



UNIVERSITÀ DEGLI STUDI DI PERUGIA

DIPARTIMENTO DI INGEGNERIA

INGEGNERIA INFORMATICA E ROBOTICA

ESAME DI SIGNAL PROCESSING AND OPTIMIZATION FOR
BIG-DATA

**Analisi Spazio-Temporale della Qualità dell'Aria
di Pechino tramite Graph Signal Processing**

Luca Fagiolo

GITHUB: [LINK](#)

DATASET: [LINK](#)

Indice

1	Introduzione	3
1.1	Contesto e Motivazione	3
1.2	Obiettivi del Progetto	4
1.3	Dataset e Area di Studio	4
2	Background Teorico	6
2.1	Teoria dei Grafi per il Signal Processing	6
2.2	Il Laplaciano di un Grafo	7
2.3	Graph Fourier Transform (GFT)	8
2.3.1	Definizione e Proprietà	8
2.3.2	Interpretazione Frequenziale	8
3	Metodologia e Implementazione	9
3.1	Architettura del Sistema	9
3.2	Struttura del Codice MATLAB	10
3.2.1	Caricamento Dati	10
3.2.2	Gestione dei Valori Mancanti	11
3.2.3	Costruzione del Grafo della Rete di Stazioni	12
3.2.4	Calcolo del Laplaciano e dei Coefficienti della GFT	13
3.2.5	Analisi della Distribuzione Spettrale di Potenza	13
3.2.6	Visualizzazione	14
4	Risultati Sperimentali	15
4.1	Selezione del Parametro k per il Grafo k-NN	15
4.2	Analisi Multi-Variabile dell’Inquinamento Atmosferico	18
4.2.1	Particolato Fine (PM2.5)	19
4.2.2	Biossido di Azoto (NO ₂)	20
4.2.3	Ozono (O ₃)	21
4.2.4	Temperatura	22
4.2.5	Precipitazioni	23
5	Conclusioni	25

1. Introduzione

Questo progetto si concentra sull’analisi della qualità dell’aria a Pechino utilizzando tecniche di Graph Signal Processing. Il dataset comprende misurazioni orarie di inquinanti atmosferici e variabili meteorologiche raccolte da 12 stazioni distribuite nell’area metropolitana della capitale cinese, coprendo un periodo di quattro anni dal 2013 al 2017. L’idea centrale è modellare la rete di stazioni come un grafo, dove ogni stazione rappresenta un nodo e le connessioni riflettono la vicinanza geografica. Applicando la decomposizione spettrale del Laplaciano e la Graph Fourier Transform, diventa possibile identificare pattern spaziali ricorrenti e analizzare come le dinamiche di inquinamento evolvano nel tempo. I risultati mostrano che questo approccio riesce a catturare pattern giornalieri, settimanali e stagionali, evidenziando correlazioni spazio-temporali che metodi tradizionali potrebbero non rilevare.

1.1 Contesto e Motivazione

L’inquinamento atmosferico urbano rappresenta uno dei problemi ambientali più significativi a livello globale, con impatti diretti sulla salute pubblica, sull’economia e sulla qualità della vita nelle aree metropolitane. La complessità di questo fenomeno deriva dalla molteplicità delle sorgenti emissive (traffico veicolare, attività industriali, riscaldamento domestico, costruzioni) e dalla loro distribuzione eterogenea sul territorio. Questa eterogeneità spaziale, combinata con le variazioni temporali legate ai cicli giornalieri delle attività umane e alle condizioni meteorologiche stagionali, genera pattern di inquinamento caratterizzati da forte variabilità spazio-temporale. Per questo motivo le reti multi-stazione sono diventate lo standard per monitorare la qualità dell’aria urbana. Tuttavia, interpretare i dati provenienti da sensori multipli presenta sfide metodologiche non banali. I metodi tradizionali tendono a trattare le misurazioni come osservazioni indipendenti, perdendo informazioni preziose sulle relazioni spaziali e sui meccanismi di trasporto degli inquinanti. In questo contesto i grafi offrono una soluzione naturale ed efficace. Le stazioni diventano nodi collegati secondo criteri di vicinanza o similitudine, permettendo di sfruttare la struttura topologica della rete per estrarre informazioni altrimenti nascoste.

1.2 Obiettivi del Progetto

L’obiettivo principale di questo progetto è l’analisi spazio-temporale dell’inquinamento atmosferico urbano attraverso l’applicazione di tecniche di Graph Signal Processing. Questo obiettivo viene raggiunto sfruttando la struttura topologica della rete di stazioni per identificare pattern non evidenti con metodi tradizionali.

Il progetto si articola in quattro fasi operative:

1. **Modellazione della rete di stazioni:** costruzione di un grafo mediante algoritmo k-Nearest Neighbors (k-NN) applicato alle coordinate geografiche. La scelta del parametro k viene valutata confrontando le proprietà spettrali dei grafi risultanti, bilanciando connettività locale e archi poco significativi.
2. **Analisi spettrale del grafo:** decomposizione del Laplaciano per identificare i modi spaziali fondamentali. I modi a bassa frequenza catturano correlazioni regionali e fenomeni diffusi, mentre quelli ad alta frequenza rivelano variazioni locali e discontinuità spaziali.
3. **Trasformata di Fourier su grafo (GFT):** proiezione dei segnali di inquinamento sulla base degli autovettori per quantificare il contributo energetico di ciascun modo spaziale.
4. **Visualizzazione e interpretazione:** sviluppo di strumenti grafici per rappresentare le dinamiche spazio-temporali identificate, inclusi spettrogrammi, mappe di correlazione e visualizzazioni geografiche dei modi autovettoriali.

1.3 Dataset e Area di Studio

Il dataset utilizzato è il *Beijing Multi-Site Air-Quality Data Set*, reso disponibile tramite Kaggle. Il dataset ha natura multivariata e include sia parametri ambientali che variabili meteorologiche. In particolare, vengono monitorati sei principali inquinanti atmosferici: PM2.5 (particolato fine), PM10 (particolato grosso), SO₂ (biossalido di zolfo), NO₂ (biossalido di azoto), CO (monossalido di carbonio) e O₃ (ozono). A questi si aggiungono sei grandezze meteorologiche: temperatura, pressione atmosferica, punto di rugiada, precipitazioni, direzione del vento e velocità del vento.

Dal punto di vista temporale, le osservazioni coprono un arco che va dal 1° marzo 2013 al 28 febbraio 2017, con una risoluzione oraria. Questa caratteristica rende il dataset adatto all’analisi di fenomeni su diverse scale temporali: giornaliere, settimanali, stagionali e annuali. La disponibilità di una serie temporale così estesa e dettagliata permette inoltre di individuare pattern ricorrenti e variazioni legate a specifici periodi dell’anno. Va tuttavia notato che in alcune stazioni si riscontrano valori mancanti che rendono necessario un adeguato preprocessing dei dati prima di procedere con l’analisi. La Tabella 1.1 riporta le statistiche di completezza.

Strutturalmente, il dataset è organizzato in 12 tabelle (una per stazione), ciascuna con 35.064 osservazioni e 17 variabili. Ogni riga corrisponde alle misurazioni registrate da una stazione in un dato intervallo orario. Le variabili includono gli identificatori temporali (Year, month, day, hour), i principali inquinanti atmosferici (PM2.5, PM10, SO₂, NO₂, CO, O₃), le grandezze meteorologiche (TEMP, PRES, DEWP, RAIN, wd, WSPM) e l'identificativo della stazione di misura (station).

Tabella 1.1: Statistiche di completezza dei dati per variabile

Variabile	Totale	Mancanti	Dati Mancanti (%)
PM2.5	420768	8739	2.1
PM10	420768	6449	1.5
SO ₂	420768	9055	2.2
NO ₂	420768	12903	3.1
CO	420768	20701	4.9
O ₃	420768	14241	3.4
TEMP	420768	998	0.2
PRES	420768	665	0.2
DEWP	420768	403	0.1
RAIN	420768	390	0.1
WSPM	420768	318	0.1

2. Background Teorico

Graph Signal Processing nasce dall'idea di estendere gli strumenti classici dell'elaborazione dei segnali ai domini non regolari, come reti e grafi. Mentre nel contesto tradizionale i segnali sono definiti su strutture regolari (ad esempio sequenze temporali o immagini su una griglia), nella GSP i segnali sono campionati sui nodi di un grafo le cui connessioni descrivono relazioni di prossimità o di similarità. Questo approccio consente di modellare fenomeni complessi come i dati ambientali raccolti da una rete di stazioni di monitoraggio.

2.1 Teoria dei Grafi per il Signal Processing

Un **grafo** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ è definito da:

- $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$: insieme di N nodi
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$: insieme di archi

I nodi rappresentano le entità di interesse (in questo caso, le stazioni di monitoraggio), mentre gli archi modellano le relazioni di vicinanza o correlazione.

La **matrice di adiacenza** $\mathbf{A} \in \mathbb{R}^{N \times N}$ è definita come:

$$\mathbf{A}_{i,j} = \begin{cases} 1 & \text{se } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{altrimenti} \end{cases} \quad (2.1)$$

Essa rappresenta in forma matriciale la struttura del grafo: un valore pari a 1 in posizione (i, j) indica la presenza di un arco tra i nodi v_i e v_j . Nel caso di un grafo non diretto la matrice di adiacenza è simmetrica, in quanto l'arco tra i nodi (i, j) è ovviamente equivalente all'arco tra i nodi (j, i) .

La **matrice diagonale dei gradi** $\mathbf{D} \in \mathbb{R}^{N \times N}$ è definita come:

$$\mathbf{D}_{i,i} = \sum_{j=1}^N \mathbf{A}_{i,j}$$

Ciascun elemento $\mathbf{D}_{i,i}$ della matrice rappresenta il grado del nodo v_i , ovvero la somma degli archi incidenti su di esso.

2.2 Il Laplaciano di un Grafo

Il Laplaciano è uno degli operatori fondamentali nello studio delle reti complesse e, in particolare, nel Graph Signal Processing. Esso fornisce un legame diretto tra la struttura topologica del grafo e le proprietà analitiche dei segnali definiti sui suoi nodi.

Dato un grafo non diretto $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, con matrice di adiacenza $\mathbf{A} \in \mathbb{R}^{N \times N}$ e matrice diagonale dei gradi $\mathbf{D} \in \mathbb{R}^{N \times N}$, allora il **Laplaciano** \mathbf{L} è definito come:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}. \quad (2.2)$$

Questo operatore possiede alcune proprietà fondamentali:

- \mathbf{L} è una matrice simmetrica e semidefinita positiva
- Il più piccolo autovalore di \mathbf{L} è sempre $\lambda_1 = 0$, con autovettore associato il vettore costante $\mathbf{1} = [1, 1, \dots, 1]^T$.
- Il numero di autovalori nulli di \mathbf{L} coincide con il numero di componenti connesse del grafo.
- La somma delle righe (e delle colonne) di \mathbf{L} è nulla, ossia $\mathbf{L}\mathbf{1} = \mathbf{0}$.

Dal punto di vista interpretativo, il Laplaciano cattura la nozione di variazione di un segnale sul grafo: due nodi collegati tendono ad avere valori simili se il segnale è "smooth". In questo senso, \mathbf{L} svolge un ruolo analogo all'operatore differenziale nel dominio continuo, rendendolo lo strumento centrale per definire concetti spettrali e trasformate nel contesto dei grafi.

Essendo \mathbf{L} una matrice simmetrica semidefinita positiva, esiste una decomposizione agli autovalori:

$$\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^T, \quad (2.3)$$

dove:

- $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \in \mathbb{R}^{N \times N}$ è una matrice ortogonale contenente gli autovettori normalizzati,
- $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ è la matrice diagonale degli autovalori, ordinati come $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$.

Dal punto di vista interpretativo, gli autovalori del Laplaciano possono essere visti come frequenze del grafo. Questa corrispondenza sta alla base della Graph Fourier Transform.

2.3 Graph Fourier Transform (GFT)

2.3.1 Definizione e Proprietà

La **GFT** estende la trasformata di Fourier classica a domini discreti e irregolari rappresentati da grafi. Essa utilizza come base ortonormale gli autovettori del Laplaciano del grafo, i quali riflettono la struttura topologica della rete.

Sia $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^T$ la decomposizione spettrale del Laplaciano, con \mathbf{U} matrice degli autovettori e Λ matrice diagonale degli autovalori. Per un segnale $\mathbf{x} \in \mathbb{R}^N$ definito sui nodi del grafo, la sua GFT è:

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}, \quad (2.4)$$

mentre la trasformata inversa è data da:

$$\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}. \quad (2.5)$$

I coefficienti $\hat{\mathbf{x}}$ rappresentano l'ampiezza delle componenti del segnale rispetto agli autovettori del Laplaciano, che svolgono il ruolo di “funzioni armoniche” del grafo.

2.3.2 Interpretazione Frequenziale

La nozione di frequenza nel dominio dei grafi è legata agli autovalori e agli autovettori del Laplaciano. In particolare, gli autovettori associati ad autovalori piccoli descrivono modalità a *bassa frequenza*, caratterizzate da variazioni graduali e globali nel grafo. Al contrario, gli autovettori corrispondenti ad autovalori elevati identificano modalità ad *alta frequenza*, che rappresentano oscillazioni rapide e localizzate tra nodi vicini.

Ad esempio, nel contesto di questo progetto, le componenti a bassa frequenza possono catturare tendenze persistenti e stagionali, condivise da più stazioni di monitoraggio, mentre le alte frequenze permettono di individuare eventi anomali o variazioni locali dovute a fonti puntuali di inquinamento. La GFT fornisce dunque un quadro teorico solido per distinguere tra dinamiche globali e fenomeni locali, rendendola uno strumento cruciale per l'interpretazione e la modellazione dei dati su reti complesse.

3. Metodologia e Implementazione

3.1 Architettura del Sistema

Il sistema sviluppato è stato progettato seguendo un approccio modulare basato su una pipeline di elaborazione che combina tecniche di preprocessing dei dati, analisi spettrale attraverso la Graph Fourier Transform, e costruzione di grafi. In particolare la pipeline di elaborazione è strutturata in cinque fasi principali:

- Caricamento e Preprocessing dei Dati: Lettura del dataset CSV, strutturazione dei dati temporali per stazione, e gestione dei valori mancanti attraverso interpolazione temporale e spaziale.
- Costruzione del Grafo k-NN: Creazione di un grafo non diretto basato sulla prossimità geografica tra le stazioni di monitoraggio, utilizzando un algoritmo k-nearest neighbors.
- Calcolo del Laplaciano e Decomposizione: Computazione del Laplaciano del grafo e sua decomposizione in autovalori e autovettori per ottenere le frequenze caratteristiche del grafo.
- Analisi mediante Graph Fourier Transform: Trasformazione dei segnali di inquinamento dal dominio spaziale al dominio delle frequenze del grafo, consentendo l'identificazione di pattern spazio-temporali.
- Visualizzazione e Interpretazione: Generazione di grafici multidimensionali per l'analisi dei risultati, inclusi spettrogrammi, mappe di correlazione e analisi dei modi autovettoriali.

Il codice è organizzato nei seguenti moduli principali:

- `main.m` - Script principale
- `carica_dati.m` - Modulo di caricamento dati
- `visualizza_grafici.m` - Modulo di visualizzazione

Lo script principale coordina l'intero processo di analisi, gestisce i parametri di configurazione (variabile di analisi, numero di vicini k, opzioni di visualizzazione) e orchestra la chiamata alle funzioni specializzate. Il modulo di caricamento dati implementa un sistema di caching intelligente che ottimizza le operazioni di I/O

salvando i dati preprocessati in formato .mat, riducendo così i tempi di caricamento nelle esecuzioni successive. Il modulo di visualizzazione genera un set completo di visualizzazioni scientifiche, inclusi grafici geografici interattivi con mappa satellitare, rappresentazioni 3D degli autovettori, spettrogrammi ad alta risoluzione e analisi dei pattern temporali.

3.2 Struttura del Codice MATLAB

3.2.1 Caricamento Dati

Il caricamento dei dati avviene tramite la funzione carica_dati() che segue questo ordine logico: innanzitutto verifica se i dati sono già disponibili nel workspace; in caso contrario controlla l'esistenza di un file .mat salvato; se nessuna di queste condizioni è soddisfatta, procede al caricamento diretto dal file CSV mediante la funzione carica_dati_da_csv(). In quest'ultimo caso, i dati vengono inoltre salvati sia in un file .mat sia nel workspace, così da poter essere riutilizzati in futuro.

```
1 function dati = carica_dati()
2     if evalin('base', 'exist(''dati_pechino'', ''var'')')
3         fprintf('Utilizzo dati precedentemente caricati
4                 dal workspace...\n');
5         dati = evalin('base', 'dati_pechino');
6         return;
7     end
8
9     if exist('dati_pechino.mat', 'file')
10        fprintf('Caricamento dati da file .mat salvato
11                ... \n');
11        load('dati_pechino.mat', 'dati_pechino');
12        dati = dati_pechino;
13        assignin('base', 'dati_pechino', dati);
14        return;
15    end
16
17    fprintf('Caricamento dati da CSV ... \n');
18    dati = carica_dati_da_csv();
19
20    dati_pechino = dati;
21    save('dati_pechino.mat', 'dati_pechino');
22    fprintf('Dati salvati in dati_pechino.mat\n');
```

```

23 assignin('base', 'dati_pechino', dati);
24 end

```

3.2.2 Gestione dei Valori Mancanti

La procedura di gestione dei valori mancanti si articola in tre fasi principali:

1. **Interpolazione temporale:** per ciascuna stazione, i valori mancanti vengono ricostruiti mediante interpolazione lineare nel tempo;
2. **Interpolazione spaziale:** se dopo la fase precedente persistono valori mancanti, questi vengono stimati utilizzando una media pesata inversamente proporzionale alla distanza tra stazioni;
3. **Fallback:** eventuali valori residui vengono sostituiti con la media globale.

```

1 ...
2 % Interpolazione lineare nel tempo
3 matrice_segnali_piena = matrice_segnali;
4 for stazione = 1:N
5     dati_stazione = matrice_segnali(:, stazione);
6     matrice_segnali_piena(:, stazione) = fillmissing(
7         dati_stazione, 'linear');
8 end
9
10 % Interpolazione spaziale
11 for t = 1:T
12     slice_temporale = matrice_segnali_piena(t, :);
13     stazioni_mancanti = isnan(slice_temporale);
14
15     if any(stazioni_mancanti) && sum(~stazioni_mancanti)
16         > 0
17         stazioni_disponibili = find(~stazioni_mancanti);
18         indici_stazioni_mancanti = find(stazioni_mancanti
19             );
20
21         for idx_mancante = indici_stazioni_mancanti
22             distanze_da_target = pdist2(coordinate(
23                 idx_mancante, :), coordinate(
24                 stazioni_disponibili, :));
25             if any(distanze_da_target > 0)
26                 pesi = 1 ./ distanze_da_target;
27                 pesi = pesi / sum(pesi);

```

```

23         matrice_segnali_piena(t, idx_mancante) =
24             sum(pesi .* slice_temporale(
25                 stazioni_disponibili));
26     end
27 end
28
29 % Media globale
30 media_globale = mean(matrice_segnali_piena(:, 'omitmissing'));
31 matrice_segnali_piena(isnan(matrice_segnali_piena)) =
32     media_globale;
33 matrice_segnali = matrice_segnali_piena;
34 matrice_segnali(isinf(matrice_segnali)) = media_globale;
35 ...

```

3.2.3 Costruzione del Grafo della Rete di Stazioni

Il grafo della rete viene costruito a partire dalle coordinate delle stazioni. Dopo aver calcolato la matrice delle distanze euclidi, per ciascun nodo vengono selezionati i k vicini più prossimi, con i quali si stabiliscono le connessioni. La matrice di adiacenza **A** risultante viene quindi resa simmetrica per ottenere un grafo non orientato. Infine, dalla stessa si ricava la matrice diagonale dei gradi **D**, che contiene il numero di connessioni associate a ciascun nodo.

```

1 ...
2 distanze = pdist2(coordinate, coordinate);
3
4 A = zeros(N);
5 [~, indici_ordinati] = sort(distanze, 2);
6
7 for i = 1:N
8     vicini = indici_ordinati(i, 2:k+1);
9     A(i, vicini) = 1;
10 end
11
12 A = double(A + A');
13 D = diag(sum(A));
14 ...

```

3.2.4 Calcolo del Laplaciano e dei Coefficienti della GFT

Il Laplaciano del grafo viene calcolato come differenza tra la matrice dei gradi **D** e la matrice di adiacenza **A**. Successivamente, si effettua la decomposizione agli autovalori, ordinando le frequenze del grafo in ordine crescente. Gli autovettori corrispondenti costituiscono la base ortogonale della GFT. Infine, si ottengono i coefficienti proiettando la matrice dei segnali sulla base spettrale.

```
1 ...
2 L = (D - A);
3
4 [U, Lambda] = eig(L, 'vector');
5 [autovalori, idx] = sort(Lambda);
6 U = U(:, idx);
7
8 coefficienti_gft = (matrice_segnali * U);
9 ...
```

3.2.5 Analisi della Distribuzione Spettrale di Potenza

A partire dai coefficienti della GFT, viene calcolata la potenza spettrale come media del modulo quadro dei coefficienti stessi. Ordinando i valori ottenuti in ordine decrescente, si identificano le frequenze del grafo più rilevanti in termini di contributo energetico. Infine, vengono riportate sul terminale le prime cinque frequenze significative, con i rispettivi indici degli autovalori e i corrispondenti livelli di potenza.

```
1 ...
2 grafo_dsp = mean(abs(coefficienti_gft).^2, 1);
3
4 [~, freq_significative] = sort(grafo_dsp, 'descend');
5
6 fprintf('\nFrequenze del grafo piu significative:\n');
7 for i = 1:5
8     idx_freq = freq_significative(i);
9     fprintf('Frequenza grafo %.4f (indice autovalore %2d)
10          : Potenza = %.4f\n', ...
11             autovalori(idx_freq), idx_freq, grafo_dsp(
12                 idx_freq));
13 end
14 ...
```

3.2.6 Visualizzazione

La fase di visualizzazione comprende cinque grafici principali, finalizzati a rappresentare sia la struttura della rete sia l'evoluzione dei segnali nello spazio e nel tempo:

1. **Struttura geografica della rete:** rappresenta le stazioni sul piano geografico e le connessioni del grafo, evidenziando la distribuzione spaziale della rete.
2. **Spettro degli autovalori:** mostra l'insieme degli autovalori del Laplaciano, interpretati come frequenze del grafo, utili per analizzare le proprietà spettrali della rete.
3. **Modi autovettoriali:** visualizza i primi sei autovettori del Laplaciano in una rappresentazione tridimensionale, fornendo un'intuizione sui modi oscillatori della rete.
4. **Spettrogramma:** mette in evidenza la dinamica temporale delle frequenze più rilevanti.
5. **Pattern spazio-temporali:** comprende diverse visualizzazioni sintetiche, tra cui le medie per stazione, i pattern diurni e stagionali, la matrice di correlazione tra stazioni e lo spettro di potenza sul dominio del grafo.

L'implementazione MATLAB dei grafici è contenuta nel file `visualizza_grafici.m`, tuttavia il loro codice non verrà aggiunto di seguito perché si reputa poco significativo. Invece i render grafici di tutte le possibili immagini realizzabili sono presenti all'interno della cartella "img" di GitHub.

4. Risultati Sperimentali

4.1 Selezione del Parametro k per il Grafo k-NN

Inizialmente sono stati valutati tre diversi valori di k per generare il grafo della rete di stazioni meteorologiche. In particolare sono stati valutati $k = 2$, $k = 4$, $k = 6$. Le reti risultanti sono le seguenti.

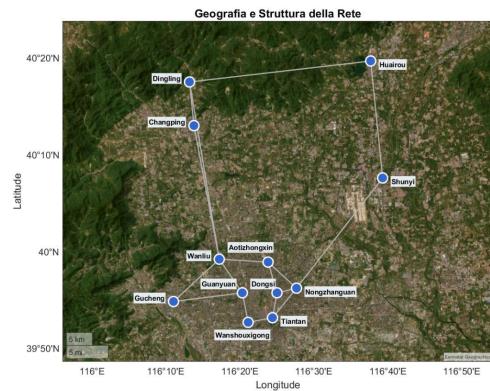


Figura 4.1: Grafo con $K = 2$

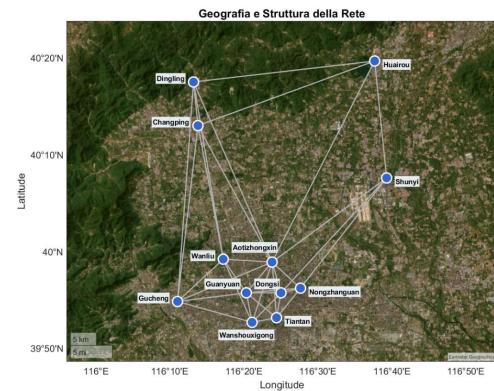


Figura 4.2: Grafo con $K = 4$

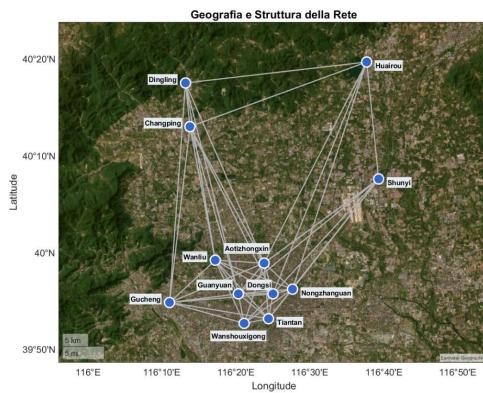


Figura 4.3: Grafo con $K = 6$

In tutti e tre i casi analizzati, il grafo risulta costituito da un'unica componente连通的. Al variare di k si notano le seguenti differenze:

- con $k = 2$ il grafo risulta relativamente scarno, con 17 archi complessivi e una media di circa 2,83 archi per nodo;

- passando a $k = 4$, il numero di archi raddoppia, arrivando a 35, con un grado medio che sale a 5,83;
- infine, con $k = 6$ la struttura diventa ancora più densa, raggiungendo i 48 archi totali e un grado medio esattamente pari a 8.

Per evidenziare meglio le differenze strutturali, vengono inoltre riportati i grafici degli autovalori, che offrono una prospettiva utile sull'evoluzione delle proprietà spettrali del grafo al crescere di k .

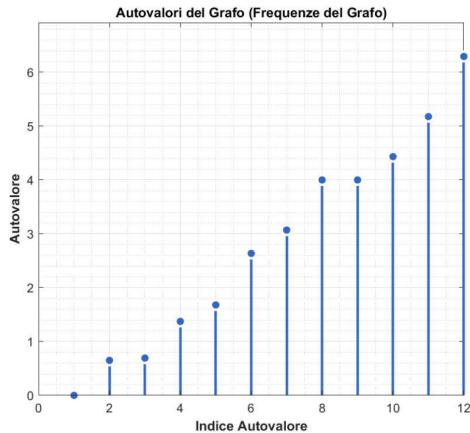


Figura 4.4: Autovalori del grafo con $K = 2$

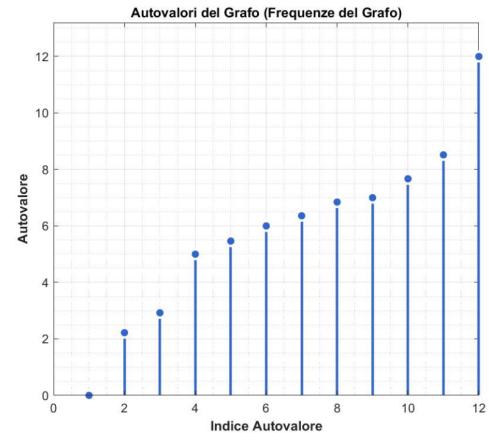


Figura 4.5: Autovalori del grafo con $K = 4$

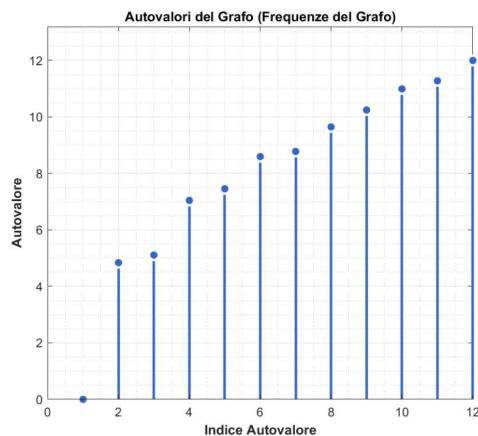


Figura 4.6: Autovalori del grafo con $K = 6$

Osservando i tre grafici degli autovalori al variare di k , emergono chiaramente le differenze nella struttura spettrale:

- Con $k = 2$, lo spettro degli autovalori si distribuisce in un intervallo relativamente ristretto, con valori che vanno da 0 fino a circa 6,3. I primi autovalori hanno valori molto ridotti, indicando una struttura con connettività limitata.
- Aumentando a $k = 4$, si osserva un allargamento dello spettro con l'autovalore massimo che cresce fino a 12, raddoppiando rispetto al caso precedente. La

distribuzione diventa più uniforme e il gap tra il terzo e quarto autovalore diventa molto evidente, suggerendo che il grafo abbia tre cluster.

- Con $k = 6$, lo spettro mantiene l'autovalore massimo a 12 ma mostra una distribuzione ancora più densa e regolare degli autovalori intermedi. La presenza di valori più elevati anche nelle posizioni centrali dello spettro testimonia la crescente densità della rete.

L'analisi spettrale conferma quantitativamente quanto osservato sulla struttura: all'aumentare di k , il grafo non solo guadagna archi, ma acquisisce anche proprietà spettrali che indicano una maggiore robustezza e interconnessione. Successivamente vengono mostrati i primi sei autovettori rappresentati sui relativi grafi.

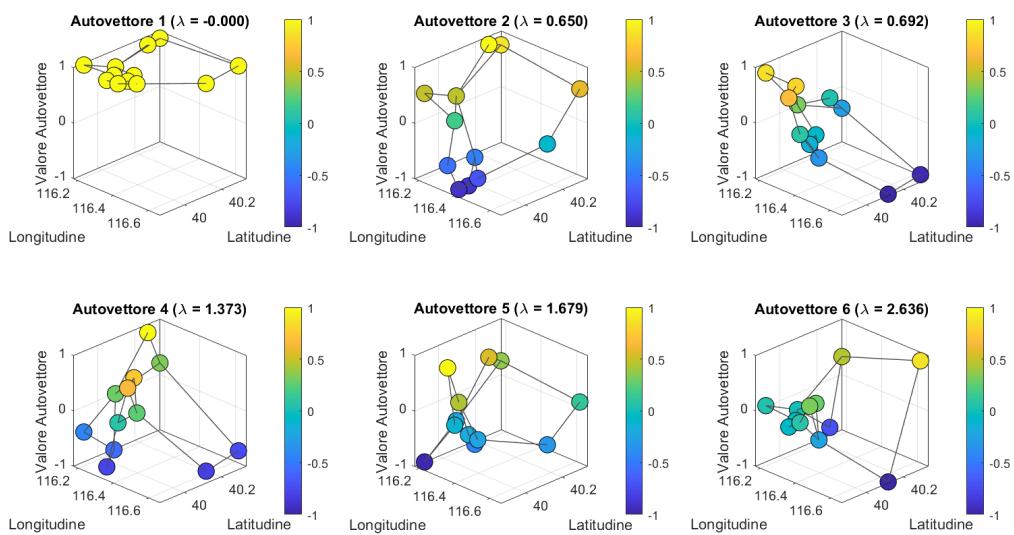


Figura 4.7: Autovettori del grafo con $K = 2$

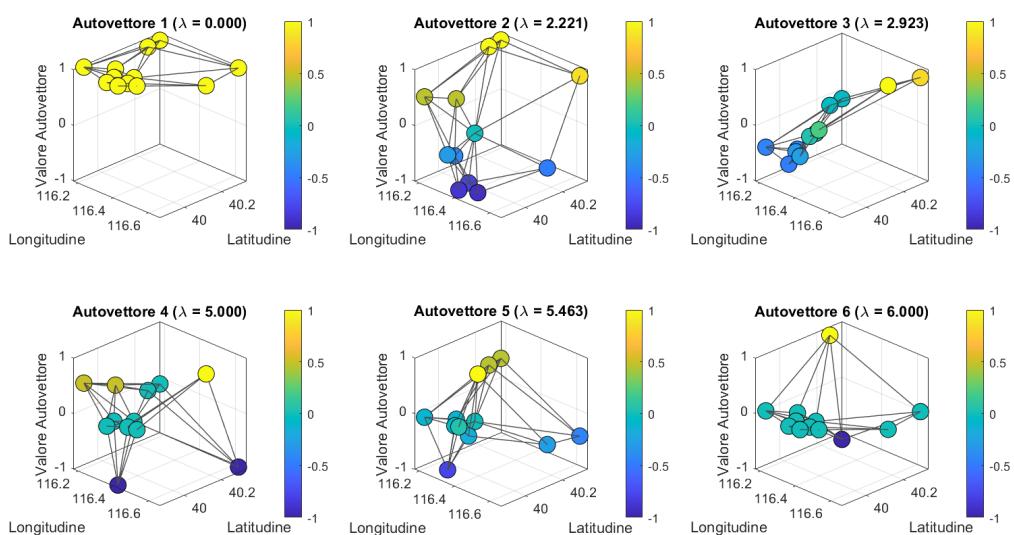


Figura 4.8: Autovettori del grafo con $K = 4$

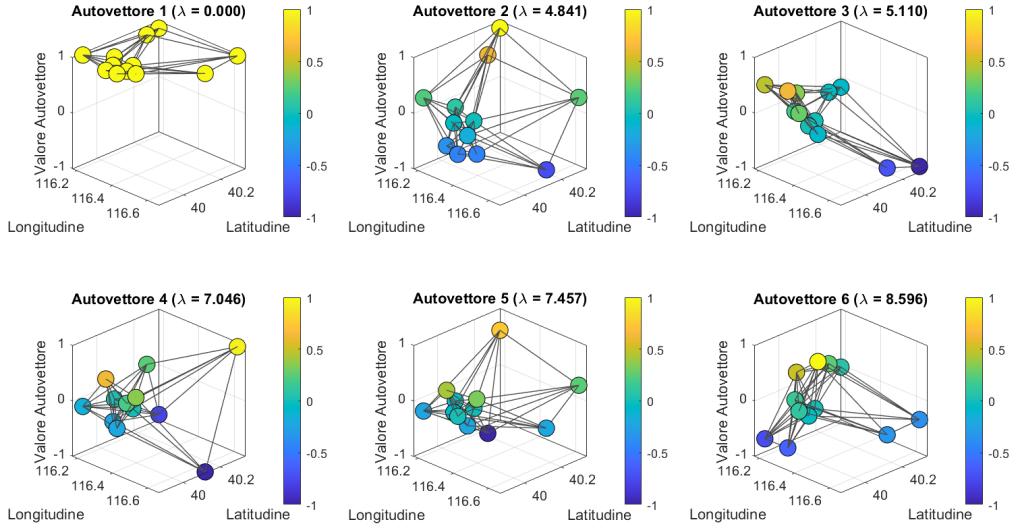


Figura 4.9: Autovettori del grafo con $K = 6$

La scelta del valore ottimale di k richiede un bilanciamento tra diversi fattori strutturali del grafo. Da un lato, un valore troppo basso rischia di produrre una rete eccessivamente sparsa, con connessioni insufficienti a catturare le relazioni significative tra i nodi; dall’altro, un k troppo elevato può generare un grafo troppo denso, in cui le connessioni diventano ridondanti e la struttura perde capacità discriminativa. Il valore $k = 4$ rappresenta un compromesso ottimale in quanto garantisce una connettività sufficiente a mantenere la coesione della rete, preservando allo stesso tempo la capacità di distinguere i pattern significativi nei dati. Per questi motivi, è stato selezionato $k = 4$ come parametro per le analisi successive.

4.2 Analisi Multi-Variabile dell’Inquinamento Atmosferico

Il dataset comprende diverse variabili di interesse, quindi per evitare ridondanze e mantenere l’analisi focalizzata sui fenomeni più significativi, si è scelto di approfondire solamente un sottoinsieme rappresentativo di queste variabili. Le variabili selezionate per l’analisi dettagliata sono il PM2.5, il NO₂, l’O₃, la temperatura e le precipitazioni. Il PM2.5 rappresenta l’inquinante più critico per la salute umana e tipicamente mostra pattern regionali ben definiti. Il NO₂, strettamente legato alle emissioni da traffico veicolare e attività industriali, presenta invece una maggiore variabilità spaziale e comportamenti più localizzati, in contrasto con il particolato fine. L’ozono è stato incluso per la sua dinamica completamente diversa dagli altri inquinanti: essendo un inquinante secondario di origine fotochimica, presenta pattern diurni e stagionali opposti rispetto agli inquinanti primari. Le variabili meteorologiche (temperatura e precipitazioni) sono state aggiunte per mostrare come

i fattori ambientali si comportino diversamente dagli inquinanti nella trasformata di Fourier sul grafo, evidenziando meccanismi di diffusione spaziale alternativi. Le variabili non trattate in dettaglio mostrano comportamenti analoghi a quelle selezionate, ad esempio il PM10 presenta pattern molto simili al PM2.5, mentre il SO₂ e il CO hanno concentrazioni generalmente basse nelle aree urbane moderne. Inoltre gli altri parametri meteorologici tendono ad avere pattern correlati con la temperatura e le precipitazioni. I grafici completi per tutte le variabili analizzate sono disponibili nel repository GitHub del progetto.

4.2.1 Particolato Fine (PM2.5)

Il PM2.5 mostra una concentrazione dell'energia nelle basse frequenze del grafo, visibile nello spettrogramma 4.10 come una banda persistente nel tempo. Questo è confermato dallo spettro di potenza (Figura 4.11) che è dominato quasi interamente dalla prima frequenza del grafo. La distribuzione spaziale appare notevolmente omogenea tra le stazioni. Questo suggerisce che il particolato fine sia influenzato da dinamiche meteorologiche su scala regionale piuttosto che da sorgenti puntuali. Durante la giornata, le concentrazioni seguono un andamento caratteristico con minimi nelle ore centrali, quando l'altezza del mixing layer favorisce la dispersione, e massimi durante la notte. Quest'ultimo comportamento riflette sia le ore di traffico intenso sia le condizioni di inversione termica tipiche di questi momenti. La matrice di correlazione inter-stazione conferma questa lettura con valori elevati e uniformi. Questo perché le diverse località sperimentano variazioni simili, rafforzando l'ipotesi di un fenomeno a carattere regionale.

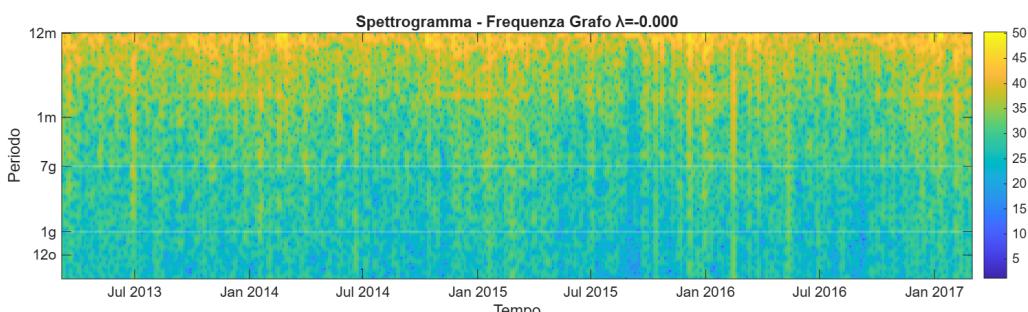


Figura 4.10: Spettrogramma PM2.5

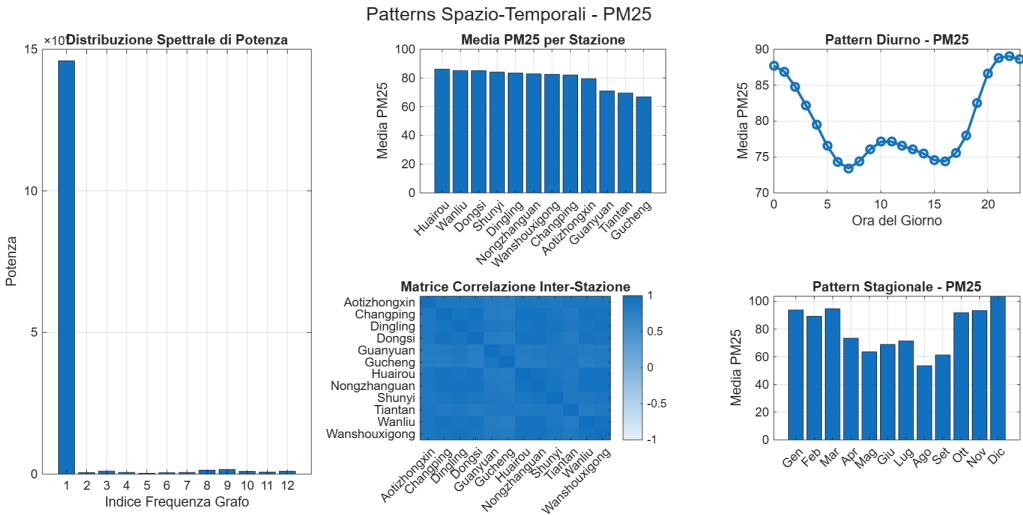


Figura 4.11: Patterns Spazio-Temporali PM2.5

4.2.2 Biassido di Azoto (NO_2)

A differenza del particolato, il biassido di azoto presenta una firma spettrale più articolata. La prima frequenza mantiene un ruolo importante ma non esclusivo, mentre le frequenze intermedie contribuiscono in modo significativo (Figura 4.13). Questa distribuzione riflette la natura più localizzata dell'inquinante, legato a sorgenti emissive specifiche come il traffico veicolare. Un'eterogeneità territoriale emerge chiaramente dalla media per stazione, ad eccezione di Dingling che registra concentrazioni quasi doppie rispetto alle stazioni meno inquinate. La conformazione geografica della zona e l'intenso flusso turistico verso i siti limitrofi potrebbero spiegare questa anomalia. Il ciclo giornaliero presenta caratteristiche peculiari. I valori massimi si osservano in piena notte (ore 22-2) e rimangono elevati fino a metà mattina, per poi crollare nel primo pomeriggio. Questo è dovuto alla radiazione solare che innesca la fotolisi del NO_2 convertendolo in altri composti e riducendone temporaneamente la concentrazione. Lo spettrogramma 4.12 mostra una distribuzione energetica meno marcata rispetto al PM2.5, con maggiore variabilità nelle frequenze intermedie, segno di eventi di inquinamento più transitori e localizzati.

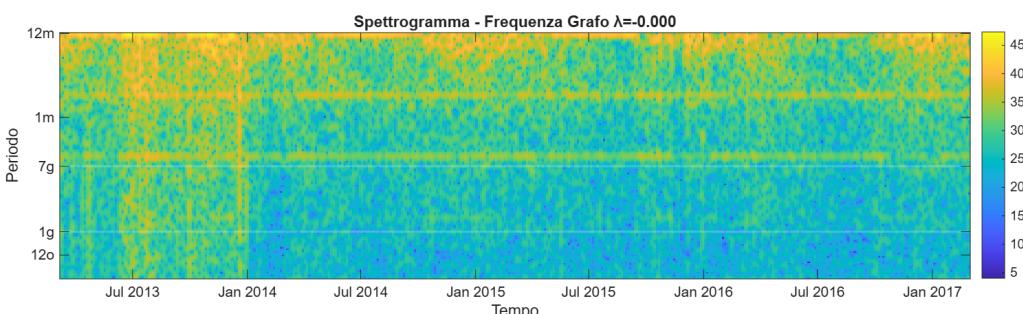


Figura 4.12: Spettrogramma NO_2

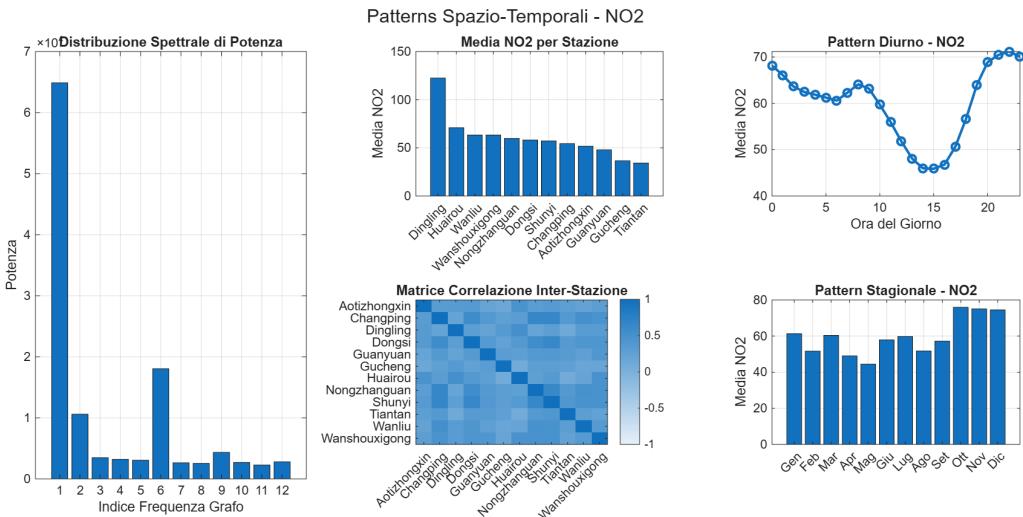


Figura 4.13: Patterns Spazio-Temporali NO₂

4.2.3 Ozono (O₃)

L’ozono si differenzia dai pattern osservati negli inquinanti precedenti. Nello spettrogramma 4.14, l’energia si concentra nei mesi estivi mentre l’inverno appare stabile e poco variabile. Questa inversione stagionale deriva dalla natura fotochimica dell’O₃ che necessita di radiazioni solari intense per innescare le reazioni fotochimiche che lo producono a partire dai suoi precursori. Il ciclo giornaliero rispecchia questa dipendenza, infatti i valori minimi caratterizzano le ore notturne quando mancano sia l’irraggiamento sia i processi fotochimici, mentre i valori massimi si raggiungono nelle ore centrali del giorno. La distribuzione spettrale presenta di nuovo una prima frequenza dominante, ma con articolazioni più evidenti nelle frequenze intermedie rispetto al PM2.5. Nonostante la formazione di ozono sia prevalentemente un fenomeno regionale, le diverse concentrazioni dei precursori introducono alcune lievi variazioni locali. Le correlazioni inter-stazione rimangono elevate, indicando che i processi su larga scala governano comunque la dinamica complessiva.

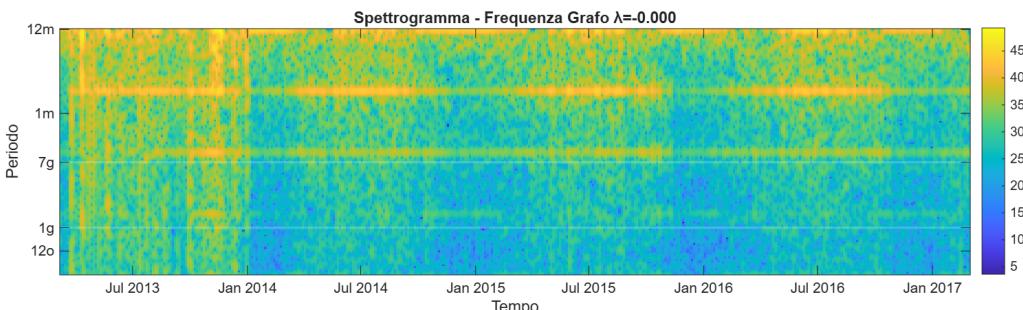


Figura 4.14: Spettrogramma O₃

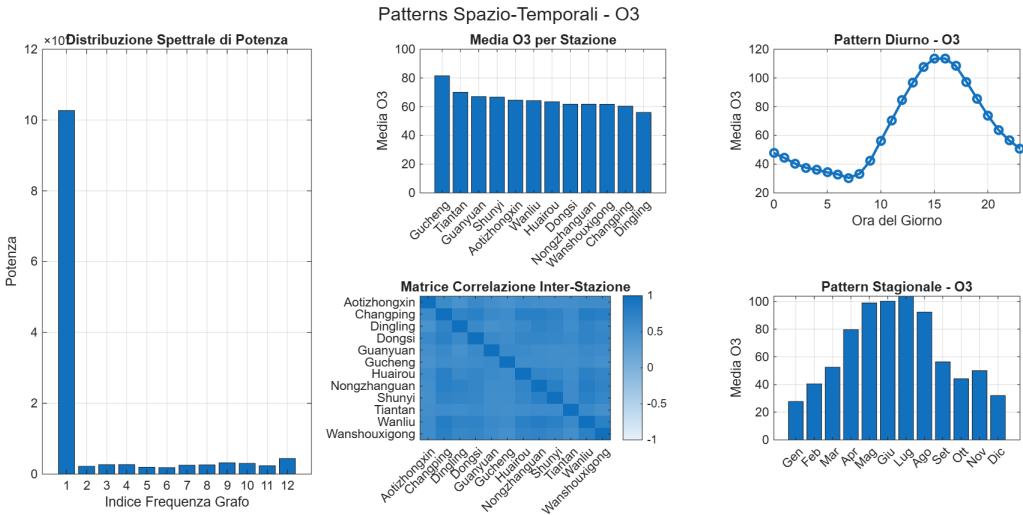


Figura 4.15: Patterns Spazio-Temporali O₃

4.2.4 Temperatura

La temperatura fornisce un interessante termine di paragone rispetto agli inquinanti. Lo spettrogramma 4.16 rivela una periodicità straordinariamente regolare e definita, con cicli giornalieri, settimanali, mensili e annuali immediatamente riconoscibili. A differenza degli inquinanti, soggetti a eventi episodici e fluttuazioni irregolari, la temperatura segue pattern strutturati e prevedibili risultando in uno spettrogramma visivamente più ordinato. La concentrazione energetica sulla prima frequenza è ancor più accentuata che per il PM2.5, con i contributi delle altre frequenze che sono praticamente trascurabili (Figura 4.17). Da questo si intuisce che la temperatura è quasi perfettamente uniforme su scala spaziale nell'intera regione considerata. Il pattern diurno segue ovviamente il ciclo solare con minimi notturni e massimi pomeridiani, mentre la matrice di correlazione mostra valori estremamente elevati. I processi atmosferici che determinano la temperatura operano su scala ben più ampia dell'area metropolitana di Pechino, interessando uniformemente tutte le stazioni.

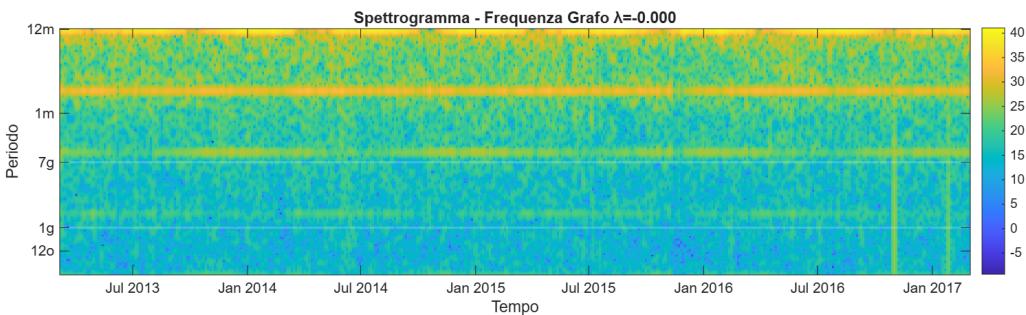


Figura 4.16: Spettrogramma Temperatura

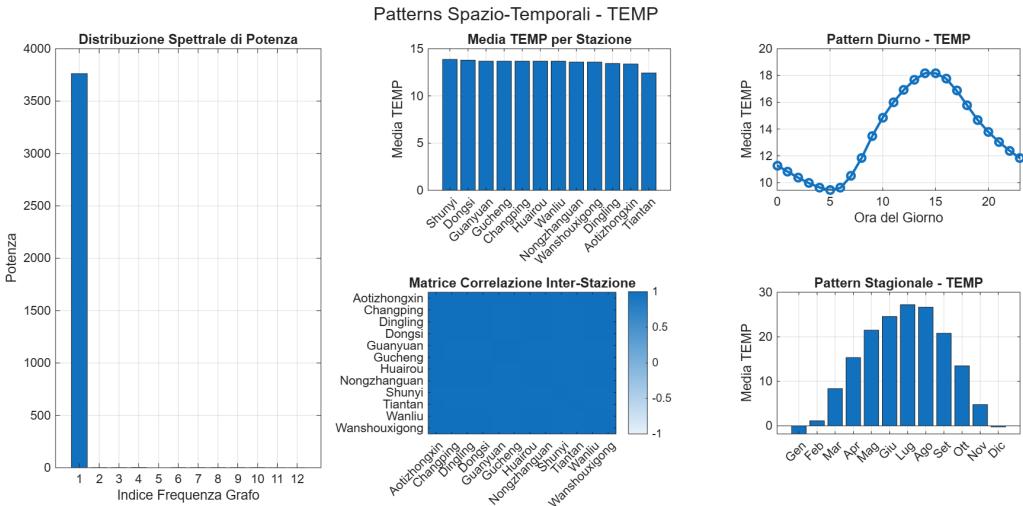


Figura 4.17: Patterns Spazio-Temporali Temperatura

4.2.5 Precipitazioni

Le precipitazioni si distinguono nettamente da tutte le altre variabili per la loro natura intrinsecamente discontinua. Lo spettrogramma 4.18 presenta un’alternanza rapida e irregolare tra eventi intensi e periodi di assenza totale. Non ci sono transizioni graduali come per la temperatura, infatti le precipitazioni sono naturalmente discrete, episodiche e localizzate temporalmente. Comunque si nota una certa stagionalità estiva che riflette l’influenza dei monsoni sul clima. Anche la distribuzione spettrale si comporta diversamente (Figura 4.19). Pur mantenendo una prima frequenza dominante, i contributi sulle frequenze intermedie e alte sono significativamente più distribuiti. Le precipitazioni hanno certamente una componente regionale ma presentano anche marcata variabilità locale. La matrice di correlazione interstazione conferma questa duplice natura con valori moderati e più variabili rispetto alle altre grandezze analizzate. L’evidente contrasto stagionale mostra massimi estivi e minimi invernali confermando nuovamente l’impronta del monsone estivo sulla capitale cinese.

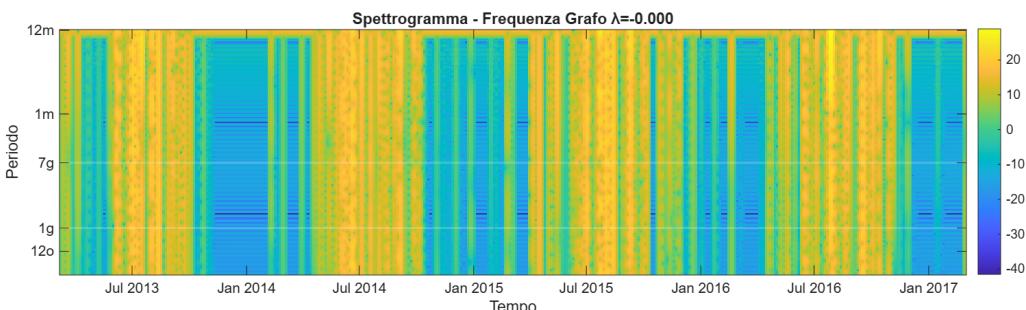


Figura 4.18: Spettrogramma Precipitazioni

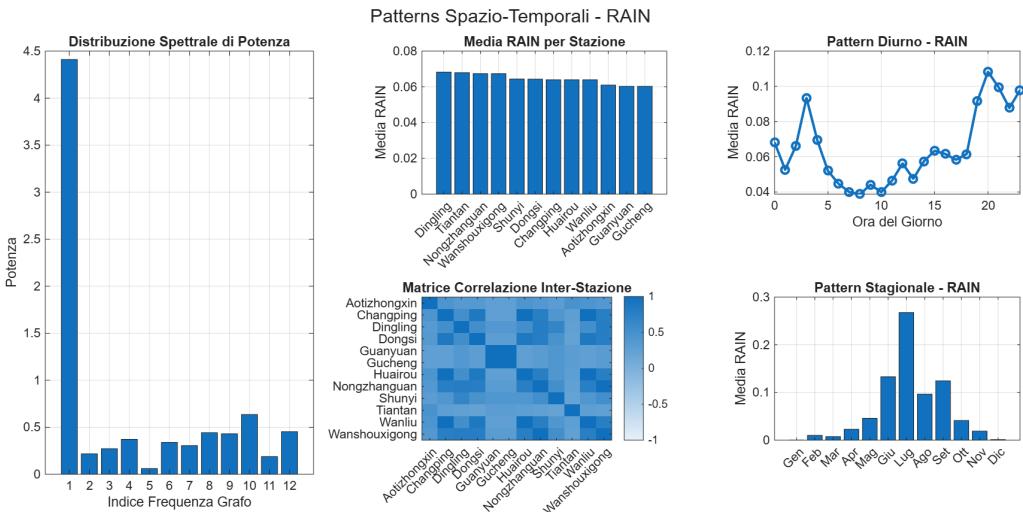


Figura 4.19: Patterns Spazio-Temporali Precipitazioni

5. Conclusioni

In questo progetto è stato affrontato il problema dell'analisi delle dinamiche temporali e spaziali degli inquinanti atmosferici a Pechino. A partire da un dataset reale, è stato costruito un grafo in cui le stazioni di monitoraggio rappresentano i nodi, mentre le relazioni spaziali sono state modellate mediante connessioni basate sulla vicinanza geografica. Attraverso l'utilizzo del Laplaciano di questa struttura è stata ottenuta una rappresentazione spettrale dei segnali ambientali.

L'elaborazione ha evidenziato come i tre inquinanti considerati presentino comportamenti distinti sia in termini di distribuzione spettrale sia nella variabilità temporale. Nel dettaglio, il particolato fine (PM_{2.5}) ha mostrato una forte eterogeneità spaziale e temporale, confermata dalle componenti armoniche dominanti. Il biossido di azoto (NO₂) ha messo in luce le differenze territoriali tra le varie stazioni, indicando l'aspetto più locale del fenomeno. L'ozono (O₃) invece presenta caratteristiche più marcate in termini di stagionalità.

Dal punto di vista metodologico, la fase di pre-processing si è rivelata cruciale per garantire l'affidabilità delle analisi, permettendo di gestire in maniera sistematica i dati mancanti e ridurre la disomogeneità temporale tra le misurazioni. L'applicazione della trasformata di Fourier su grafi ha permesso di evidenziare pattern non immediatamente osservabili nel dominio del tempo, fornendo una chiave di lettura diversa rispetto a metodi statistici tradizionali.

Nel complesso l'elaborato ha dimostrato l'efficacia della GSP come approccio per lo studio dei fenomeni ambientali su reti di sensori distribuiti. L'analisi ha messo in evidenza non solo la possibilità di rappresentare in modo compatto l'informazione, ma anche di isolare le componenti strutturali e dinamiche che caratterizzano le diverse sostanze inquinanti.