Multi-scale Assessment of Water Quality

Structure, Multivariate Regimes, and Heavy Metal Hotspots

A Comprehensive Analysis of Chicago Waterways

The Urban River Challenge

Urban river systems like Chicago's face a complex matrix of pressures. Unlike pristine environments, they are impacted by a mix of nutrient enrichment, industrial suspended sediments, and toxic heavy metals.

The Paradigm Shift:

Traditional monitoring often isolates these pollutants (measuring nitrate alone, or lead alone). This report adopts a **multivariate, process-oriented approach** to understand how these pollutants interact and cluster in space and time.

Core Research Questions

RQ1: Structure

Do pollutant variables form coherent categories? Which variables dominate each group?

Focus: Internal Cohesion

RQ2: Patterns

Can unsupervised clustering reveal unique water quality patterns or "regimes"?

Focus: PCA & K-Means

RQ3: Hotspots

When and where do heavy metal concentrations abnormally increase?

Focus: Spatiotemporal Analysis

Data & Analytical Framework

Dataset Overview

- **Source:** Multi-year monitoring dataset from Chicago waterways.
- **Variables:**
  - Water Quality (TEMP, DO, pH, SS, VSS)
  - Nutrients (Nitrogen species, Total P)
  - Metals/Minerals (Fe, Mn, Cu, Zn, Ni, Hg, As, Ca, Mg)
- **Preprocessing:** Z-score standardization to eliminate unit effects.

Workflow

1. Categorical Structure Analysis

↓

2. PCA & K-Means Clustering

↓

3. Outlier & Hotspot Detection

Finding 1: Pollutant Structure & Cohesion

Nutrients

Most Internally Consistent

cohesion=0.348

Nitrogen and Phosphorus levels tend to rise and fall together, forming a tight system.

Water Quality

Moderate Consistency

cohesion=0.257

Driven primarily by physical factors like Suspended Solids (SS) and pH.

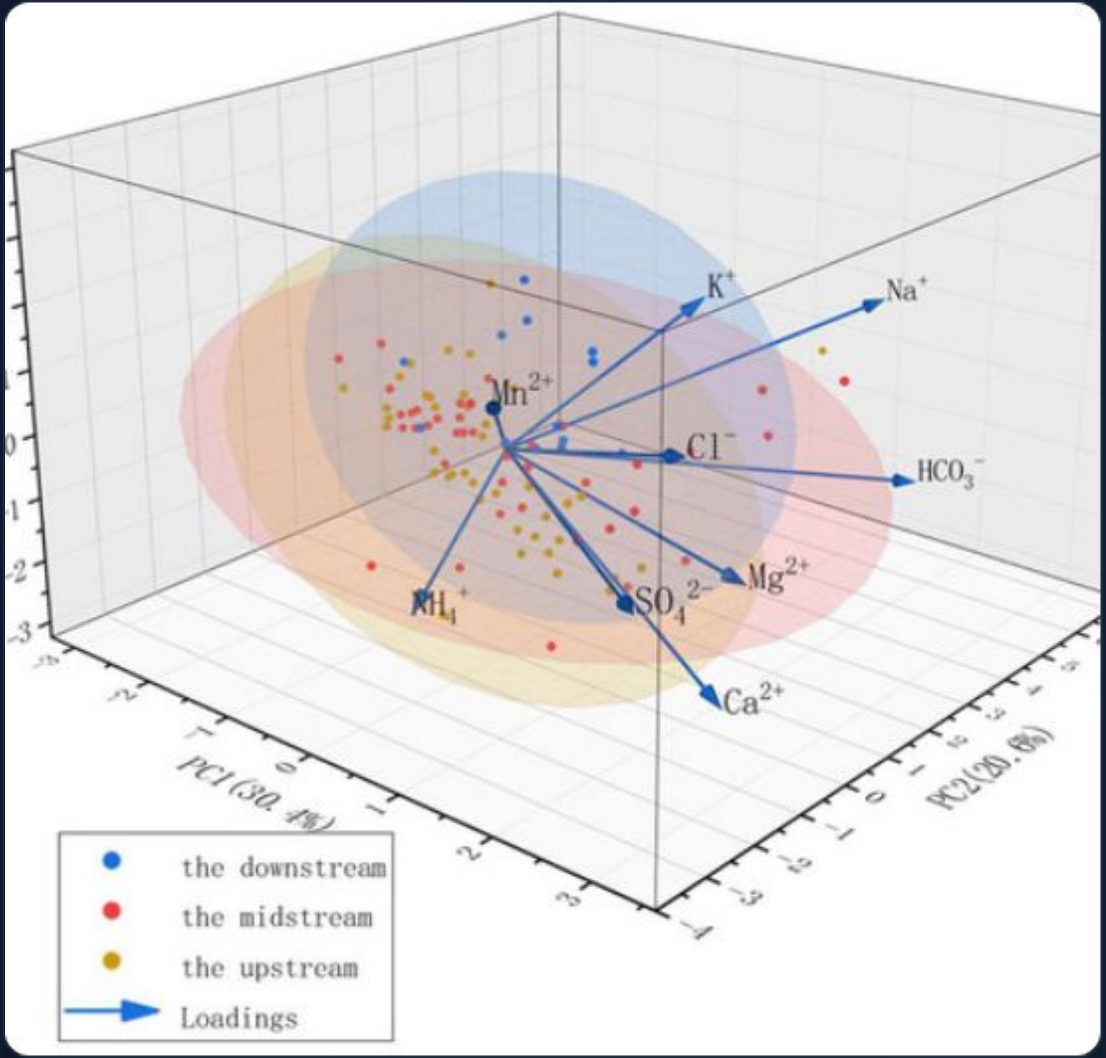
Heavy Metals

Highest Impact, Lowest Cohesion

cohesion=0.157

Metals act independently but exert the strongest statistical influence on overall water quality.

Finding 2: Multivariate Regimes (K-Means)



Two Distinct Patterns

Cluster 0: Metal-Dominated

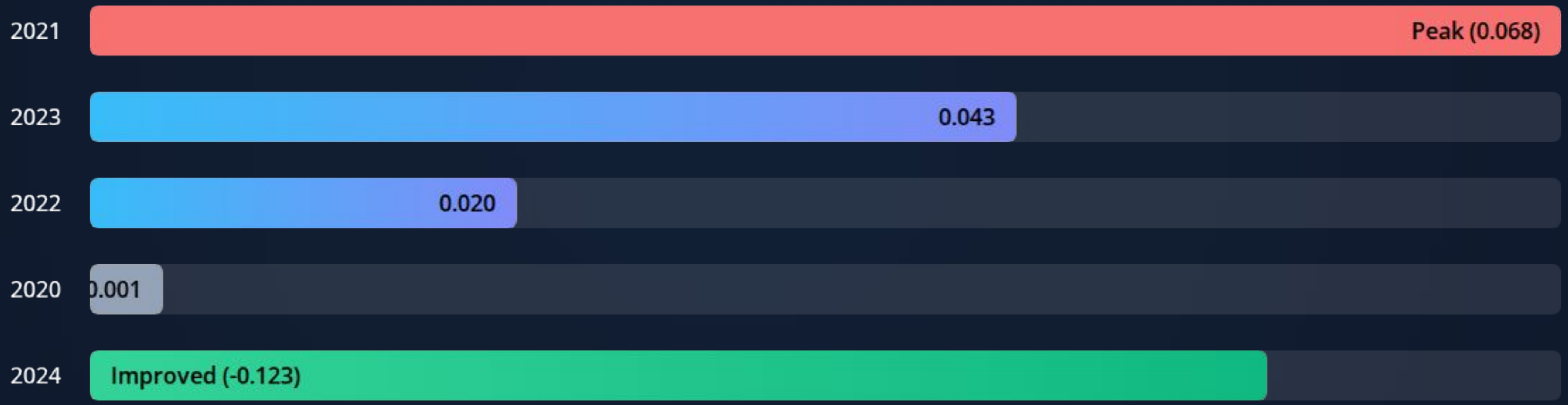
Small, discrete group (n≈165). Characterized by high Metal Pollution Index (HPI ≈ 58.8). Represents localized, acute contamination events.

Cluster 1: Multi-Stressor

Widespread group (n≈2,781). Lower average metal risk (HPI ≈ 33.8) but reflects the broad, diffuse background of nutrients and solids.

Finding 3: Temporal Anomalies (Yearly)

Standardized heavy metal pollution index by year. **2021 was the most polluted year**, while 2024 showed significant improvement.

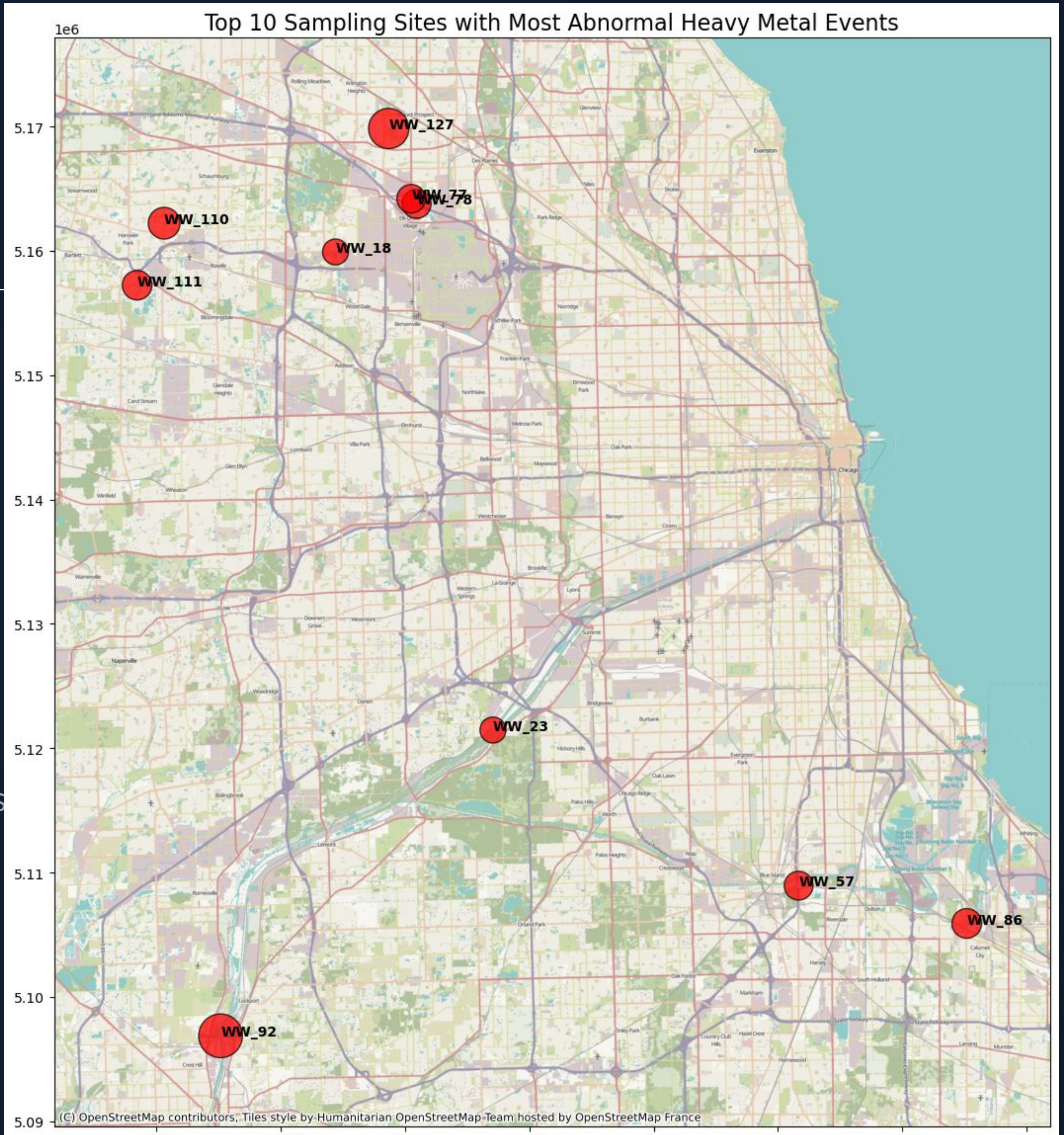


Seasonal Insight: Anomalies consistently peak in February and March, likely due to hydrological conditions favoring metal resuspension.

Spatial Hotspots

Heavy metal pollution is not evenly distributed. It concentrates in specific "hotspot" zones.

- **Top Hotspots:** Sites like WW_110 and WW_18 recorded over 800 anomalies.
- **Key Contaminants:** Manganese, Iron, Zinc, and Barium are the most frequent outliers.
- **Implication:** These sites likely represent locations with historical industrial sediment or active infrastructure inputs.



Integrated Discussion

Nutrients

Nutrients form the "backbone" of the chemical system—internally consistent and widespread. They represent long-term ecological risks like eutrophication.

Metals

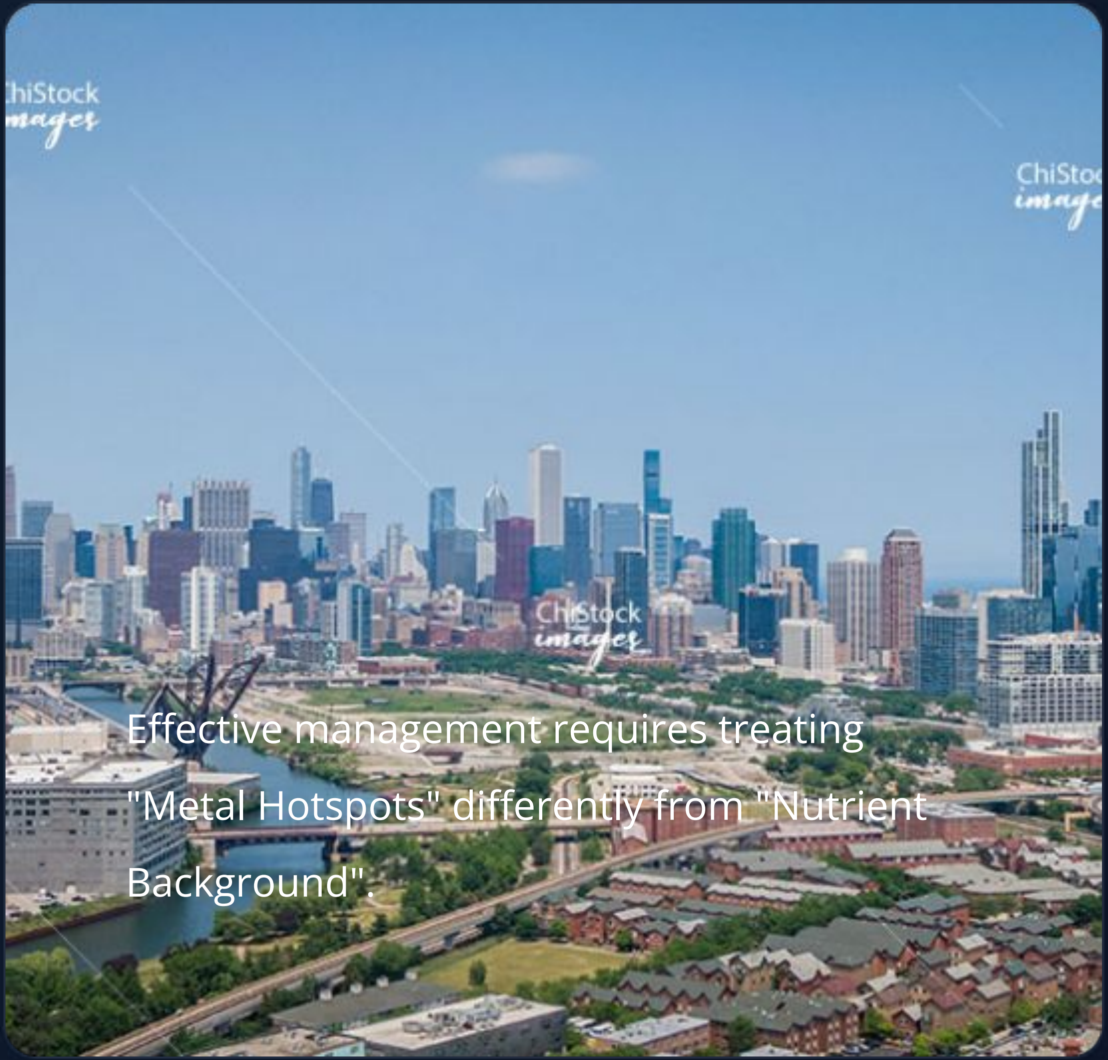
Metals are "disruptors". They are less cohesive but drive acute water quality drops. They appear as distinct peaks (Cluster 0) rather than a constant baseline.

Synthesis

The river network is largely a "diffuse multi-stressor" system (Cluster 1), punctuated by intense, localized "metal events" (Cluster 0) at specific hotspots.

Management Implications

- **Target Hotspots:** Focus remediation on the top 10 identified sites (e.g., WW_110).
- **Seasonal Vigilance:** Increase sampling frequency in late Winter/Early Spring (Feb-Mar).
- **Integrated Strategy:** Combine sediment management (for solids/metals) with runoff control (for nutrients).
- **Early Warning:** Use Calcium/Magnesium trends to predict shifts in metal dominance.



Effective management requires treating "Metal Hotspots" differently from "Nutrient Background".

Limitations & Future Work

Current Limitations

- **Temporal Resolution:** The dataset may miss rapid pulse events (storms) or diurnal cycles.
- **Biotic Gap:** Lack of biological indicators (algae, fish) limits ecological impact assessment.
- **Linear Methods:** PCA is linear; non-linear methods (UMAP) might reveal finer structures.

Future Directions

Integration of Biological Data:
Future studies should correlate these chemical clusters with chlorophyll-a or macroinvertebrate health to validate ecological harm.