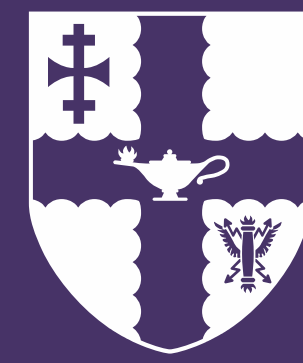


Source Apportionment AutoEncoder (SAAE): A Deep Learning Framework for Environmental Receptor Modelling

Amirhossein Parsi

Supervisors: Dr. Andrew Watson, Prof. Kambiz Ebrahimi



Loughborough University

1. Introduction

What is source apportionment?

Source apportionment techniques connect ambient air pollution back to its original emission sources.

What is receptor modelling?

Operating on the principle of mass conservation, receptor models mathematically separate chemical measurements into underlying source profiles and their time-based contributions.

What are state-of-practice receptor models?

Common methods include Chemical Mass Balance (CMB), Non-Negative Matrix Factorization (NMF), and Positive Matrix Factorization (PMF). PMF is a cornerstone of air quality research, with thousands of published applications worldwide [1].

2. Limitations of traditional models

■ Subjective interpretation:

Traditional solvers extract purely mathematical constructs, forcing researchers to manually inspect and label sources, which introduces subjective bias [2].

■ No predictive capability:

Because the source profiles and temporal contributions are solved simultaneously for a specific dataset, traditional models cannot compute contributions for unseen observations without re-running the model [3].

■ Lack of flexibility:

As strict mathematical solvers, traditional models cannot capture complex, non-linear atmospheric interactions or easily integrate external features like meteorological data to refine the solution [4].

3. Source Apportionment AutoEncoder (SAAE)

SAAE is a novel deep learning framework that executes receptor modelling entirely within a neural network environment. The main equation solved in receptor modelling can be described as:

$$X = GF + E$$

Where X represents receptor measurements, G temporal contributions, F source profiles, and E reconstruction error. SAAE solves this using a custom autoencoder-like architecture:

Encoder: A non-linear function f , parametrized by θ , maps X to G :

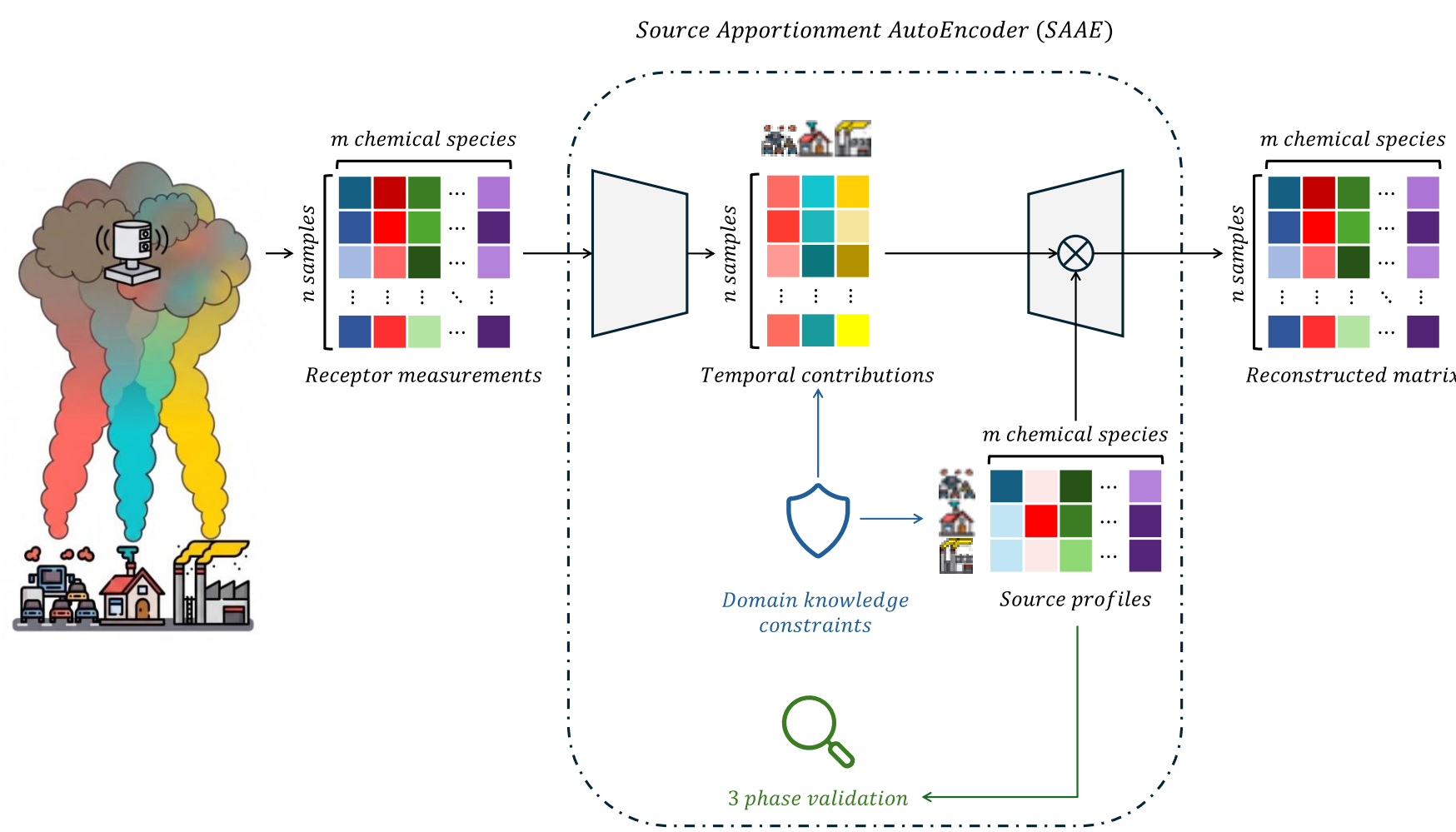
$$G = f_{\theta}(X)$$

Decoder: A single layer decoder function d , with weight matrix W_{dec} and no bias, extracts F :

$$F = d(W_{\text{dec}})$$

Reconstruction: The network natively reconstructs the original dataset through direct matrix multiplication:

$$f_{\theta}(X)d(W_{\text{dec}}) = GF \approx X$$



Domain knowledge integration: Known physical constraints can be hard-coded directly into the architecture and auxiliary penalty terms can be added to the loss function to pull the solution toward physical validity.

Three-phase validation protocol: A three-phase protocol validates the results across stability (initializations), statistical uncertainty (bootstrapping), and rotational ambiguity (boundary optimization).

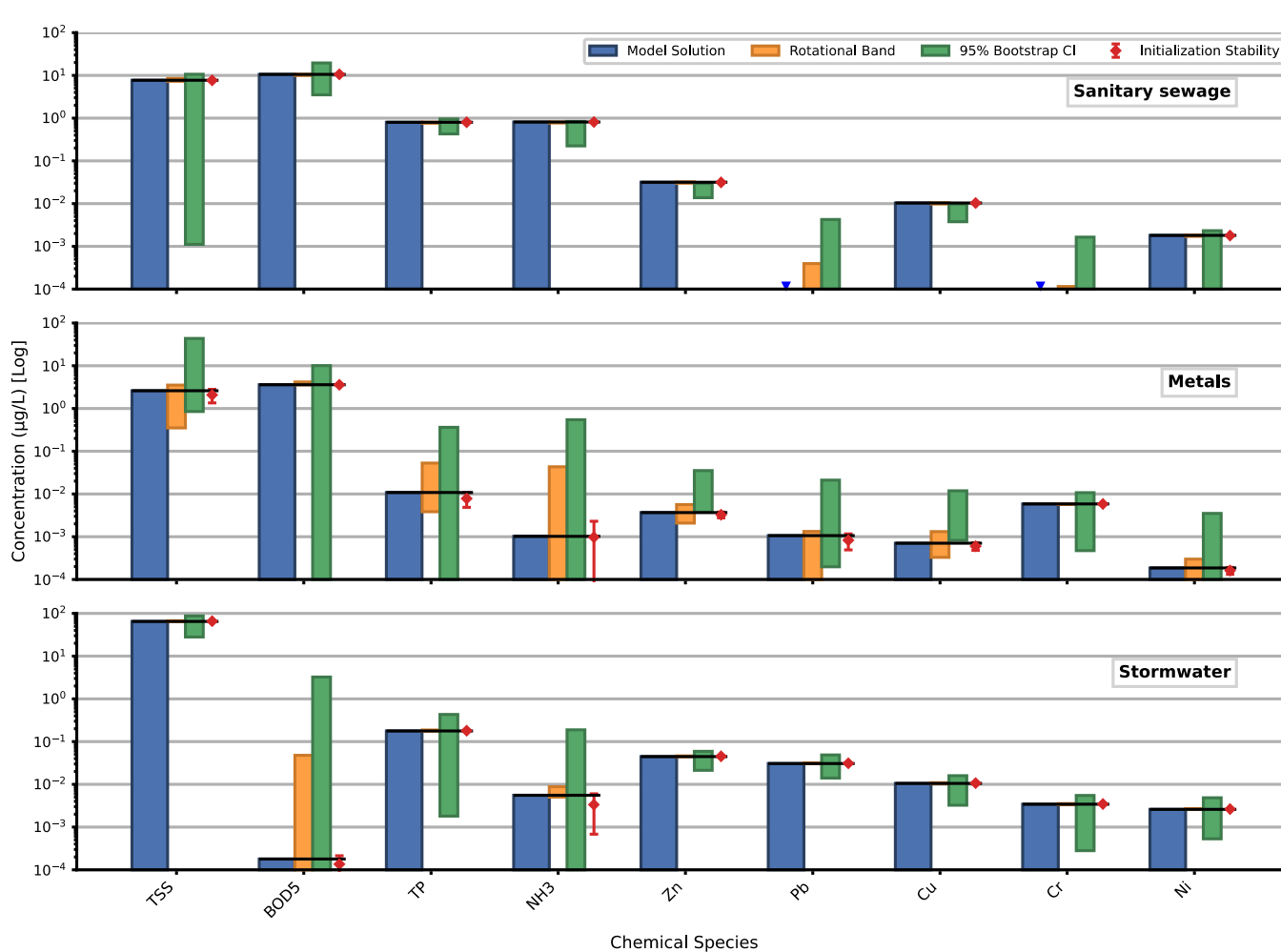
4. Benchmarking & results

SAAE was evaluated against PMF and NMF across three distinct datasets: Milwaukee Water, St. Louis Supersite PM_{2.5}, and Baton Rouge PAMS VOC. Reconstruction quality is reported using Mean Squared Error (MSE; lower is better). Because PMF and NMF benchmarks follow different experimental configurations, MSE values are not comparable between the two baselines.

Resolved source profiles and three-phase validation for the Milwaukee Water dataset are shown on the right.

MSE reconstruction error

Dataset	PMF	SAAE	NMF	SAAE
Milwaukee Water	93.31	94.69	80.67	80.47
St. Louis PM _{2.5}	7.44	7.42	1.13	1.14
Baton Rouge VOC	184.97	181.71	97.99	98.02



5. Future work

Developed as a flexible, data-driven alternative to traditional receptor models, SAAE matches state-of-the-practice performance on identical source apportionment tasks. Future work focuses on two key upgrades: integrating **temporal dynamics** to enable forecasting and suppress unrealistic fluctuations across consecutive contributions, and incorporating **auxiliary features** (e.g., meteorology, sectoral activity data) to better guide training and ensure physically valid solutions.

REFERENCES

- Hopke, P.K., 2016. Review of receptor modeling methods 3. for source apportionment. Journal of the Air & Waste Management Association 66, 237–259. doi:10.1080/10962247.2016.1140693.
- Galvão, E.S., et al., 2020. The role of receptor models as tools for air quality management: a case study of an industrialized urban region. Environ Sci Pollut Res 27, 35918–35929. doi:10. 1007/s11356-020-07848-8.
- Liu, Y., et al., 2025. Source apportionment of PM10 particles in the urban atmosphere using PMF and LPO-XGBoost. Environmental Research, 278, 121659. doi:10.1016/j.envres.2025.121659.
- Oshakbayeva, A., et al., 2026. From receptor models to machine learning: Advancing source apportionment of carbonaceous aerosol. Atmospheric Environment: X 30, 100470. doi:10.1016/j.aeoa.2026.100470.

CONTACT INFORMATION

Aeronautical and Automotive Engineering Department
Loughborough University
Leicestershire LE11 3TU UK
a.parsi@lboro.ac.uk
www.aqdt.net
www.lboro.ac.uk/departments/aae