DELFI: Deep Mixture Models for Long-term Air Quality Forecasting in the Delhi National Capital Region

Naishadh Parmar 1  Raunak Shah 1  Tushar Goswamy 1  Vatsalya Tandon 1  Ravi Sahu 1  Ronak Sutaria 2  Purushottam Kar 1  Sachchida Nand Tripathi 1

Abstract

The identification and control of human factors in climate change is a rapidly growing concern and robust, real-time air-quality monitoring and forecasting plays a critical role in allowing effective policy formulation and implementation. This paper presents DELFI, a novel deep learning-based mixture model to make effective long-term predictions of Particulate Matter (PM) 2.5 concentrations. A key novelty in DELFI is its multi-scale approach to the forecasting problem. The observation that point predictions are more suitable in the short-term and probabilistic predictions in the long-term allows accurate predictions to be made as much as 24 hours in advance. DELFI incorporates meteorological data as well as pollutant-based features to ensure a robust model that is divided into two parts: (i) a stack of three Long Short-Term Memory (LSTM) networks that perform differential modelling of the same window of past data, and (ii) a fully-connected layer enabling attention to each of the components. Experimental evaluation based on deployment of 13 stations in the Delhi National Capital Region (Delhi-NCR) in India establishes that DELFI offers far superior predictions especially in the long-term as compared to even non-parametric baselines. The Delhi-NCR recorded the 3rd highest PM levels amongst 39 mega-cities across the world during 2011-2015 and DELFI’s performance establishes it as a potential tool for effective long-term forecasting of PM levels to enable public health management and environment protection.

1. Introduction

The global challenge of climate change demands a multi-faceted response. Rapid, robust, and real-time identification of human sources of climate change such as combustion, mining, and other activities is a key aspect in enabling agile and adaptive policy and regulatory decisions. Climate change positively correlates with air pollution levels, with air pollutants such as black carbon that constitute particulate matter (PM), methane, tropospheric ozone, and aerosols, also affecting the amount of incoming sunlight and contributing to global temperature rise and glacial degradation (Mantisidis et al., 2020). For instance, fossil fuel and biomass combustion are the biggest source of black carbon aerosols that contribute to both PM levels as well as accelerating glacier melting in the Himalayas (Patella et al., 2018).

In particular, PM2.5 refers to particulate matter with a diameter less than 2.5 \( \mu m \) and includes combustion by-products, metals and organic materials. PM2.5 is classified as an atmospheric pollutant of high concern since owing to its small size and comparatively larger surface area, it can remain suspended for extended periods, be easily transported and infiltrate the pulmonary and circulatory systems, if inhaled. Chronic exposure to high PM2.5 levels has been linked to, heart attacks and strokes (PopeIII, 2002), and respiratory diseases such as lung cancer (Kampa and Castanas, 2008). Maternal exposure to high PM2.5 levels elevates the risk of congenital heart defects in infants (Zhang et al., 2016). Effective monitoring and regulatory control of PM2.5 levels presents the need for a robust long-term forecasting model for PM2.5, especially in high pollution regions such as the Delhi National Capital Region (NCR) in India where extremely high PM2.5 levels (484 on a scale on 500) led to the local government triggering a state of public health emergency on November 1, 2019 (Dasgupta, 2019).

Technical Contributions and Impact: The primary technical contribution of this paper is DELFI, a novel deep learning-based mixture model to perform long-term predictions of PM2.5 levels with key technical contributions:

1. Adopting a novel novel multi-scale forecasting strategy employing probabilistic predictions for forecasts with horizons longer than 6 hours
2. A light-weight technique employing pre-computed NEF features (see Sec 2) that allow spatio-temporal effects to be incorporated without additional expense
DELFI: Deep Mixture Models for Long-term AQ Forecasting in the Delhi National Capital Region

3. A mixture model architecture with 3 components each comprised of a stack of LSTM networks with attention and an alternating optimization-based training strategy.

4. Forecasts for time horizons as large as 24 hours into the future that offer significantly improved accuracy compared to baseline methods while offering short-term (1-4 hour) predictions within sensor error levels.

DELFI makes point short-term predictions (for up to 6 hours in the future) and probabilistic long-term predictions (for up to 48 hours in the future). Experiments indicate that the model can be used to reliably design long-term air quality forecasting and early-warning systems that enable rapid regulatory action of preventative nature. Given the close link between air-quality and the drivers of climate change, this can not only safeguard the health of citizens against the harmful effects of air pollution, but also mitigate the adverse impact of human activities on climate change.

Related Works and Contributions of DELFI in Context:
At a high level, most existing works use a monolithic Long Short Term Memory (LSTM)-based model to make short-term predictions, often next-hour but up to 6 hours. In contrast, DELFI offers accurate probabilistic predictions up to 24 hours in advance that make it suitable for use in designing early warning systems. Bansal et al. (2019) uses data on CO, NO2, NO, Ozone, PM2.5 and SO2 levels from high-grade monitors to perform next hour prediction of the same levels using LSTMs. In contrast, DELFI uses PM1, PM10 and PM2.5 concentrations from low-cost sensors in addition to meteorological features to offer predictions over much longer time scales as well as MAE values that are 35% lower. Chaudhary et al. (2018) use LSTMs to make next-hour predictions but use additional sources of seasonal data such as holidays and traffic data to which DELFI does not assume access. Kumar and Goyal (2011) apply principal component regression to make short-term predictions at a single air quality monitoring station. In contrast, DELFI ingests data from, and makes simultaneous predictions at other stations given wind direction and velocities. Ensemble models have been also considered such as (Liu et al., 2019) which incorporates 5 models but only offer predictions up to 6 hours in advance, and (Bai et al., 2019) that uses an ensemble LSTM network to make next-hour predictions. In contrast, DELFI makes use of fewer but diverse components in its mixture, and does a careful assignment of data used to train each component, finally using an aggregator network to assign attention weights to each component.

2. DELFI: Deep Mixture models for Long-term air-quality Forecasting

Data: DELFI is trained on air pollution as well as meteorological data from 13 stations in the region of Delhi-NCR, India, latitudes 28° 27’ 15” to 28° 38’ 40” and longitudes 77° 4’ 26” to 77° 19’ 40”. Data collection was done at intervals of 1 hour in the time period of 1 November 2018 00:00:00 to 28 March 2019 23:00:00. PM2.5 concentrations that appeared extremely elevated were not removed as outliers from the data so as to enable the model to be trained and tested on predicting spikes in PM2.5 levels.

Training Features: DELFI uses a total of 9 features, each available as a time-series at each station, to perform forecasting: (i) PM1, (ii) PM10, (iii) PM2.5, (iv) Temperature, (v) Humidity, (vi) Visibility, (vii) Wind speed, (viii) Wind direction and (ix) the Net external flow (NEF). Of these, the first 8 features are standard and available from the monitoring stations or else standard APIs. However, the last NEF feature (described below) was engineered to allow DELFI to take spatio-temporal effects of long-range air flow into account. A data-point \( w_{i,t} \) is defined as a six-hour window (ending at timestamp \( t \)) of these pollutant and meteorological features taken at one-hour intervals at station \( i \). The true PM2.5 concentration value at station \( i \) at timestamp \( t \) will be denoted by \( P M_{i,t} \).

The NEF Feature: This feature attempts to capture in a relatively inexpensive manner, the effect of PM2.5 concentrations at other stations given wind direction and velocities. For a station \( a \) and time \( t \), this feature is defined as:

\[
NEF_{a,t} = \frac{1}{1 + e^{-x}} = \frac{1}{1 + e^{-\sum_i PM_{i,t} \times V_{i,t} \times \cos(\theta_a - \phi_{i,t})}},
\]

where \( PM_{i,t}, V_{i,t} \) and \( \phi_{i,t} \) are respectively the true PM2.5 concentration, wind speed and wind bearing at station \( i \) at time \( t \) and \( \theta_a \) is the bearing between station \( a \) and station \( i \) based on the World Geodetic System (WGS84).

A Key to Long-term Forecasting: As discussed in Section 1, most existing work attempts to make short term forecasts (up-to 6 hours). Long term forecasting rapidly deteriorates in quality possibly due to the lack of complete knowledge of all factors affecting air quality as well as the chaotic nature of aerodynamic systems. However, DELFI makes an observation that when making longer-term forecasts, say 24 hours in advance, a point prediction becomes less critical. Specifically, if making a 24 hour forecast at 1100hrs today, it is not critical to know the PM2.5 levels at exactly 1100hrs tomorrow. Rather, the distribution of PM2.5 levels around 1100hrs (e.g. in the period 0900-1300 hrs, are the levels likely to remain mild or can they spike) becomes more important both from the point of policy formulation as well as modulating personal behavior. To exploit this observation, DELFI uses air-quality categorization set forth by regulatory authorities in India, namely 0-30 (Good), 30-60 (Satisfactory), 60-90 (Moderately polluted), 90-120 (Poor), 120-250 (Very poor) and 250+ (Severe) to create 6 bins. Thus, when making long-term forecasts, say in the above example, DELFI predicts a discrete probability
were prominent PM2.5 hotspot candidates whereas others
short-term predictions (see Figure A in the appendix), each
used as the aggregator one for long-term predictions. A fully connected layer is
distinct mixture models: one for short-term predictions and
input and offers a sequential output. DELFI develops two
stacked LSTMs which take in features
components, each being a stack of LSTM networks with attention
that a single model may struggle to address these extremes.
However, given the diversity of the stations (some
components before starting the main training step. As a
decide the data points used to pre-train the three individual
component best suited to make predictions for its data. These large-capacity archi-
tectures allow the 13 diverse stations to adaptively choose a
component best suited to make predictions for its data.

**Model and architecture:** Initial experiments were con-
ducted with 13 separate LSTM-based model being trained on
data from each of the 13 stations in the deployment. However, this strategy neither took advantage of the much larger amount of overall data, nor did models learnt for one station do well in predicting values for another station. However, given the diversity of the stations (some were prominent PM2.5 hotspot candidates whereas others reported much milder PM2.5 values), it was also expected that a single model may struggle to address these extremes. DELFI’s solution is mixture model that consults 3 com-
ponents, each being a stack of LSTM networks with attention
weights being learnt for each component (see Figure 1).
More specifically, each component consists of a series of
stacked LSTMs which take in features $w_{i,t}$ as sequential input and offers a sequential output. DELFI develops two
distinct mixture models: one for short-term predictions and
one for long-term predictions. A fully connected layer is
used as the aggregator that offers the attention weights. For
short-term predictions (see Figure A in the appendix), each
component predicts a certain residual value, for example
$ΔPM_{i,t+1}$ for next-hour predictions and weights are as-
signed to the output of each component. For long-term
predictions, the output is used to scale the concatenated
sequences and passed through another fully connected layer
to get the final histogram output. These large-capacity archi-
tectures allow the 13 diverse stations to adaptively choose a
component best suited to make predictions for its data.

**Pre-training:** The variance in the residual values
$ΔPM_{i,t+1}$ was computed for all stations. Using this statistic, stations were grouped into three categories. Each group was assigned one component in the mixture and data from within the group was used to pre-train model parameters for that particular component. This division was used to
decide the data points used to pre-train the three individual
components before starting the main training step. As a
result, stations prone to spiking PM2.5 values were likely
to cluster together into one group, with other stations with
more gentle variations in PM2.5 levels falling into another.

**Training:** An alternating optimization procedure reminis-
cent of the EM algorithm was adopted to train both the fully
connected layer (aggregator) and individual components of
the mixture described in Algorithm 1. The mixture model

![Figure 1. Model architecture used by DELFI for long-term probabilistic predictions. $B$ is batch-size. The blue and orange portions of the network are trained alternately as described in Algorithm 1.](image-url)
Table 2. KL divergence between predicted and actual histograms for various horizon lengths. Linear models were unable to provide meaningful predictions and were excluded from comparison. DELFI offers KL divergence values that are at least 42% and up to 76% smaller than those offered by the KNN algorithm.

| Time (hrs) | KNN | DELFI |
|-----------|-----|-------|
| 1±3       | 1.33| 0.32  |
| 8±4       | 1.80| 0.89  |
| 12±6      | 2.40| 1.38  |
| 24±12     | 1.40| 0.59  |
| 48±24     | 1.17| 0.53  |

4. Discussion and Impact on Climate Change

This paper presents DELFI, a novel algorithm that introduces several technical innovations such as the use of probabilistic predictions for long-term forecasting and using a mixture model trained using an alternating strategy. Intuitively, a low bias model is preferred for short-term predictions to encode local features properly that explains KNN’s good performance on small horizons (1-4 hrs). However, for long-term predictions, as unpredictability of the system grows, a low-variance method is preferable instead that KNN does not offer. DELFI seems to offer a suitable balance between bias and variance allowing it to perform well in both regimes. Experimental results suggest that DELFI
offers predictions reliable enough to design early-warning systems based on long-term forecasts. In highly polluted regions of the globe such as the Delhi-NCR, such systems offer citizens a chance to modulate their own personal behavior e.g. avoiding outdoor activities, but can also enable regulatory authorities to take immediate preventative action as well as effect long-term policy shift. Given the close link between air-quality and the drivers of climate change discussed in Section 1, this can not only safeguard the health of citizens but also mitigate the adverse impact of human activities on climate change in the longer term.

References

Yun Bai, Bo Zeng, Chuan Li, and Jin Zhang. An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting. *Chemosphere*, 222:286–294, May 2019. doi: 10.1016/j.chemosphere.2019.01.121. URL https://doi.org/10.1016/j.chemosphere.2019.01.121.

Mohit Bansal, Anirudh Aggarwal, Tanishq Verma, and Apoorvi Sood. Air quality index prediction of delhi using lstm, 11 2019.

Vidushi Chaudhary, Anand Deshbhratar, Vijayanand Kumar, and Dibyendu Paul. Time series based lstm model to predict air pollutant’s concentration for prominent cities in india. 2018.

Neha Dasgupta. New delhi declares public emergency as pollution at year’s worst, 2019. URL https://www.reuters.com/article/us-india-pollution/new-delhi-declares-public-emergency-as-pollution-at-years-worst-idUSKBN1XB3S1.

Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental Pollution*, 151(2):362–367, January 2008. doi: 10.1016/j.envpol.2007.06.012. URL https://doi.org/10.1016/j.envpol.2007.06.012.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

Anikender Kumar and Piyush Goyal. Forecasting of air quality in delhi using principal component regression technique. *Atmospheric Pollution Research*, 2:436–444, 10 2011. doi: 10.5094/APR.2011.050.

Hui Liu, Zhu Duan, and Chao Chen. A hybrid framework for forecasting PM2.5 concentrations using multi-step deterministic and probabilistic strategy. *Air Quality, Atmosphere & Health*, 12(7):785–795, April 2019.

doi: 10.1007/s11869-019-00695-8. URL https://doi.org/10.1007/s11869-019-00695-8.

Ioannis Manisalidis, Elisavet Stavropoulou, Agathangelos Stavropoulos, and Eugenia Bezirtzoglou. Environmental and health impacts of air pollution: A review. *Frontiers in Public Health*, 8, February 2020. doi: 10.3389/fpubh.2020.00014. URL https://doi.org/10.3389/fpubh.2020.00014.

Vincenzo Patella, Giovanni Florio, Diamira Magliacane, Ada Giuliano, Maria Crivellaro, Daniela Bartolomeo, Arturo Genovese, Mario Palmieri, Amedeo Postiglione, Erminia Ridolo, Cristina Scaletti, Maria Ventura, and Anna Zollo. Urban air pollution and climate change: “the decalogue: Allergy safe tree” for allergic and respiratory diseases care. *Clinical and Molecular Allergy*, 16, 09 2018. doi: 10.1186/s12948-018-0098-3.

C. Arden PopeIII. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, 287(9):1132, March 2002. doi: 10.1001/jama.287.9.1132. URL https://doi.org/10.1001/jama.287.9.1132.

Ravi Sahu, Ayush Nagal, Kuldeep Dixit, Harshavardhan Unnibhavi, Srikanth Mantravadi, Srijith Nair, Yogesh Simmhan, Brijesh Mishra, Rajesh Zele, Ronak Sutaria, Purushottam Kar, and Sachchida Tripathi. Robust statistical calibration and characterization of portable low-cost air quality monitoring sensors to quantify real-time o3 and no2 concentrations in diverse environments, 06 2020.

Bin Zhang, Shengwen Liang, Jinzhu Zhao, Zhengmin Qian, Bryan A Bassig, Rong Yang, Yiming Zhang, Ke Hu, Shunqing Xu, Tongzhang Zheng, and Shaoping Yang. Maternal exposure to air pollutant PM2.5 and PM10 during pregnancy and risk of congenital heart defects. *Journal of Exposure Science & Environmental Epidemiology*, 26(4):422–427, February 2016. doi: 10.1038/jes.2016.1. URL https://doi.org/10.1038/jes.2016.1.

Yu Zheng, Xiwen Yi, Ming Li, Ruiyuan Li, Zhanqing Shan, Eric Chang, and Tianrui Li. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th SIGKDD conference on Knowledge Discovery and Data Mining*. KDD 2015, August 2015. URL https://www.microsoft.com/en-us/research/publication/forecasting-fine-grained-air-quality-based-on-big-data/.
A. Appendix

Figure 2. Model architecture used by DELFI for short-term point predictions. $B$ is batch-size. The blue and orange portions of the network are trained alternately as described in Algorithm 1.