NO}_x$

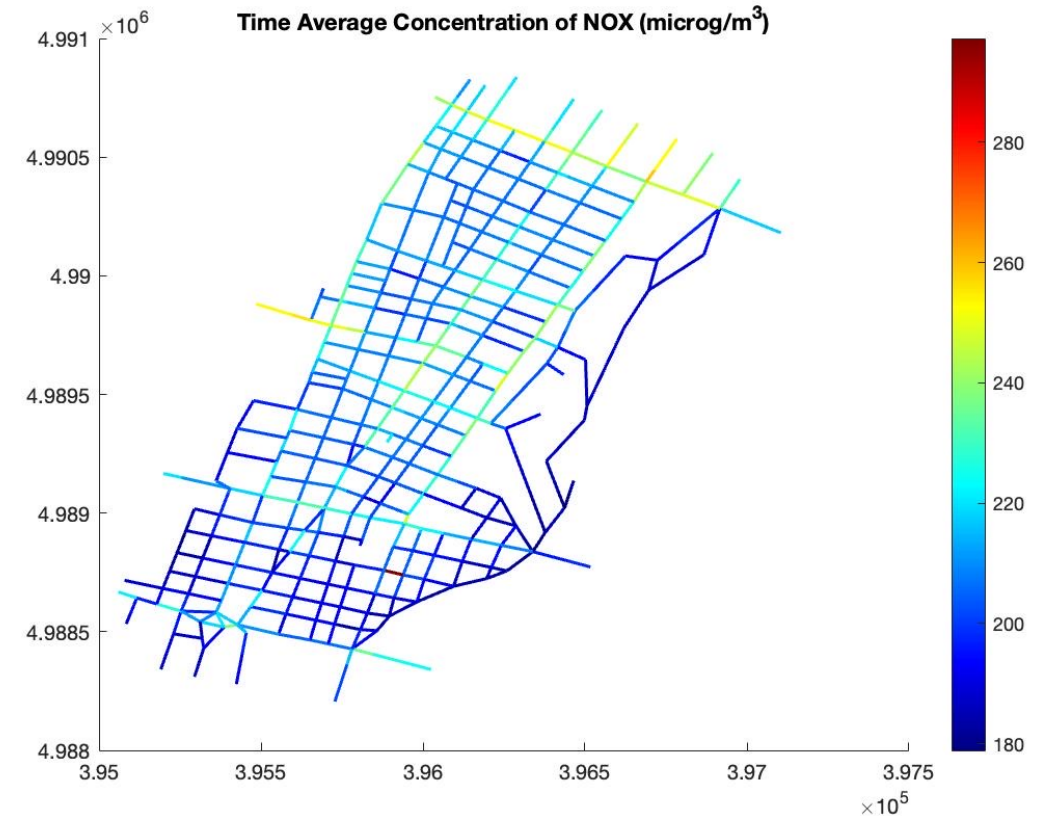
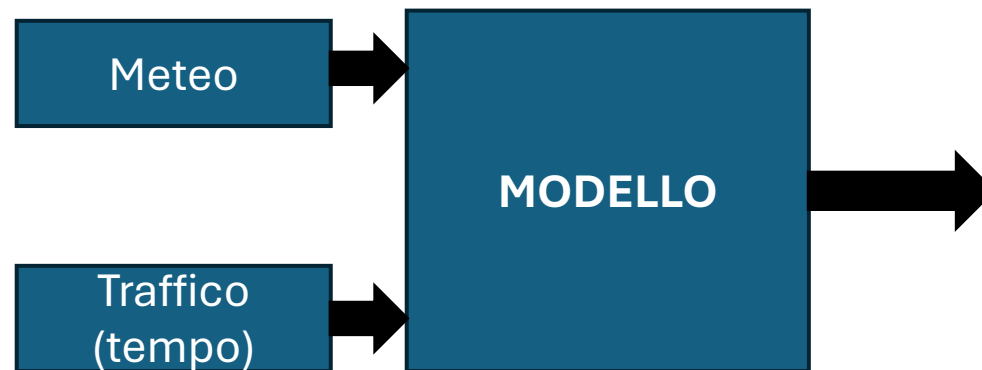
- **Previsione rapida** delle concentrazioni inquinanti
- Supporto alla **pianificazione urbana**
- **Identificazione** delle **strade** a rischio sanitario elevato
- **Scalabilità** del modello per scenari futuri
- Strumento per valutare **politiche** di **riduzione** delle **emissioni**
- **Integrazione** con **sistemi** di **alerta** e informazione al cittadino

$\text{NO}_x$  si comporta come **scalare passivo**  $\Rightarrow$  no reazioni chimiche non lineari

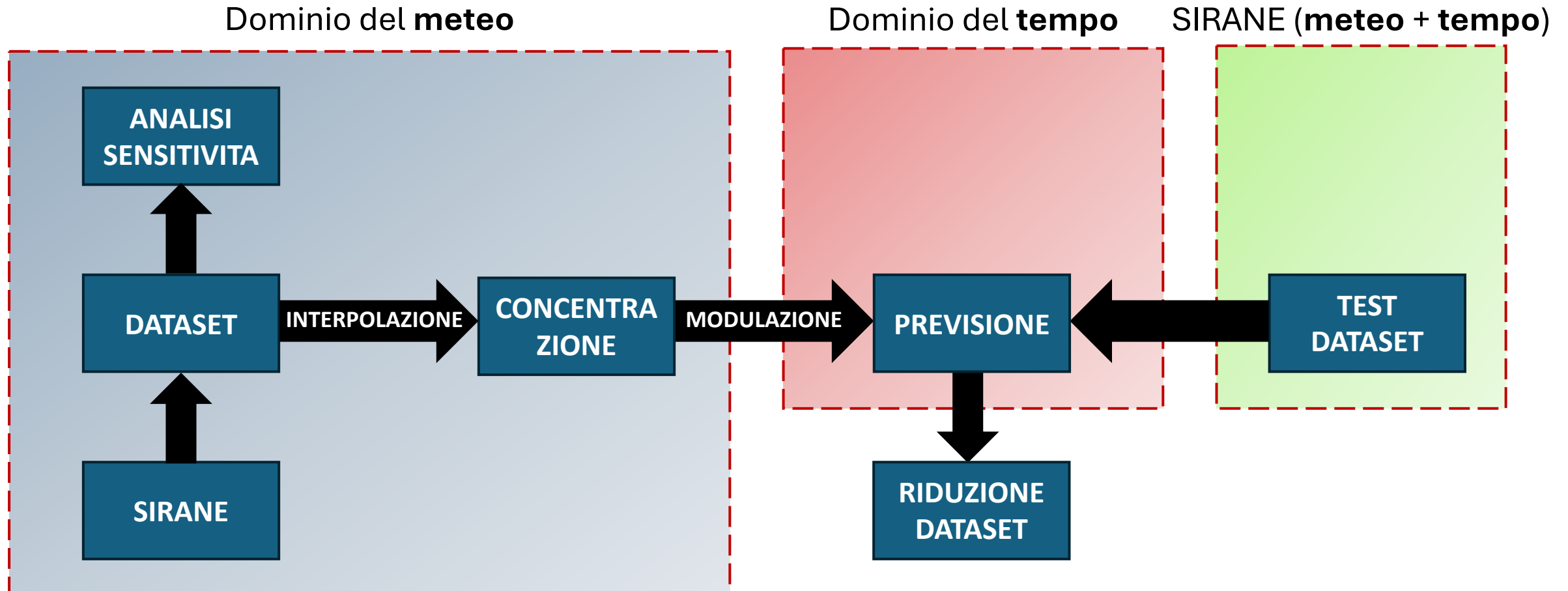
# Scopo della tesi

Scelto come caso studio il quartiere di **San Salvario**:

1. Geometria griglia regolare
2. Facile da isolare (ferrovia ad ovest, fiume ad est)
3. Dominio ridotto, tempi computazionali minori



## Work flow – Riduzione dataset

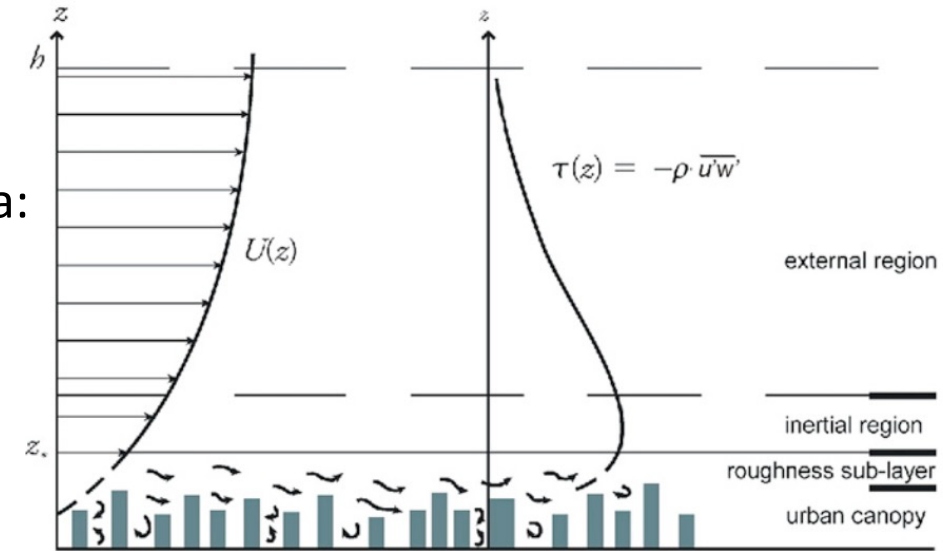


# Fisica meteorologica

**Friction velocity** ( $u_*$ ), è uno sforzo di taglio che si correla direttamente a:

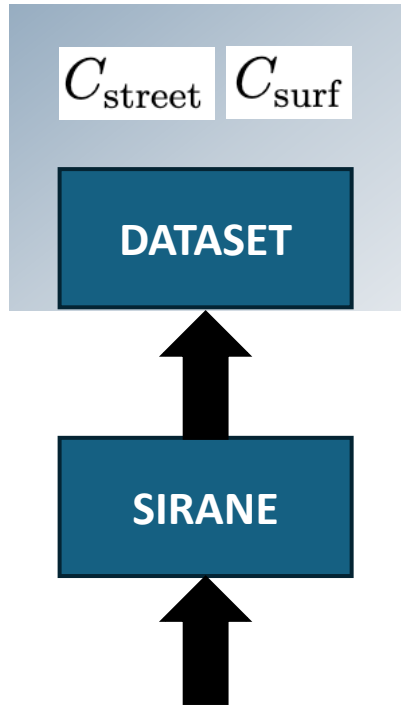
- **velocità** 'effettiva' del vento vicino al suolo
- quanto il vento riesce a **generare turbolenza**

Vento forte  $\Rightarrow$  Maggiore **shear**  $\Rightarrow$  Maggiore mescolamento aria



**Lunghezza di Monin-Obukhov** ( $L_{MO}$ ), parametro di **stabilità** dell'atmosfera:

- $L_{MO} < 0 \rightarrow$  atmosfera **instabile**: prevale trasporto verticale moti convettivi (buoyancy)
- $L_{MO} > 0 \rightarrow$  atmosfera **stabile**: prevale turbolenza meccanica vento (shear)
- $L_{MO} \rightarrow \infty \rightarrow$  atmosfera **neutra**: conta solo turbolenza meccanica vento (shear)

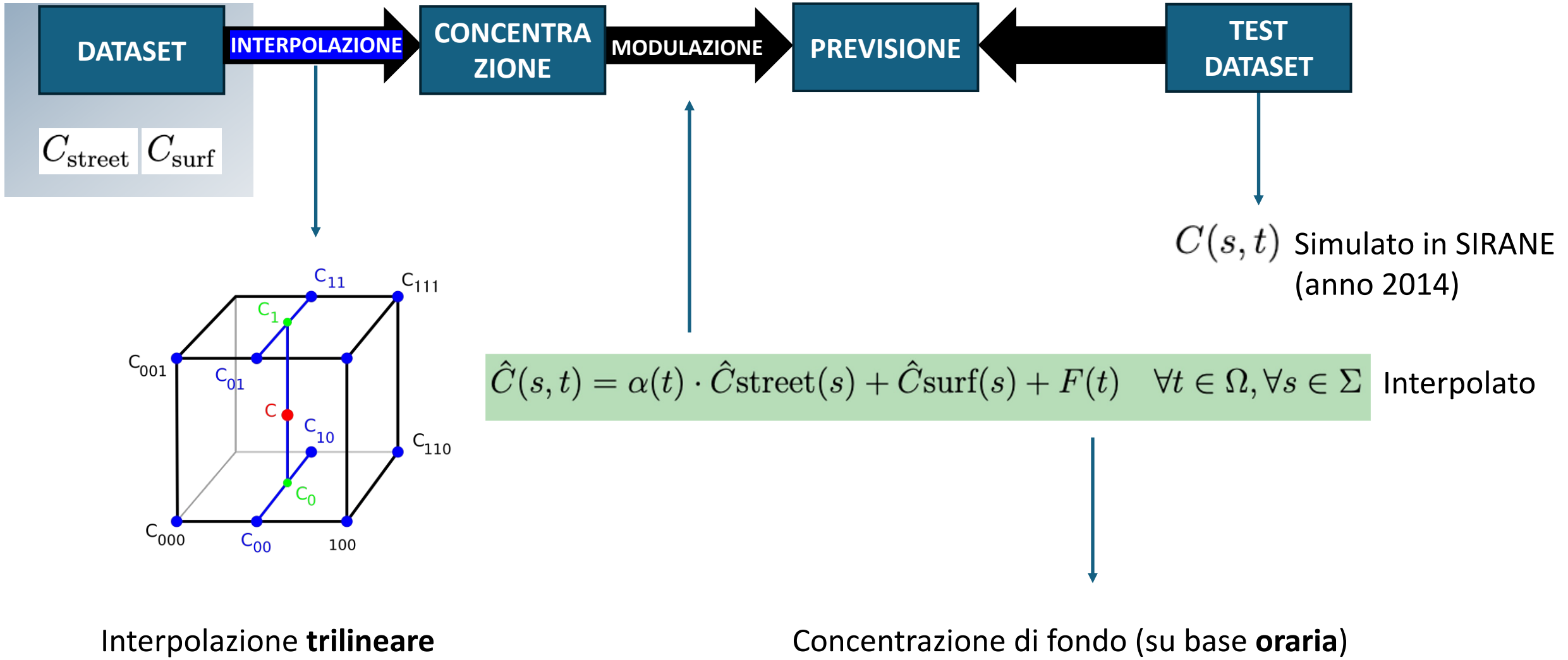


Focus **4 parametri** meteorologici. Simulazione concentrazioni al variare dei parametri in uno spazio delle fasi discreto:

- $L_{MO}^{-1}: [-0.3, -0.2, -0.1, -0.004, -0.019, 0, 0.004, 0.012, 0.05] 1/m \in I$
- $T: [-20: 5: 45]^\circ \in T$
- $Dir: [0: 5: 360]^\circ \in \Theta$
- $u_*: [0.05: 0.05: 1.2] m/s \in \Phi$

$$C_{street}(s) = f(s, Dir, T, u_*, L_{MO}^{-1}), \quad C_{surf}(s) = f(s, Dir, T, u_*, L_{MO}^{-1}) \quad \forall s \in \Sigma$$

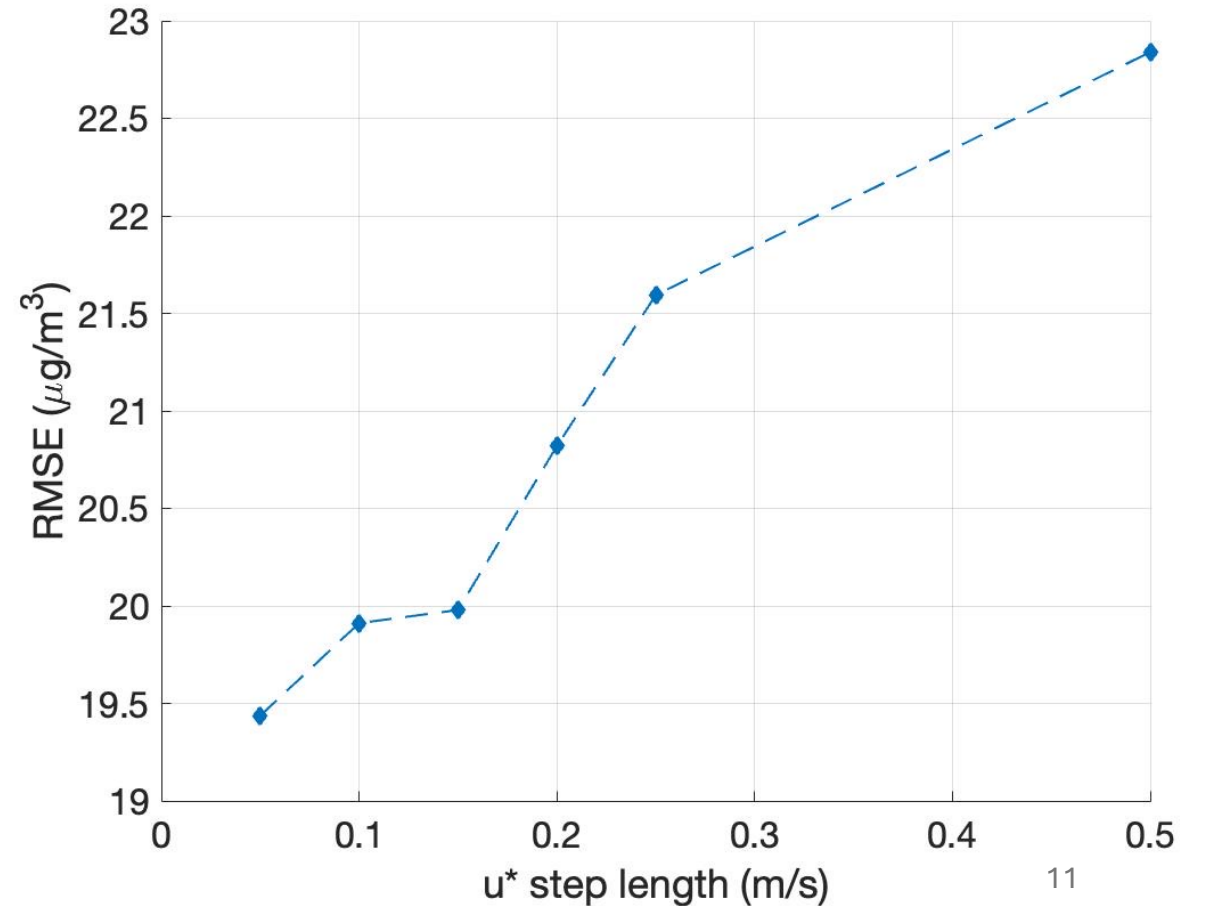
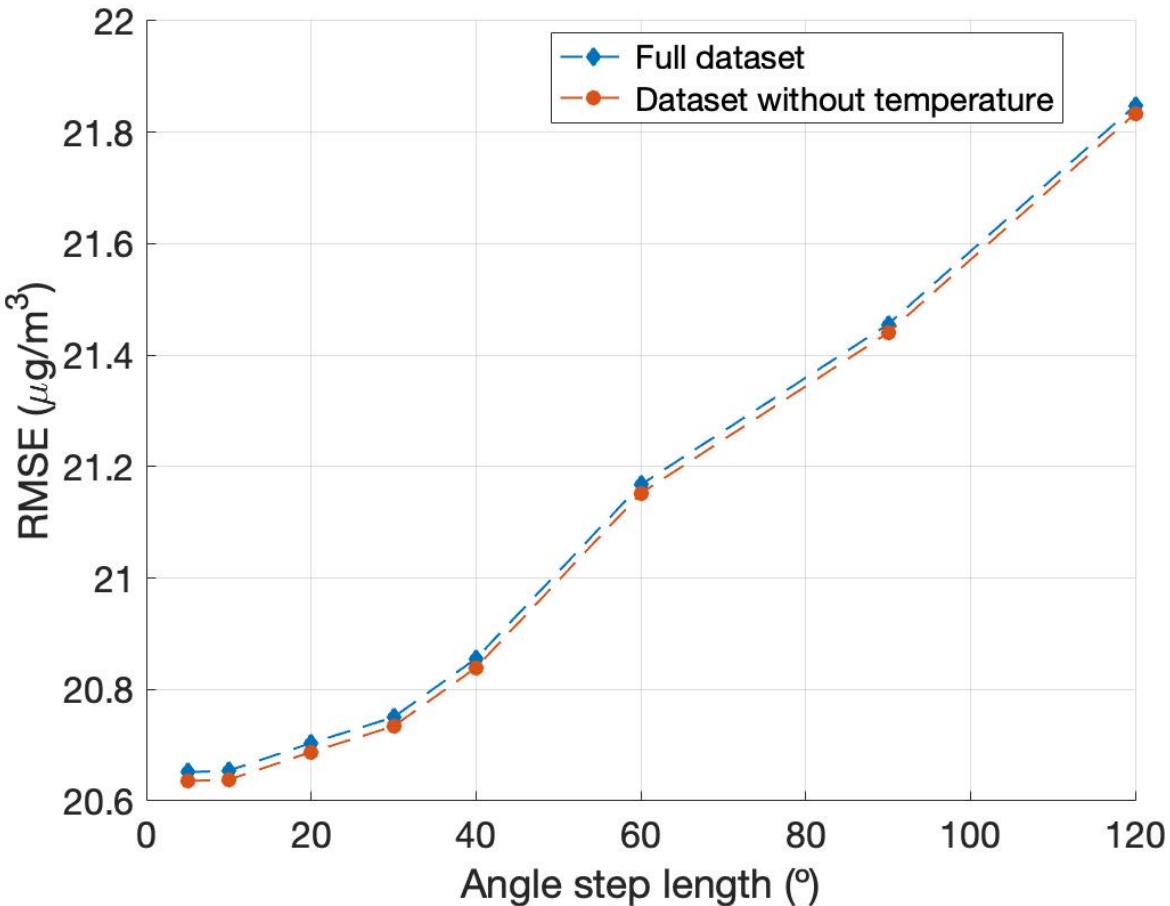
Indice di strada



**RIDUZIONE DATASET**

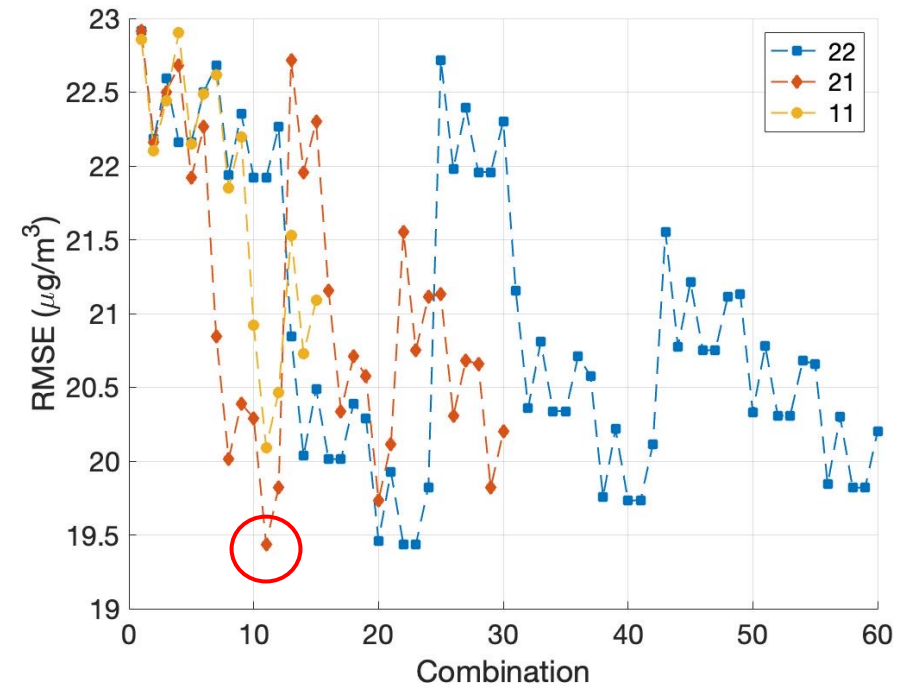
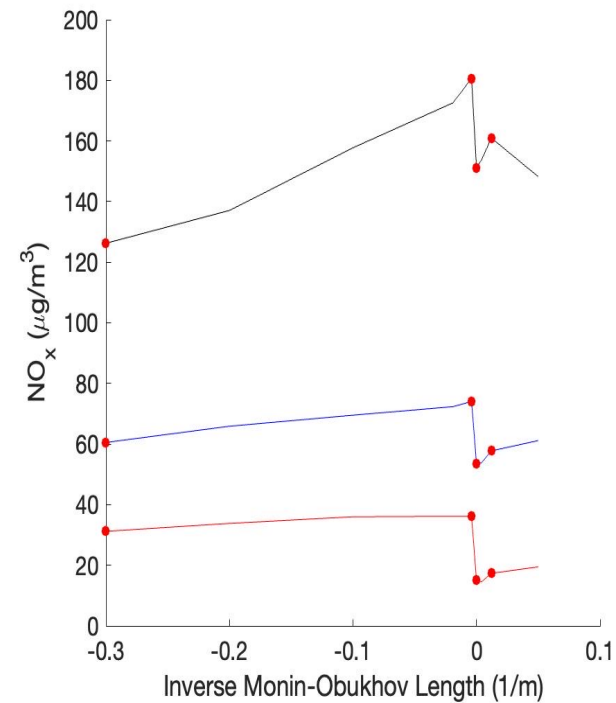
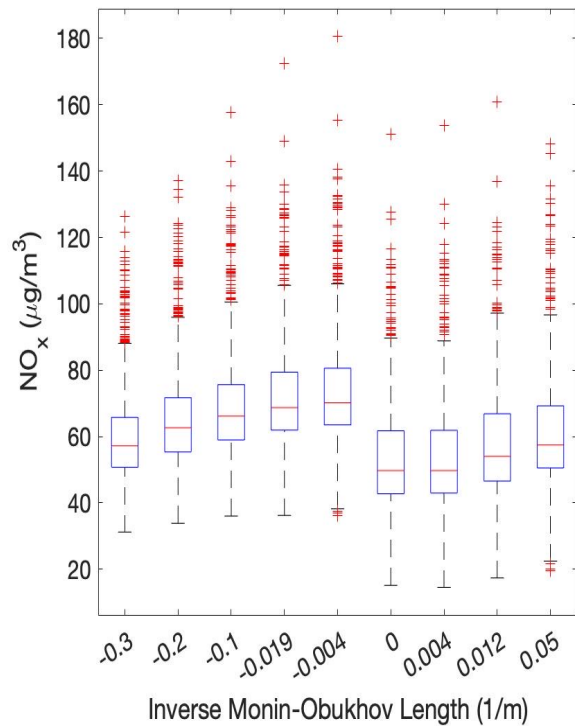
$$RMSE_{\Omega} = \sqrt{\frac{1}{|\Omega|} \sum_{t \in \Omega} (\bar{C}(t) - \hat{\bar{C}}(t))^2}$$

Errore quadratico medio nel dominio del tempo ( $\Omega$ ) sulla media spaziale ( $\bar{C}$ ).



RIDUZIONE  
DATASET

$$L_{MO}^{-1}: [-0.3, -0.2, -0.1, -0.004, -0.019, 0, 0.004, 0.012, 0.05] \text{ 1/m} \in I$$



## RIDUZIONE DATASET

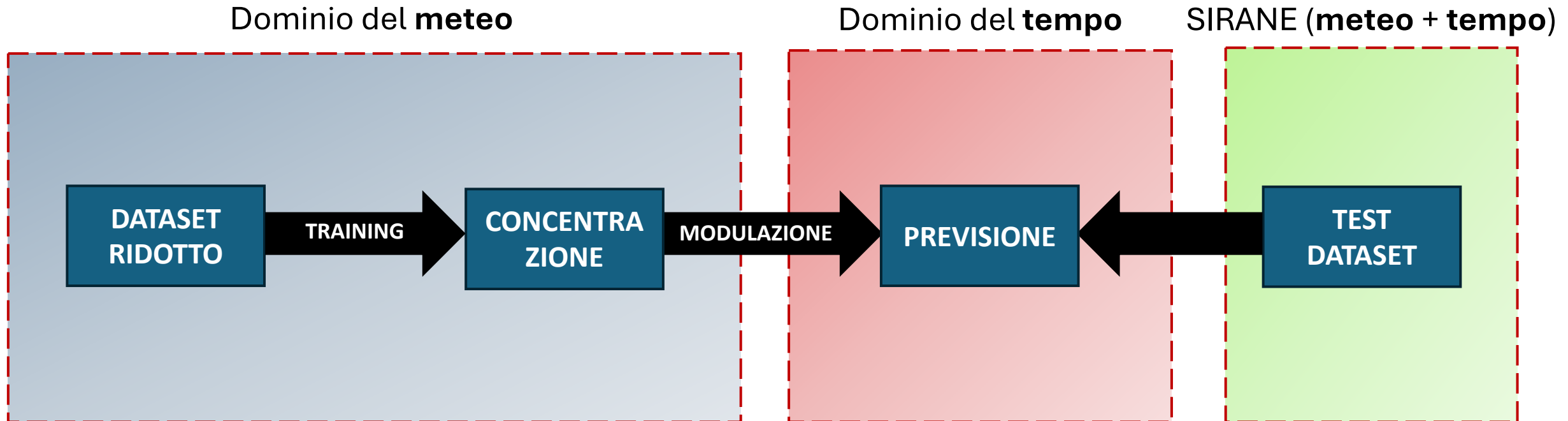
Riduzione finale:

- $L_{MO}^{-1}: [-0.3, -0.004, 0, 0.012] \ 1/m \subset I$
- $T$ : rimossa
- $Dir: [0: \mathbf{30}: 360]^\circ \subset \Theta$ . (**+0.48%** RMSE)
- $u_*: [0.05: \mathbf{0.05}: 1.2] \ m/s \in \Phi$

Risultati pratici:

- Da 189216 a **1248** combinazioni
- **0.66%** dataset iniziale
- Da 700+ GB a **4.6+ GB**
- **14.25%** test dataset

## Work flow – Training e testing



## Modello di regressione lineare

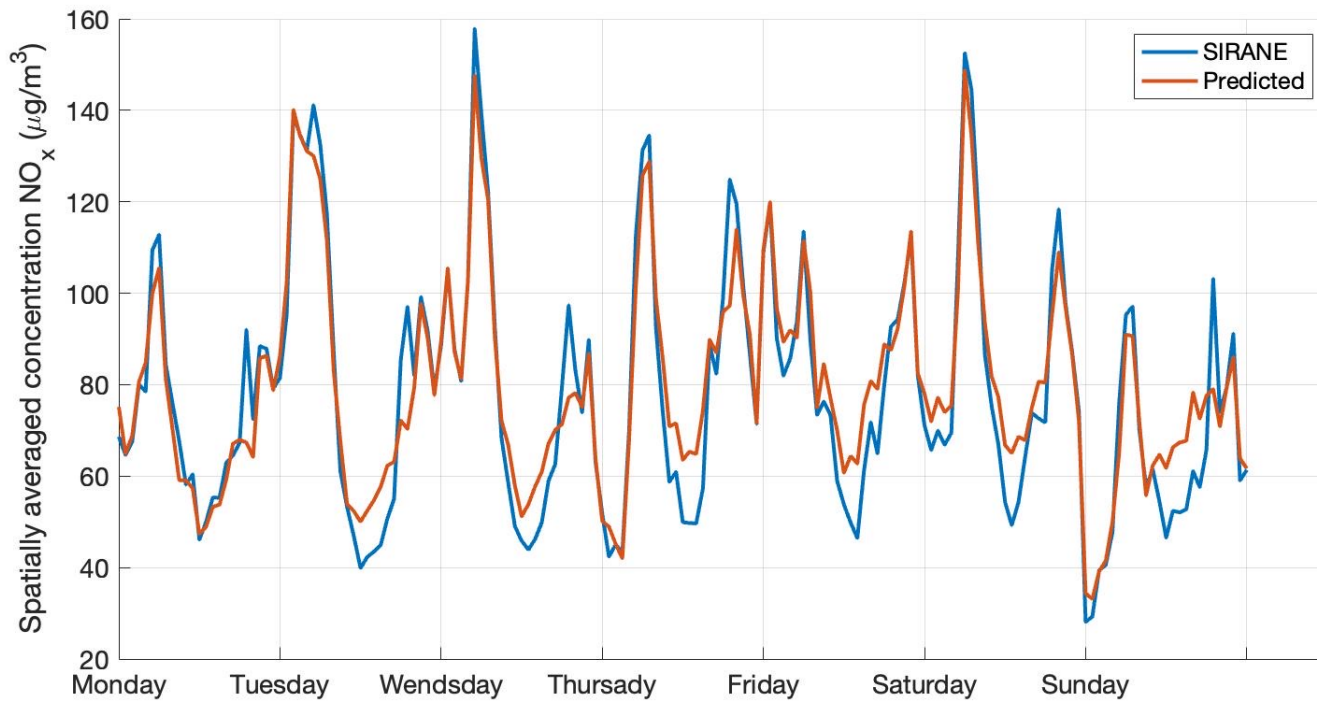
**Model 1 (NO\_Road\_ID):**  $NO_x = \beta_0 + \beta_1 \sin(\text{Dir}) + \beta_2 \cos(\text{Dir}) + \beta_3 u_* + \beta_4 L_{MO}^{-1}$

**Model 2 (Road\_ID):**  $NO_x = \beta_0 + \beta_1 \sin(\text{Dir}) + \beta_2 \cos(\text{Dir}) + \beta_3 u_* + \beta_4 L_{MO}^{-1} + \sum_{i=1}^{585} \gamma_i \cdot \text{Road\_ID}_i$

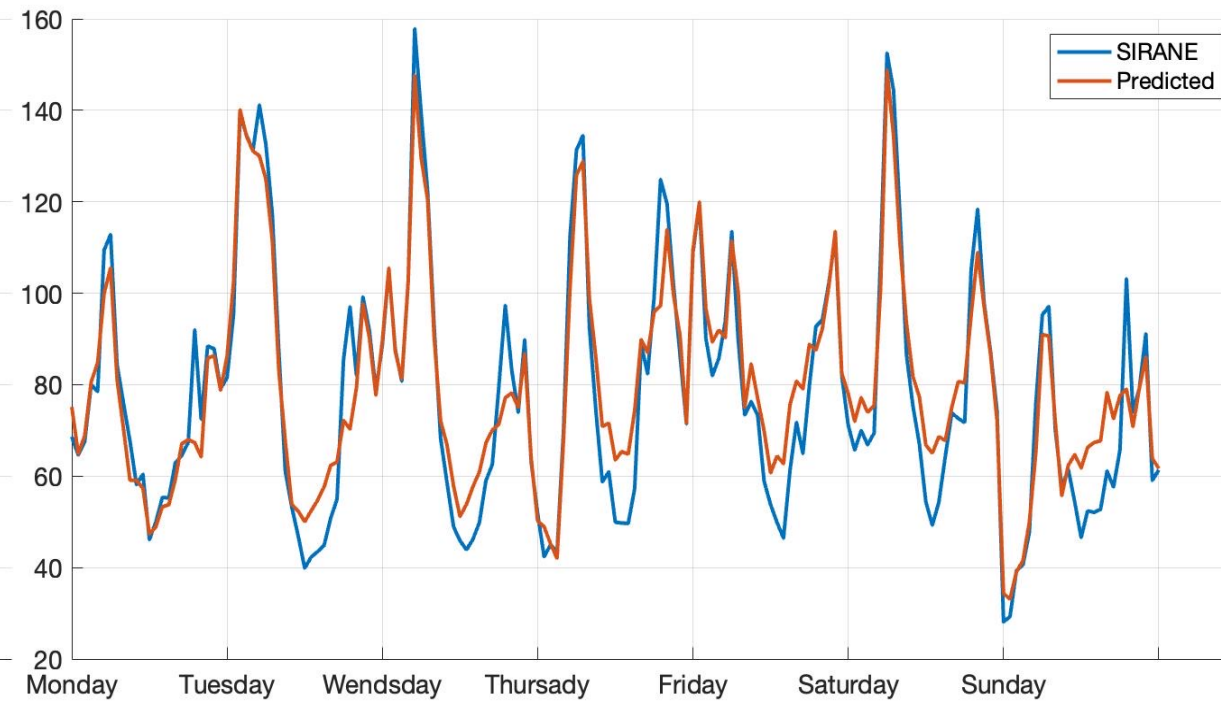
Metric	Model 1 (No Road_ID)	Model 2 (Road_ID)
MAE	18.26 $\mu\text{g}/\text{m}^3$	16.16 $\mu\text{g}/\text{m}^3$
RMSE	29.38 $\mu\text{g}/\text{m}^3$	26.45 $\mu\text{g}/\text{m}^3$
NMSE	0.08 $\mu\text{g}/\text{m}^3$	0.05 $\mu\text{g}/\text{m}^3$
MAPE	16.60%	13.97%
R <sup>2</sup>	0.85	0.88
$\rho$	0.94	0.96

# Modello di regressione lineare

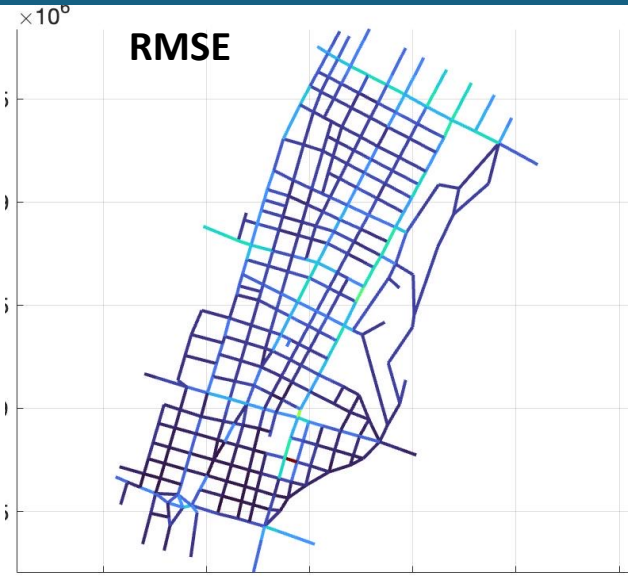
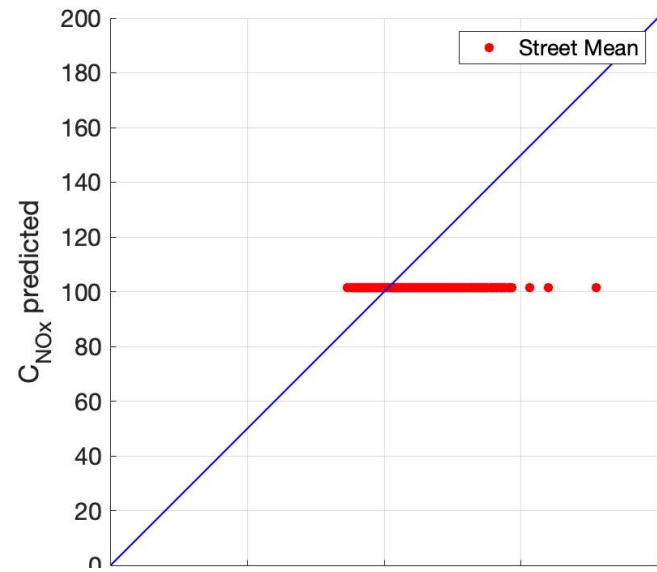
## Model 1



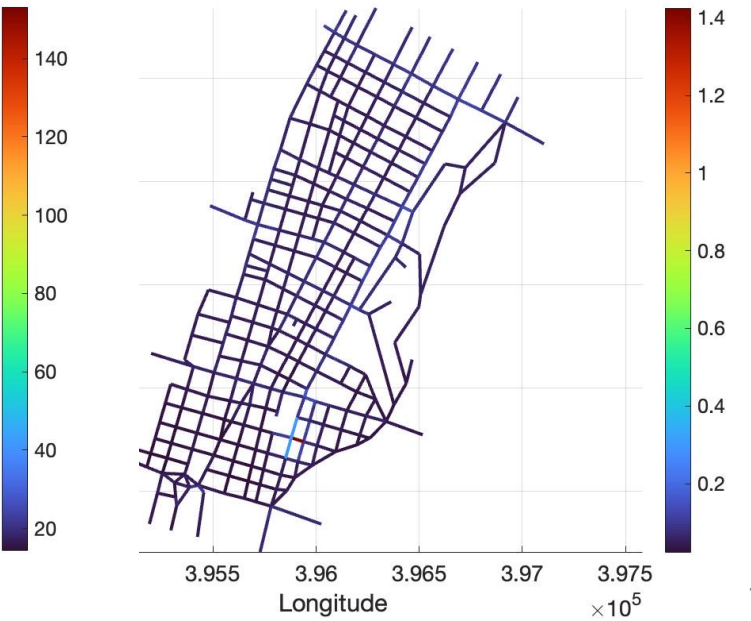
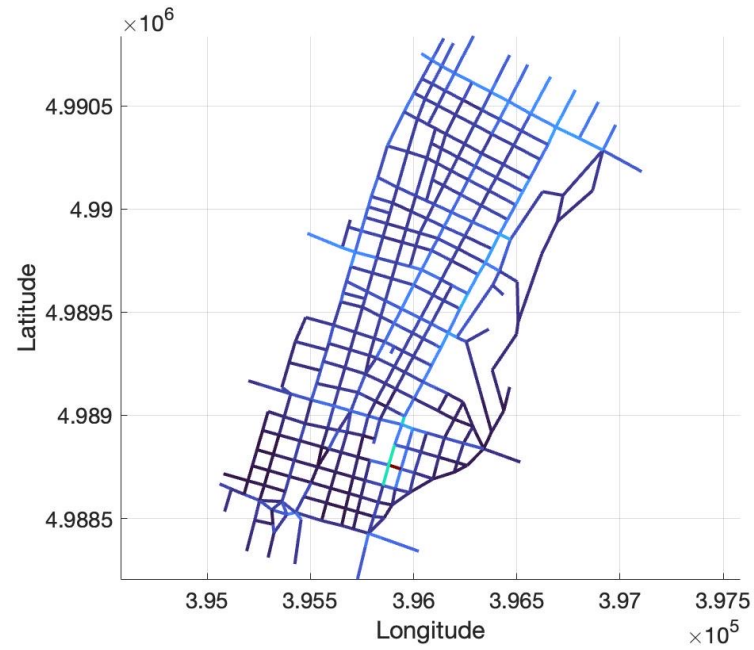
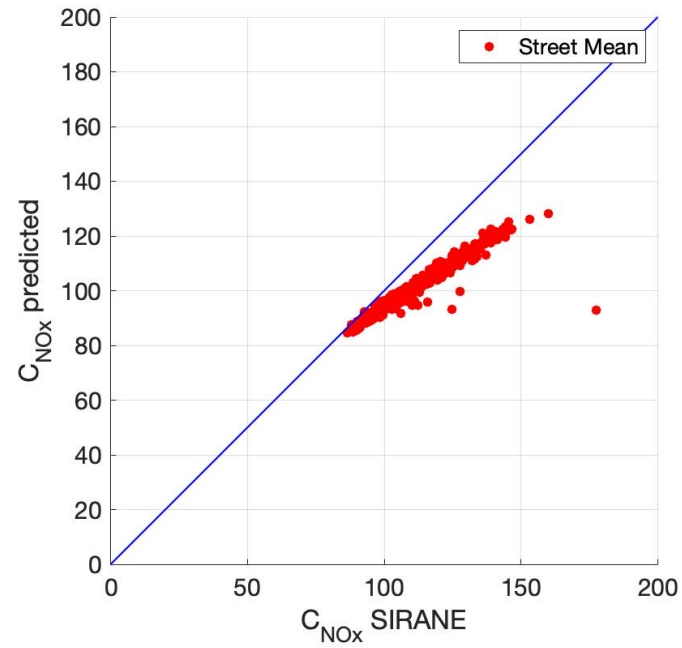
## Model 2



### Model 1

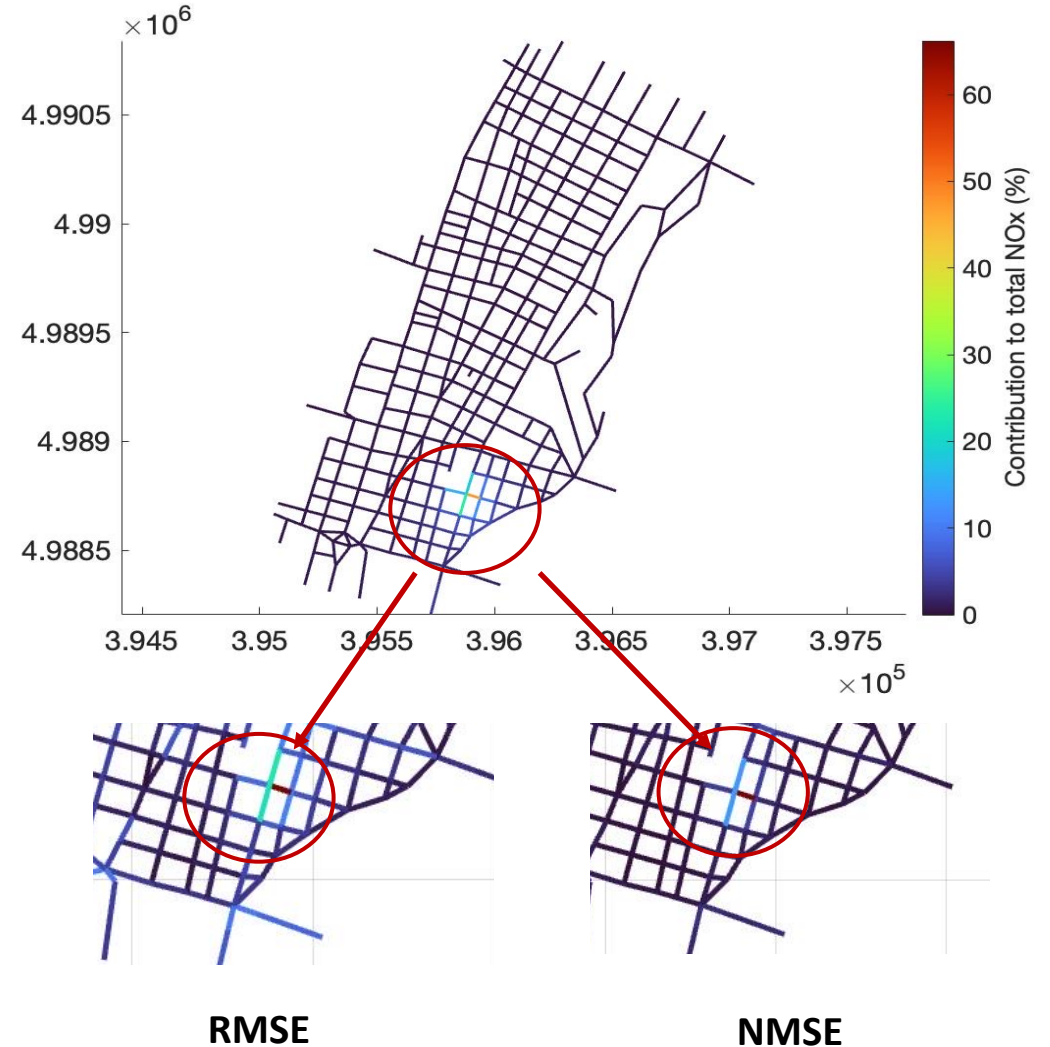


### Model 2



## Considerazioni finali

1. **Sottostima** dell'errore dovuta **sorgente puntuale** (non considerata costruzione database)
2. Entrambi **modelli regressivi** si **correlano** molto bene (effetto modulazione da parte dell'utente)
3. Modello 1 senza Road\_ID ha pressoché la **stessa performance** per la **media spaziale** ma manca accuratezza previsioni locali (singole strade)
4. Informazione sulla **geometria** codificata nel **database stesso**



## Conclusioni e sviluppi futuri

1. L'**Interpolazione** modello più preciso ma meno portabile
2. Nessuno dei modelli sviluppati è realmente **general purpose**:
  - La **dipendenza temporale** imposta dall'utente
  - Il modello è **overfittato** sul **dominio** della simulazione e non in grado di predire accuratamente su geometrie urbane arbitrarie
3. Il lavoro costituisce una base preliminare:
  - Espansione a **domini** urbani più **ampi**
  - Introduzione di **predittori strutturali e temporali**
  - **Interazione** tra predittori
  - **Modelli machine learning** più avanzati (CNN, GNN)