

UNIVERSITÀ DEGLI STUDI DI FIRENZE
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
TESI DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

Progettazione e sviluppo di componenti per la piattaforma
AirQino dedicata al monitoraggio della qualità dell'aria

Candidato
Edoardo D'Angelis



 Consiglio Nazionale delle Ricerche
Istituto per la BioEconomia
Dipartimento di Scienze Bio-Agroalimentari

Relatori
Prof. Andrew D. Bagdanov
Prof. Pietro Pala

Correlatori
Dott. Walter Nunziati
Dott.ssa Alice Cavaliere

Anno Accademico 2020/2021

Contesto

L'**inquinamento atmosferico** è uno dei principali problemi che interessano le aree urbanizzate.

- ▶ Può portare a **problemi di salute** causati dall'esposizione a lungo termine a sostanze nocive (PM, NO₂, CO₂, O₃, etc.)
- ▶ Il **monitoraggio** è essenziale per la tutela della salute pubblica
 1. Con reti regionali di rilevamento fisse, gestite da ARPA (DLgs. n.155 del 13/08/2010)
 2. Con nuove reti di sensori *low cost* ad alta portabilità per l'acquisizione di misure aggiuntive, anche a minor precisione (es. **AirQino**)

La piattaforma AirQino (1/3)

- ▶ Monitoraggio ambientale ad alta precisione
- ▶ Configurabile ed estendibile
- ▶ Dati in tempo reale (PM, NO₂, CO₂, O₃, etc.)



Figura: Una centralina AirQino

<https://airqino.magenta-lab.it>

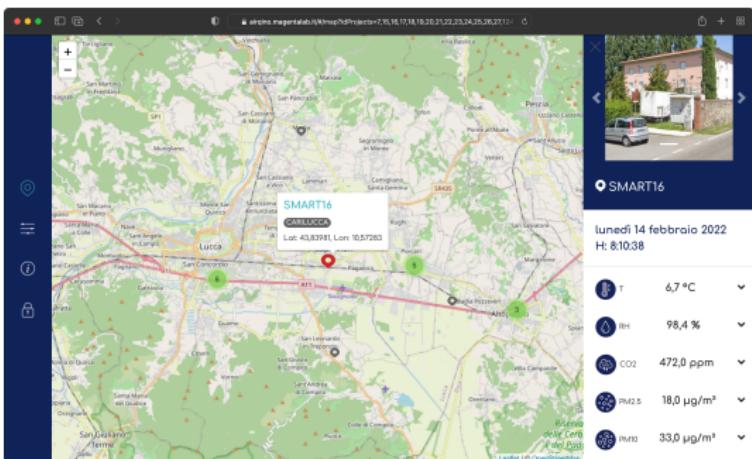


Figura: Pagina web AirQino

La piattaforma AirQino – Architettura (2/3)

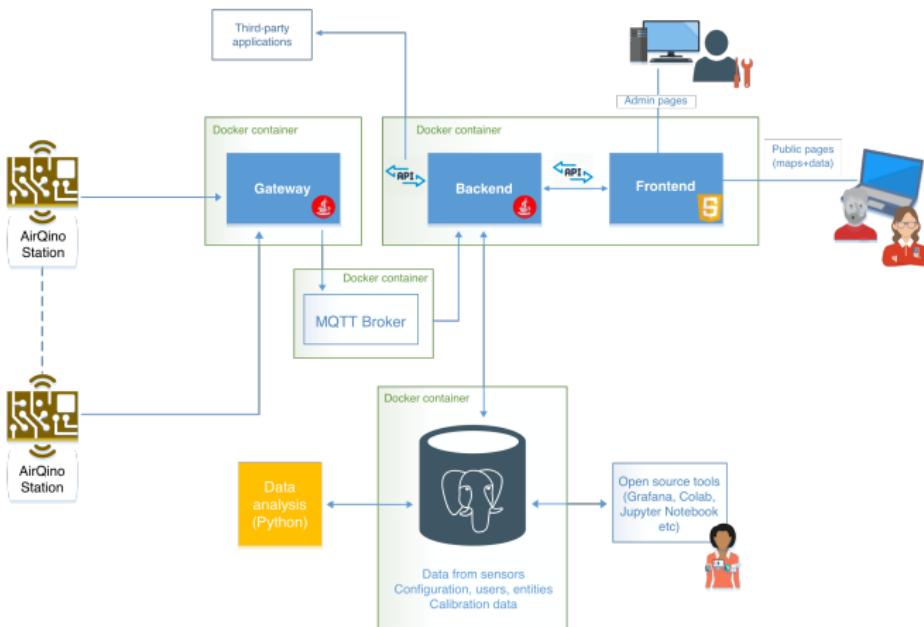
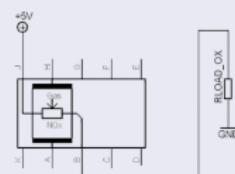


Figura: Architettura della piattaforma

La piattaforma AirQino – Sensori (3/3)

MiCS-2714 per NO₂

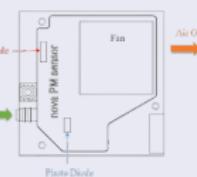


(a) Sensore

(b) Circuito

- ▶ Di tipo MOS, basato su *ossidoriduzione*
- ▶ Uscita in *counts*
- ▶ Costo: meno di 5€

SDS011 per PM_{2.5} e PM₁₀



(c) Sensore

(d) Componenti

- ▶ Basato su *scattering*
- ▶ Uscita in $\mu\text{g}/\text{m}^3$
- ▶ Costo: 25€ circa

Obiettivi

- ▶ Sviluppi tecnologici alla piattaforma **AirQino**
 1. Miglioramento dell'**affidabilità** dei dati provenienti dai sensori
 2. Riduzione dei **tempi di risposta** dal database
- ▶ Studio e confronto tra diverse tecniche volte a migliorare l'accuratezza del processo di **calibrazione** dei sensori (sia NO₂ che PM)
- ▶ Sviluppo di un'**interfaccia web** per facilitare la calibrazione *massiva* di centraline

Sviluppi – Replica del database (1/4)

Replica

Tutti i dati presenti nel database vengono copiati e distribuiti su un altro spazio fisico. La nuova istanza agisce come nodo secondario, alleggerendo il carico dal database primario.

Vantaggi:

- ▶ Maggiore **affidabilità**
- ▶ Miglioramento delle **prestazioni**
- ▶ Maggiore **sicurezza** dei dati

Sviluppi – Replica del database (2/4)

Streaming replication

Funzionalità che consente di replicare i dati in tempo reale da una istanza di database Postgres a un'altra

- ▶ Replica di sola lettura
- ▶ Basata su **WAL** (*Write Ahead Log*)
- ▶ Automazione con Docker

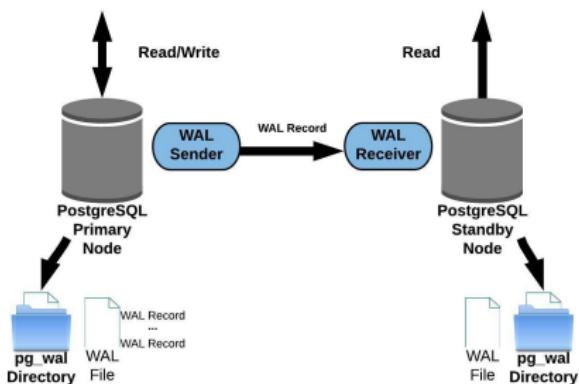


Figura: Streaming replication

Sviluppi – Ottimizzazione di query temporali (3/4)

Continuous aggregates

Funzionalità di **Timescale** per aggregare dati in tempo reale in maniera incrementale

- ▶ Miglioramento delle **performance**
- ▶ Aggiornamento **automatico** in background
- ▶ Risparmio di **spazio**

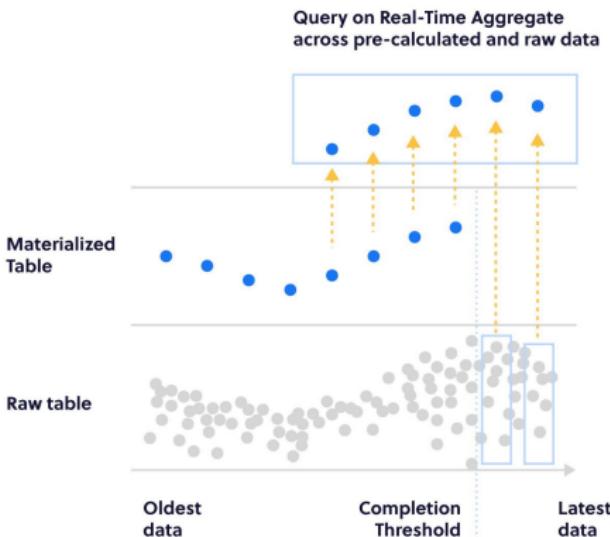
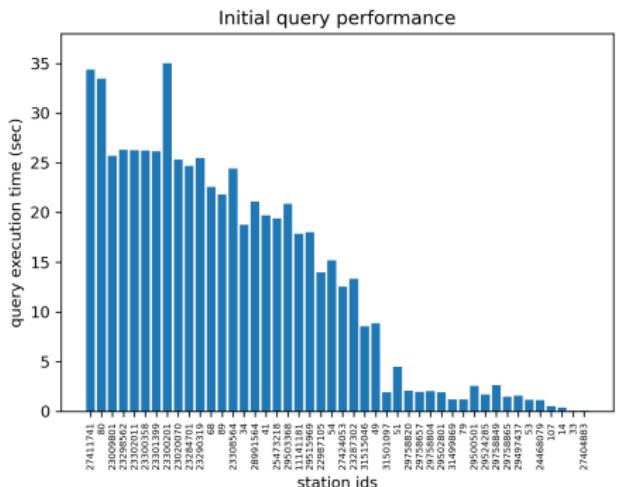


Figura: Continuous aggregates

Sviluppi – Ottimizzazione di query temporali (4/4)

Tempi di risposta della query per estrarre la media oraria di NO₂ dell'ultima settimana da tutte le centraline **AirQino**:



Calibrazione – Centraline (1/5)

- ▶ SMART16 (AirQino), in co-locazione con la stazione ARPAT di Capannori (Lucca)
 - ▶ Parte del progetto Carilucca
 - ▶ Sito di particolare interesse data la varietà di fonti di emissione^[1]
- ▶ Calibrazione: analisi della risposta dei sensori di NO₂ e PM sotto diverse concentrazioni di gas (test *indoor*) e comparazione verso sensori fissi di riferimento (test *outdoor*)



Figura: Posizione della centralina SMART16 (AirQino) e stazione ARPAT a Capannori (Lucca)

^[1] L. Brilli, F. Carotenuto, B. P. Andreini et al., "Low-Cost Air Quality Stations' Capability to Integrate Reference Stations in Particulate Matter Dynamics Assessment", *Atmosphere*, vol. 12, n. 8, 2021

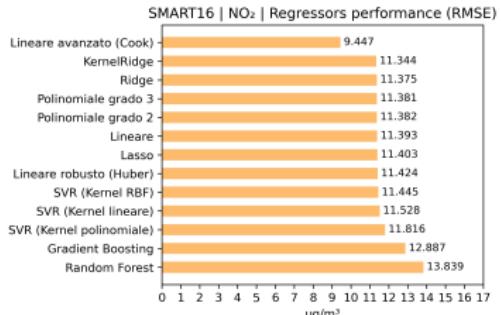
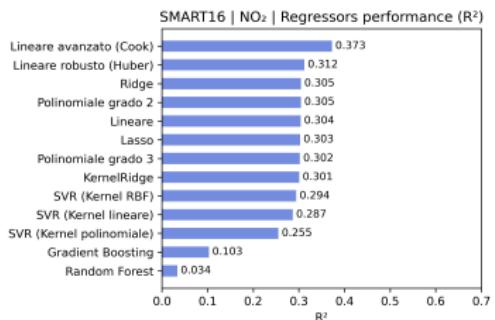
Calibrazione – Procedura (2/5)

Procedura per ogni sensore (NO_2 , $\text{PM}_{2.5}$, PM_{10})

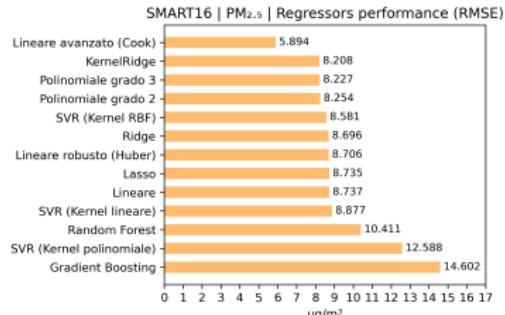
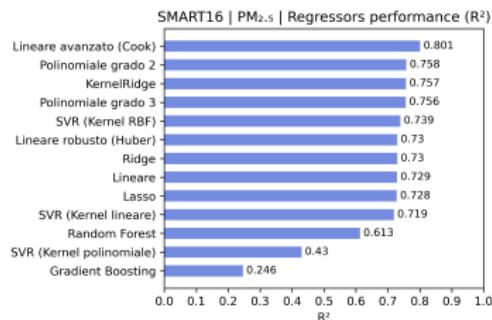
1. Preprocessamento e creazione di un **dataset** (dati della centralina AirQino SMART16 e di riferimento ARPAT)
2. **Allineamento** temporale e ricampionamento
3. **Scatterplot** del segnale di riferimento e del segnale del sensore
4. Analisi dei **residui**
5. Applicazione di tredici diversi modelli di **regressione**
 - ▶ Su tutto il dataset
 - ▶ Con cadenza mensile
6. Valutazione della **performance** in termini di R^2 e $RMSE$
7. **Validazione** del modello più performante

Calibrazione – Risultati (3/5)

Risultati NO₂ (R^2 , RMSE)



Risultati PM_{2.5} (R^2 , RMSE)



Calibrazione – Modello più performante (4/5)

Regessione lineare avanzata

Con rimozione dei valori anomali (*outlier*) in base a determinate metriche.

Distanza di Cook [2]:

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_{j(i)})^2}{p * MSE^2}$$

Soglia di cut-off [3]: $D_i > \frac{4}{n}$

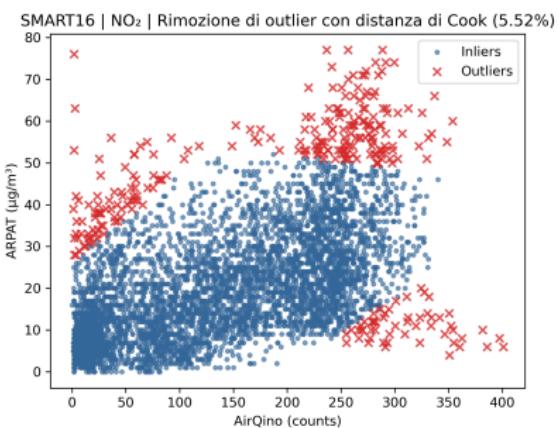


Figura: Rilevamento di *outlier* (in rosso) tramite distanza di Cook

[2] R. D. Cook, "Detection of Influential Observation in Linear Regression", *Technometrics*, 1977

[3] J. Fox, "Applied Regression Analysis and Generalized Linear Models", *Sage Publications*, 1997

Calibrazione – Validazione (5/5)

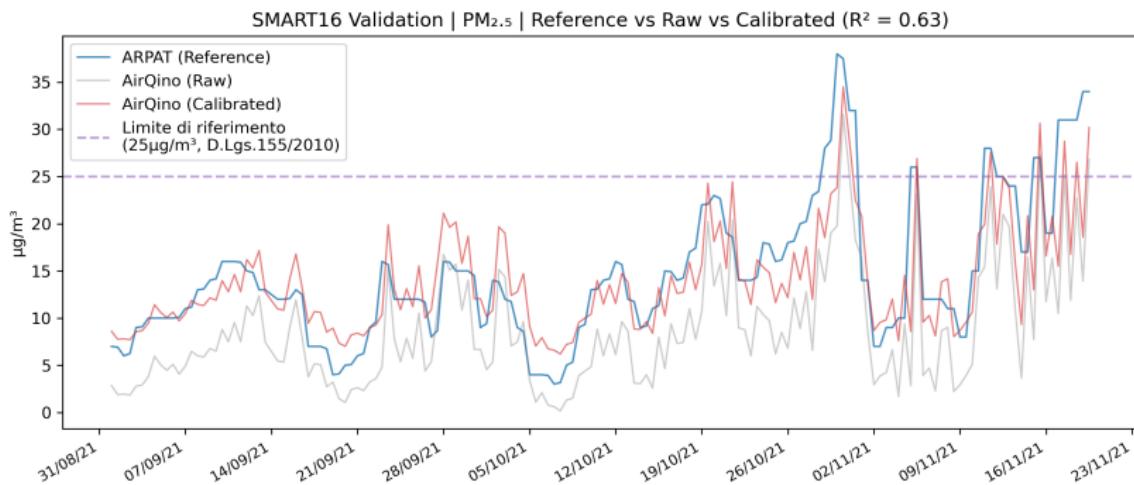


Figura: Confronto tra l'andamento temporale delle concentrazioni di PM_{2.5} misurate dalla stazione AirQino calibrata e non calibrata con riferimento alla stazione ARPAT di Capannori (LU). Medie a otto ore. Periodo dal 01/09/2021 al 20/11/2021.

Interfaccia utente web (1/2)

Soluzione per la calibrazione massiva di centraline **AirQino**:

- ▶ Per permettere di inserire più **coefficienti** contemporaneamente
- ▶ Basata su caricamento di file **csv** appositi
- ▶ Con riepilogo degli esiti in forma di tabella
 - ▶ Statistiche sui coefficienti medi

The screenshot shows a web browser window titled "Airqino Calibration Dashboard". At the top, there is a file input field labeled "Choose File" with the path "test_Formule_uia_caproni_test.csv" and a note about the header format: "Header format: id_airqino,param,a,b,c,q,score,[d],[formula],[regression_type]". Below this is a checkbox "Open a new session for every station" and a green "Uploaded!" button. The main area is titled "Results" with the subtext "0 rows created, 93 rows updated, 0 rows failed". A table follows, with columns: #, Result, Station, Sensor, R, k1, k2, k3, k4, and Last SQL Query. The table contains 7 rows of data, each showing an update query for a specific sensor (smart42, smart111) across different stations (O3, NO2, CO) with various regression parameters.

#	Result	Station	Sensor	R	k1	k2	k3	k4	Last SQL Query
1	Updated	smart42	O3	0.98	-0.1936			283	UPDATE field_calibration SET k1=-0.1936,
2	Updated	smart42	NO2	0.55	0.04068			-7.75	UPDATE field_calibration SET k1=0.04068,
3	Updated	smart42	CO	0.39	0.00952	-0.00001633		-1.321	UPDATE field_calibration SET k1=0.00952,
4	Updated	smart111	O3	0.29	-0.1342		3.451	-6.56	UPDATE field_calibration SET k1=-0.1342,
5	Updated	smart111	NO2	0.19	0.1029		-0.7563	15.47	UPDATE field_calibration SET k1=0.1029,
6	Updated	smart111	CO	0.1	0.0002663		-0.003038	0.1405	UPDATE field_calibration SET k1=0.0002663,
7	Updated	smart41	O3	0.73	0.2724		0.24	81.2	UPDATE field_calibration SET k1=0.2724,

Figura: Pagina web calibrazione

Interfaccia utente web (2/2)

- ▶ Distinzione visiva per esiti diversi di ciascun sensore
- ▶ Con un meccanismo di **autenticazione** per proteggere da utenti di terze parti (*Keycloak*)
- ▶ Processo di rilascio automatizzato (*CI/CD*) con isolamento delle dipendenze (*Docker*)

The screenshot shows a web browser window titled "Airqino Calibration Dashboard". At the top, there is a form for uploading a CSV file, with a file named "test_Formula_usa_caproni_test.csv" selected. Below the file input, there is a header format definition: "Header format: id_airqino,param,a,b,c,q,score,[d],[formula],[regression_type]". There is also a checkbox for "Open a new session for every station". A green button labeled "Uploaded!" is visible. Below the form, the word "Results" is followed by "1 rows created, 1 rows updated, 1 rows failed". A table displays the following data:

#	Result	Station	Sensor	R	k1	k2	k3	k4	d	Last SQL Query
1	Updated	smart41	NO2	0.41	0.1266	0.1323	0.6323	-1.944	-0.3245	UPDATE field_calibration SET R = 0.41, k1 = 0.1266, k2 = 0.1323, k3 = 0.6323, k4 = -1.944 WHERE station_id = 41 AND sensor_id = 2
2	Failed: station SMART999 does not exist	smart999	NO2	0.73	0.1596	0.7429	0.1243	-1.098	-0.7531	SELECT id FROM stations WHERE name = 'SMART999'
3	Created	smart104	CO2	0.61	0.1366	0.1546	0.2435	-0.3245	-0.8754	INSERT INTO field_calibration (station_id, sensor_id, R, k1, k2, k3, k4) VALUES (104, 2, 0.61, 0.1366, 0.1546, 0.2435, -0.3245)

Figura: Calibrazione con esiti diversi

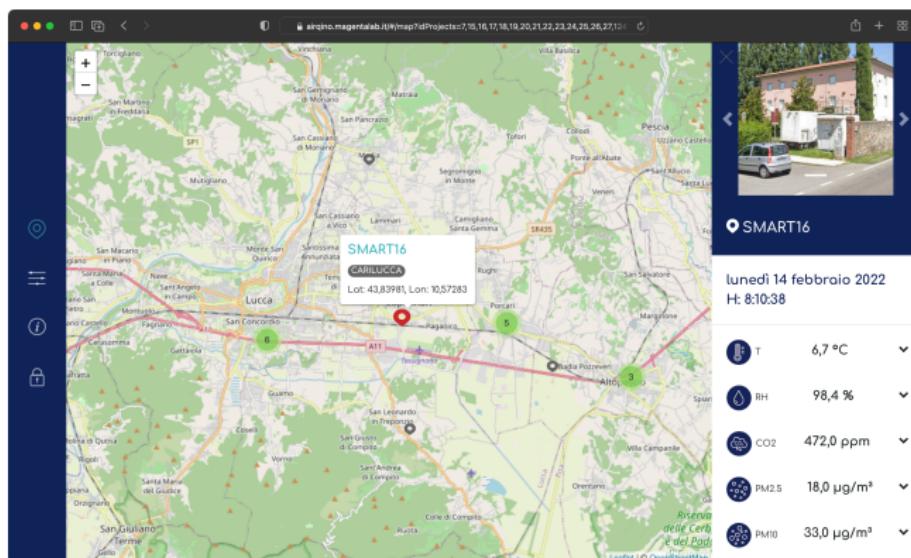
Conclusioni

- ▶ Reti di sensori *low-cost* per il monitoraggio della qualità dell'aria (es. **AirQino**) rappresentano una soluzione efficace per il rilevamento dell'inquinamento atmosferico
 - ▶ Anche in ottica di **integrazione** con le reti di monitoraggio regionali già esistenti, fornendo un quadro più completo della qualità dell'aria in ambiente urbano
- ▶ Anche con sensori a basso costo è possibile misurare inquinanti come NO₂, PM_{2.5} e PM₁₀
 - ▶ Con accuratezza ancora migliore se si ha a disposizione un **segnale di riferimento** con cui correggere i dati provenienti dai sensori
 - ▶ L'applicazione di tecniche di **regressione robusta** in fase di calibrazione ha riportato miglioramenti significativi

Sviluppi futuri

- ▶ Perfezionare il processo di calibrazione delle centraline, ad esempio implementando regressioni *multivariate* (includendo fattori metereologici quali **temperatura** e **umidità** relativa)
 - ▶ La **temperatura** tende a correlare negativamente con i PM (all'aumentare della temperatura le polveri sottili tendono a calare)
 - ▶ L'**umidità** presenta l'effetto contrario e influisce anche sull'elettronica del sensore
- ▶ Realizzare una procedura di **anomaly detection** sul segnale dei sensori
 - ▶ Con un controllo **in tempo reale** sull'andamento e monitorando la differenza con il segnale di riferimento (es. stazione ARPAT)
 - ▶ **Post processing analysis**

Grazie per l'attenzione!



Consiglio Nazionale delle Ricerche
Istituto per la BioEconomia
Dipartimento di Scienze Bio Agroalimentari