Improving Performance of Fine-Grained Air Quality Modeling with Domain-Inspired Predictors: A Case Study over Delhi, India

Zeel B Patel[1] , **Nipun Batra**[1] and **Sarath Guttikunda**[2]
[1]IIT Gandhinagar, [2]UrbanEmissions.info
{patel_zeel, nipun.batra}@iitgn.ac.in, simair@urbanemissions.info

Abstract

Air pollution kills 7 million people every year. Nations have deployed air quality sensors but they are costly and thus have spatially sparse coverage. Air quality modeling (interpolation, forecasting) studies have been conducted in the past to increase spatial coverage, but i) physics-driven models do not take sensor readings into account and thus exhibit high bias and ii) machine learning models focus mainly on empirical performance and lack actionable insights, such as, potential reduction in air pollution by removing a source. In this paper, we present an experimental case study on a fine-grained curated dataset for machine learning based $PM_{2.5}$ modeling. We describe the proxies we combine with sensor data for various sources of air pollution including but not limited to population, fire counts and road networks. We show the improvement in modeling performance by including these additional domain-inspired variables; and how these allow us to study what-if scenarios towards mitigation.

1 Introduction

Air pollution takes 7 million lives a year and 9 out of 10 people breathe air that violates WHO standards [WHO, 2024]. Air pollution is considered as one of the major environmental risk to health worldwide by the United Nations. Accordingly, it is addressed for mitigation in multiple United Nations Sustainable Development Goals (UN SDGs), such as the 3.9 and 11.6 [Rafaj *et al.*, 2018].

Countries have ramped up monitoring networks that measure various pollutants, but primarily particulate matter (PM). However, owing to the expensive nature of deploying and maintaining such sensors; and the hyperlocal nature of air pollution, the coverage of sensors in most countries is low.

Researchers have looked at various approaches to increase coverage. The machine learning community has often tackled this problem as spatio-temporal interpolation, akin to filling in missing pixels in an image or video. However, such approaches are often limited if they only consider space and time as features. Thus, researchers often use other related meteorological variables such as: wind speed, etc. The addition of these meteorological features often improves the interpolation performance. However, as we can not control meteorology, such approaches still fail to validate the efficacy of a mitigation strategy.

In contrast, domain experts, start from the first principles: they painstakingly study different sources of pollution (fires, vehicles), and then use physical modelling (condensation, advection, diffusion, etc.) to model the mixing of gases, formation of particles, etc. Despite the soundness of the underlying physics, these approaches often under count emission as it is nearly impossible to account for all sources, and thus exhibit a high bias from the sensor readings measuring air pollution. Secondly, the physical modelling requires a high degree of domain expertise and thus do not scale well or leverage the recent advancement in machine learning.

In this paper, we propose a gray-box modelling framework with the main aim of studying what-if styled mitigation strategies. We incorporate some of the inputs taken by the physical models via proxies: NO_2 can be proxied by the population density (as more population would generally imply more household fuel consumption and more vehicular usage, both increasing NO_2. We discuss with domain experts and identify several such proxies. Finally, we incorporate these proxies in our models and show the improvement in our predictive performance. Importantly, we show that if the NO_2 emissions from human activity is reduced by a factor of 50% (say by improving from low-quality solid fuels still used in rural households to high-quality fuels such as natural gas), then it may lead to a 5% reduction in $PM_{2.5}$.

We summarise our contributions below:

1. We contribute several domain-inspired proxies for various pollutants and their sources.
2. We develop gray-box modelling schemes that leverage data from air pollution sensors, meteorology, and importantly, from various polluting sources.
3. As a result of the above, we show the usage of the above models towards what-if analysis targeted towards understanding the impact of reducing emission from a particular source towards overall air quality in the city.
4. We plan to make the data, model and analysis scripts public at camera ready towards reproducibility and reducing the barrier to entry.

2 Ground-based Sensors Dataset

Delhi, the capital of India is one of the highly polluted cities in the world. Along with the sources, we also discuss about meteorology derived from Weather Research and Forecasting (WRF) model.

2.1 Air Quality Data

Delhi and its surrounding area (known as Delhi, NCR) has 65 continuous ambient air quality monitoring stations (CAAQMS) managed by central pollution control board (CPCB) as shown in Figure 1a. These stations measure various pollutants such as $PM_{2.5}$, PM_{10}, NO_2, and SO_2 among others. It also measures meteorological features such as temperature, pressure, humidity and wind direction and speed. Full list of variables is described in Table 1. Temporal measurement cycle of these stations is 15 minutes but we upscale them to 1 hour to match them with other variables. We have downloaded the entire CPCB data from 2017 to 2023. The increase in number of sensors is shown in Figure 1b. It is important to note that the number of sensors has been nearly stagnant in the recent years. Thus, the remaining coverage needs to be improved via modelling.

Pollution variables	Meteorological variables
$PM_{2.5}$, PM_{10}, NO, NO_2, NOx, NH_3, SO_2, CO, Ozone, Benzene, Toluene, Xylene, Eth-Benzene, MP-Xylene	AT (Air Temperature), RH (Relative Humidity), WS (Wind Speed), WD (Wind Direction), RF (Rainfall), TOT-RF (Total Rainfall), SR (Solar Radiation), BP (Barometric Pressure), VWS (Vertical Wind Shear)

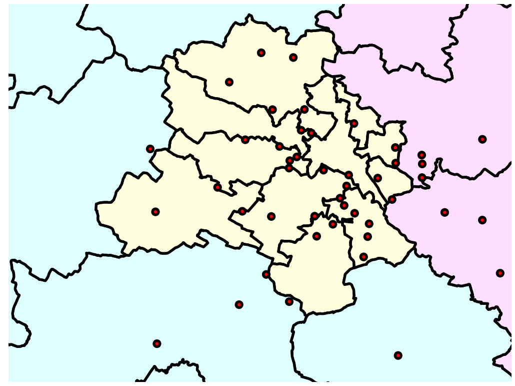
Table 1: Air pollution and weather variables measured by CAAQM stations deployed by central pollution control board (CPCB).

2.2 Correlations among the Pollutants

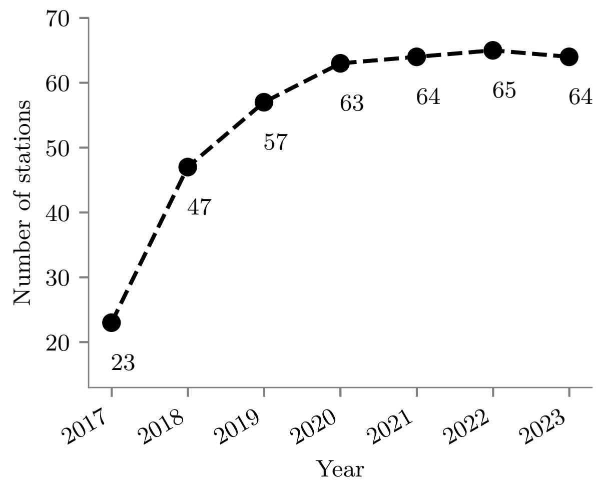
Sources at the current time-stamps may not be the best predictors of the pollution at current time-stamp. We verify this by studying the correlations of major pollutants with PM_{25} at different lags. This behavior suggests that lagged predictors can become useful for spatio-temporal modeling of air pollution.

2.3 Meteorology Data from Weather Research and Forecasting (WRF) Model

We combine the CPCB reported meteorological data with the air quality data but a huge chunk of data is missing in CPCB meteorology. From 65 stations, only 33 stations have sufficient meteorology data. Furthermore, across the world, air quality sensors may not measure meteorology. Thus, we started retrieving the WRF fields starting from 17^{th} July, 2023 from UrbanEmissions.info. These fields are available at 0.01 degree spatial resolution and cover 46 stations around Delhi region. It provides 3D meteorology which can be useful for complex modeling of air quality.



(a) CPCB stations (red dots) in and around Delhi. Left: Haryana, Right: Uttar Pradesh, Center: Delhi.



(b) Evolution of CPCB stations over years around Delhi region.

Figure 1: Locations and the evolution of number of air quality monitors in and around Delhi, India.

3 Spatio-Temporal Modeling

In this section we explore the spatio-temporal modeling of air pollution with meteorological features and analyse their results.

3.1 Attentive Neural Processes

Neural processes are inspired from Gaussian processes [Rasmussen *et al.*, 2006] and have recently got attention for air quality modeling as well [Hu *et al.*, 2023]. However, since goal of this paper is not to compete against the best model but to improve the predictors of air quality with domain-inspired features, we use a reasonably accurate attention based method, attentive neural processes (ANPs) [Kim *et al.*, 2019] for all the experiments. Given the pairs of contexts $(\mathcal{X}_c, \mathcal{Y}_c) = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), ..., (\boldsymbol{x}_n, y_n)\}$, a ANP predicts the conditional mean-field Gaussian distribution as the posterior over a target point $\boldsymbol{x}_*$: $p(y_*|\boldsymbol{x}_*, (\mathcal{X}_c, \mathcal{Y}_c)) = \mathcal{N}(\mu(\boldsymbol{x}_*, (\mathcal{X}_c, \mathcal{Y}_c)), \sigma(\boldsymbol{x}_*, (\mathcal{X}_c, \mathcal{Y}_c)))$. The only key difference between ANPs and CNPs (Conditional Neural Pro-

	NO_2	CO	NH_3	SO_2	NO_x
Lag (hour)					
0	0.362	0.410	0.264	**0.148**	0.399
1	0.383	**0.439**	**0.265**	0.145	**0.421**
2	**0.384**	0.438	0.256	0.127	0.419
3	0.375	0.421	0.244	0.109	0.403
4	0.361	0.397	0.232	0.097	0.380
5	0.345	0.372	0.221	0.088	0.354
6	0.331	0.349	0.212	0.083	0.330

Table 2: Lagged correlations between $PM_{2.5}$ and other pollutants. We can see that most of the times, instantaneous (Lag=0) variable is not having the highest correlation with $PM_{2.5}$ but Lag=1 and/or Lag=2 are having the highest correlations.

cesses) is that CNPs apply a mean aggregation to the representation vector whereas ANPs apply a weighted mean aggregation to the representation vector where weights are the attention weights which can be produced by any suitable method. In this paper, we use cross-attention between context and target inputs to generate the weights.

3.2 Baselines

We compare our results with other well-known spatial interpolators i.e. Kriging, IDW, and a general purpose regressor, Random Forest.

3.3 Experimental setup

We divide the stations into 4-fold cross validation setup. We assume test stations are unmonitored locations and evaluate the metrics on them. We consider the month of Dec 2023 to practically allow ourselves hundreds of ablation experiments on various types of features.

3.4 Evaluation

First, we run the spatial baselines on each of the time-stamp over 4-fold cross validation. As shown in Table 3, Kriging turns out as the best spatial interpolator because it has a good inductive bias. However, we can not use it further since it can not adapt to more features.

We observe that particularly on the last fold, ANP/RF models are not performing well. Upon further investigation, we found that some of the stations are either highly under-predicted or over-predicted by the model as shown in Figure 2a and 2b. This behavior motivates the need to include more predictors (proxies of air pollution sources) to potentially capture such stations. Choice of the predictors can be inspired from domain knowledge. For example,

- Anand Vihar is a bus station and diesel combustion by buses leads to high NO_2 concentrations which can potentially contribute to air pollution.
- Arya nagar, Haryana is relatively less populated region and thus likely to have fewer human activities such as traffic, contributing to air pollution.

Thus, in the next section, we discuss Domain-Inspired predictors.

Model	Fold 0	Fold 1	Fold 2	Fold 3	Mean
Kriging	53.80	49.57	49.04	71.17	55.89
IDW	55.16	48.77	51.80	73.01	57.18
RF-M	55.09	51.07	52.55	75.62	58.58
RF	53.66	50.62	54.73	78.22	59.31
ANP-M	55.25	54.27	54.86	71.65	59.01
ANP	57.08	52.45	50.19	88.15	61.97
Mean	58.36	59.19	59.13	73.10	62.45
1NN	73.21	60.23	65.47	81.56	70.12

Table 3: Baselines fitted only on the spatial co-ordinates (latitude and longitude). Kriging is performing the best due to good inductive bias however it can not be used for what-if analysis since it is unable to incorporate more features. RF-M and ANP-M additionally use meteorological features and get the improvement from it.

4 Domain-Inspired Predictors

In the previous section, we touched upon the role of features such as population density and diesel combustion in prediction of air pollution. Some of the features can be approximated by large-scale proxies provided by various agencies. We will look into them in the further subsections. **We specifically focus on this section to also discuss the what-if scenarios such as X% mitigation in a particular source can result in Y% reduction in air pollution**.

4.1 Human Activities

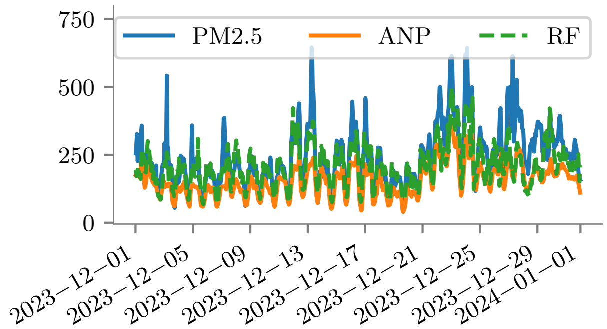
High population in a region naturally implies activities such as transport, heating, cooking, industries etc. [Guttikunda *et al.*, 2023]. Such activities can contribute to multiple sources of air pollution such as NO2, CO and BC. Mitigation in this context would require actions such as better traffic management and use of less polluting household fuels.

Proxy: 'Landscan' population dataset

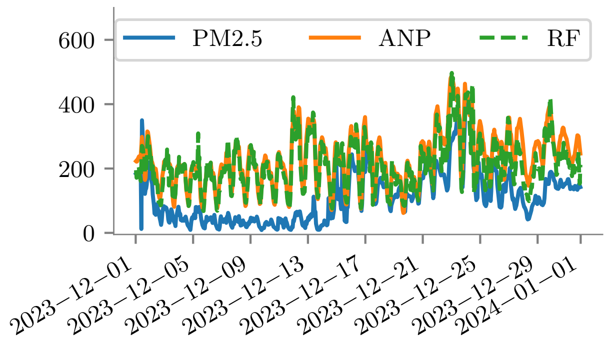
Landscan is an award winning high-resolution global population dataset produced every year by Oak Ridge National Laboratory [Sims *et al.*, 2023]. The dataset is available at approximately 0.008 degree resolution across the world, which means we have approximately 14.4k points within our study region. We create features based on number of people within X kilometer radius of a station, where $X \in \{1, 2, 3, 4, 5, 10\}$. To check the effect of these features, we fit both ANP and RF on the created features along with the meteorology data. Table 4 shows the results from our experiments. For ANP, population within 10 km is the best feature whereas for RF, population within 1 km is the best feature. **Noticably, ANP's RMSE for the Arya Nagar, Haryana station, which is less populated, reduced from 140 (ANP-M) to 78 (ANP-M-P10) signifying the importance of this feature.**

What-if: Impact of reducing emission via human activity on PM

We now study the impact of reducing the emission via human activity. This could be done via several schemes already recommended by experts, such as transitioning from low-grade



(a) A station near Anand Vihar bus station which exhibits high pollution but models finds it hard to capture.



(b) A station near Arya nagar, Haryana which is generally cleaner but models predict it as polluted.

Figure 2

Model	Fold 0	Fold 1	Fold 2	Fold 3	Mean
RF-M-P1	54.55	50.50	51.13	72.02	**57.05**
RF-M-P2	54.29	50.42	51.99	72.96	57.42
RF-M-P3	54.50	51.18	51.92	72.66	57.57
RF-M-P4	54.46	52.29	52.23	72.50	57.87
RF-M-P5	54.90	51.66	52.17	71.14	57.47
RF-M-P10	55.75	51.17	52.28	74.08	58.32
RF-M	55.09	51.07	52.55	75.62	58.58
ANP-M-P1	56.53	56.69	52.62	68.73	58.64
ANP-M-P2	58.16	60.47	56.41	76.58	62.91
ANP-M-P3	57.46	65.12	58.70	67.60	62.22
ANP-M-P4	56.17	61.89	61.70	69.12	62.22
ANP-M-P5	55.96	61.01	59.00	64.95	60.23
ANP-M-P10	58.20	52.76	51.32	68.66	**57.74**
ANP-M	55.25	54.27	54.86	71.65	59.01

Table 4: Results from RF and ANP when fitted on population data along with meteorological data. Both models have a noticable reduction in RMSE after incorporating population features. Additionally, ANP's RMSE for Arya Nagar, Haryana station, which is less populated, reduced from 140 (ANP-M) to 78 (ANP-M-P10) signifying the importance of this feature. P1: population within 1 km, P2: population within 2 km and so on.

% reduction in source	% reduction in $PM_{2.5}$
10%	0.6%
30%	2.77%
50%	5.03%

Table 5: Percentage reduction in $PM_{2.5}$ by simulating reduction in emissions via efficient human activity.

household fuels such as wood to higher-grade fuels such as natural gas. We assume that if by certain actions, if we reduce the emission in human activity by $x\%$, then how much we can reduce the PM. To do such an analysis, we feed in the Landscan population proxy feature 'reduced' by $x\%$ to our models and we look at the corresponding reduction in PM. We study these results in Table 5 and find that by reducing the emissions by 50% we may be able to decrease PM by around 5%. These numbers are comparable to purpose-built studies by domain experts [Guttikunda *et al.*, 2023].

4.2 Open Waste Burning (OWB)

Open Water Burning is one of the most underrepresented source of air pollution in pollution inventories [Guttikunda *et al.*, 2023]. Incomplete combustion generally releases CO in the environment which then forms OC and BC constituents of $PM_{2.5}$. In the following section, we present a small case study of Ghazipur landfill in Delhi:

The Trash Mountain of Delhi: Ghazipur landfill

Ghazipur landfill in Delhi is around 29 hectares sized landfill area of East Delhi. Around 2200 MT of municipal solid waste is dumped daily at this landfill [Babbar *et al.*, 2017]. It contributes heavily to the health [Yadav and Negi, 2023], water quality [Babbar *et al.*, 2017] and air quality [Guttikunda and Calori, 2013]. It releases toxic gases in air including methane which is flammable and can lead to open fire [Ranjan *et al.*, 2014] resulting in incomplete combustion and generation of CO. We verified this by analysing the time-series data around the same time-frame from three nearest stations. As shown in Figure 4b, we see big spikes in CO levels in the nearby stations during the burning period.

On 28^{th} March, 2022, Ghazipur landfill caught an open fire (see Figure 4a) which lasted over 48 hours and was reported by multiple media houses [NDTV, 2022; IndiaToday, 2022; HindustanTimes, 2022b]. More fire incidents from the same site were reported on 9^{th} April, 2022 and 20^{th} April, 2022 [HindustanTimes, 2022a].

Studies have estimated the contribution of such open waste burning from 5 to 15% [Guttikunda *et al.*, 2023]. Thus, it is important to consider open waste burning and their indicators in air quality modeling.

Proxy: Fire counts

NASA's Fire Information for Resource Management System (FIRMS) publishes Visible Infrared Imaging Radiometer

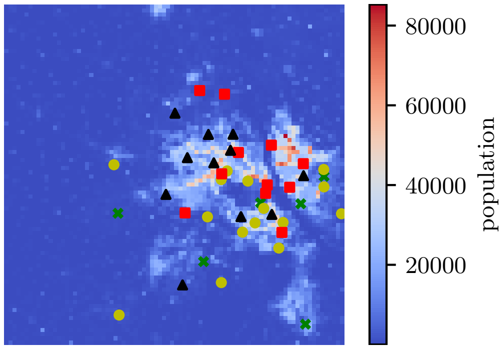


Figure 3: Landscan population distribution over Delhi along with Normalized pollution (0 to 100) from CPCB stations. We can see that highly populated areas are likely to have high air pollution while the converse may not be true due to other sources of air pollution. Black triangles: $\%PM_{2.5} > 75$; Red squares: $75 > \%PM_{2.5} > 50$; yellow circles: $50 > \%PM_{2.5} > 25$; green crosses: $25 > \%PM_{2.5}$

Suite (VIIRS) dataset to monitor active fire counts data across the world at 375 m resolution. We retrieved the counts data from VIIRS and merged it with our dataset in form of 'fire counts' and 'fire brightness'. Table 6 shows the results for RF and ANP. We did not see any improvement in results after adding these features. This could be attributed to very less number of fire activities during the experiment time-period of data.

Model	Fold 0	Fold 1	Fold 2	Fold 3	Mean
RF-M-P1	54.55	50.50	51.13	72.02	57.05
RF-M-P10	55.75	51.17	52.28	74.08	58.32
RF-M-P10f	55.75	51.16	52.28	74.13	58.33
RF-M	55.09	51.07	52.55	75.62	58.58
ANP-M-P1	56.53	56.69	52.62	68.73	58.64
ANP-M-P10	58.20	52.76	51.32	68.66	57.74
ANP-M-P10f	56.96	53.21	51.23	74.41	58.95
ANP-M	55.25	54.27	54.86	71.65	59.01

Table 6: Results from RF and ANP when fitted on population, fire count and meteorological data. We noticed that these features did not improve the metrics. We suspect the sparsity of fire counts to be responsible for this.

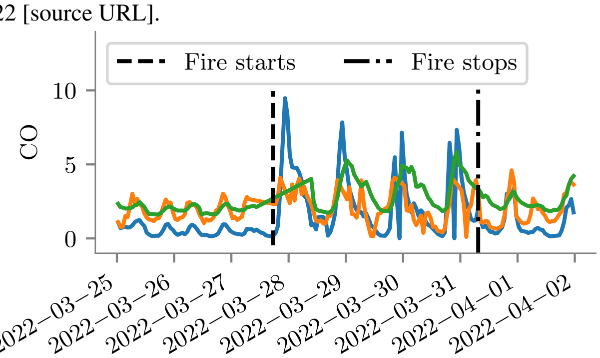
4.3 Road Networks

Vehicular emissions release various gases including but not limited to NOx, NO_2 and CO. Diesel driven vehicles are among the most polluting vehicles [Guttikunda and Mohan, 2014].

Proxy: distance from road networks

Road networks of Delhi were mapped and converted to geo-coded dataset under the NCAP program[1]. We select a category called 'motorway' which is a major circular road around

(a) Ghazipur landfill was on fire from 28^{th} March to 31^{st} March 2022 [source URL].



(b) Due to incomplete combustion in Ghazipur fire event, CO has spiked up in the closest three monitoring stations during the burning period.

Figure 4

Delhi, where a significant proportion of vehicles ply on. We take the distance from this road as a feature. Our hypothesis is that the points closer to these roads will pick up more emission from the vehicles and thus be more polluted. All other things constant, we would expect the pollution from vehicles to decrease as we move away from such roads. We show the results from our experiments incorporating these features in Table 7. We can note that incorporating the road information helps reduce the RMSE in our predictions.

5 Limitations & Future work

We now discuss the limitations and future work.

1. In the current work, we have looked at attentive neural processes. These models are not purpose built for explainability/additivity. In the future, we propose to leverage variants such as neural additive models [Agarwal *et al.*, 2021], that inherently are additive in nature, and thus would make them highly suitable for modelling what-if scenarios.
2. In the current work, we took the population and road network features, both of which are proxies for number of vehicles. In the future, we plan to look at more direct/correlated proxies for the number of vehicles. Google Traffic API could be one such source, but, it does

[1] https://prana.cpcb.gov.in/#/home

Model	Fold 0	Fold 1	Fold 2	Fold 3	Mean
RF-M-P1	54.55	50.50	51.13	72.02	57.05
RF-M-P10	55.75	51.17	52.28	74.08	58.32
RF-M-P10r	53.71	50.75	52.25	69.84	**56.64**
RF-M	55.09	51.07	52.55	75.62	58.58
ANP-M-P1	56.53	56.69	52.62	68.73	58.64
ANP-M-P10	58.20	52.76	51.32	68.66	57.74
ANP-M-P10r	54.49	54.85	55.20	62.78	**56.83**
ANP-M	55.25	54.27	54.86	71.65	59.01

Table 7: Results from RF and ANP when fitted on population (**P10**), distance from road (**r**) and meteorological data (**M**). We see an improvement in results after including this feature.

not give the vehicle count, but rather the congestion level (1, 2, 3, 4).

3. In the current work, our models are relatively simple; the focus was more on the inputs. In the future, we plan to take inspiration from the physical models that look at the mixing of gases, movement of pollutants and model them via diffusion like methods.

6 Conclusion

Air pollution is an important problem impacting a significant proportion of the population. Existing studies using machine learning models have largely focused on the interpolation problem: estimating air pollution at places where we do not have sensors. In this paper, we argue such approaches have limited benefit in understanding and developing mitigation strategies. In contrast, we propose modelling with both air quality sensors as well as sources (via proxies). We are able to show this helps in improving the predictive performance of the interpolators; and more importantly helps us understand the potential benefits of controlling pollution via particular sources.

References

[Agarwal *et al.*, 2021] Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. *Advances in neural information processing systems*, 34:4699–4711, 2021.

[Babbar *et al.*, 2017] Puneet Babbar, Swatantra Verma, and Gauhar Mehmood. Groundwater contamination from non-sanitary landfill sites–a case study on the ghazipur landfill site, delhi (india). *International Journal of Applied Environmental Sciences*, 12(11):1969–1991, 2017.

[Guttikunda and Calori, 2013] Sarath K Guttikunda and Giuseppe Calori. A gis based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in delhi, india. *Atmospheric Environment*, 67:101–111, 2013.

[Guttikunda and Mohan, 2014] Sarath K Guttikunda and Dinesh Mohan. Re-fueling road transport for better air quality in india. *Energy Policy*, 68:556–561, 2014.

[Guttikunda *et al.*, 2023] Sarath K Guttikunda, Sai Krishna Dammalapati, Gautam Pradhan, Bhargav Krishna, Hiren T Jethva, and Puja Jawahar. What is polluting delhi's air? a review from 1990 to 2022. *Sustainability*, 15(5):4209, 2023.

[HindustanTimes, 2022a] HindustanTimes. Ghazipur landfill burns for 3rd time in a month — hindustantimes.com. https://www.hindustantimes.com/cities/delhi-news/ghazipur-landfill-burns-for-3rd-time-in-a-month-101650492209448.html, 2022. [Accessed 22-02-2024].

[HindustanTimes, 2022b] HindustanTimes. Highlights: Ghazipur landfill fire in Delhi continues to rage after 10 hours — hindustantimes.com. https://www.hindustantimes.com/india-news/breaking-news-today-march-28-2022-101648429099578.html, 2022. [Accessed 22-02-2024].

[Hu *et al.*, 2023] Junfeng Hu, Yuxuan Liang, Zhencheng Fan, Hongyang Chen, Yu Zheng, and Roger Zimmermann. Graph neural processes for spatio-temporal extrapolation. *arXiv preprint arXiv:2305.18719*, 2023.

[IndiaToday, 2022] IndiaToday. Massive blaze breaks out at Delhi's Ghazipur landfill site; fire tenders rushed to spot — indiatoday.in. https://www.indiatoday.in/cities/delhi/story/delhi-ghazipur-garbage-landfill-site-fire-pictures-video-1930495-2022-2022. [Accessed 22-02-2024].

[Kim *et al.*, 2019] Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. *arXiv preprint arXiv:1901.05761*, 2019.

[NDTV, 2022] NDTV. 48 Hours On, Fire At Delhi's Ghazipur Landfill Still Not Under Control — ndtv.com. https://www.ndtv.com/india-news/delhi-ghazipur-fire-48-hours-on-fire-at-delhis-ghazipur-landfill-still-no 2022. [Accessed 22-02-2024].

[Rafaj *et al.*, 2018] Peter Rafaj, Gregor Kiesewetter, Timur Gül, Wolfgang Schöpp, Janusz Cofala, Zbigniew Klimont, Pallav Purohit, Chris Heyes, Markus Amann, Jens Borken-Kleefeld, et al. Outlook for clean air in the context of sustainable development goals. *Global Environmental Change*, 53:1–11, 2018.

[Ranjan *et al.*, 2014] Manju Rawat Ranjan, AL Ramanathan, Ashutosh Tripathi, and Pawan Kumar Jha. Landfill mining: a case study from ghazipur landfill area of delhi. *International Journal of Environmental Sciences*, 4(5):919–925, 2014.

[Rasmussen *et al.*, 2006] Carl Edward Rasmussen, Christopher KI Williams, et al. *Gaussian processes for machine learning*, volume 1. Springer, 2006.

[Sims *et al.*, 2023] Kelly Sims, Andrew Reith, Edward Bright, Jason Kaufman, Joe Pyle, Justin Epting, Jack Gonzales, Daniel Adams, Eric Powell, Marie Urban, and Amy Rose. Landscan global 2022, 2023. CY 2022.

[WHO, 2024] WHO. Air pollution — who.int. https://www.who.int/health-topics/air-pollution#tab=tab_2, 2024. [Accessed 22-02-2024].

[Yadav and Negi, 2023] Shilpi Yadav and Mahabir S Negi. The impacts on environment and health of residents: A case study of ghazipur dumpsite (landfill), east delhi. In *Population, Sanitation and Health: A Geographical Study Towards Sustainability*, pages 125–135. Springer, 2023.