

# Spatio-Temporal Pollution Forecasting using Hybrid Networks

**Pritthijit Nath**

Jadavpur University

**Pratik Saha**

SRM University Department of Nursing: SRM College of Nursing

**Asif Iqbal Middya**

Jadavpur University

**Sarbani Roy** (✉ [sarbani.roy@jadavpuruniversity.in](mailto:sarbani.roy@jadavpuruniversity.in))

Jadavpur University <https://orcid.org/0000-0002-7598-8266>

---

## Research Article

**Keywords:** Spatio-Temporal Modeling, Hybrid Networks, Stacked Ensemble, Deep Learning, Air Pollution

**Posted Date:** November 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-863931/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Spatio-Temporal Pollution Forecasting using Hybrid Networks

Pritthijit Nath · Pratik Saha · Asif Iqbal Middy · Sarbani Roy\*

Received: Date / Accepted: Date

**Abstract** Rising real estate prices to set up infrastructure coupled with expensive maintenance costs and lack of spares during times of instrument failure have become major issues for statutory bodies when dealing with real-time pollution monitoring stations. As a possible solution to these problems, a novel class of hybrid spatio-temporal pollution forecasting networks which are a combination of various widely used temporal forecasting methods and spatial interpolation methods have been proposed in this paper. In addition, a novel multi-site Multi Layer Perceptron based Ensemble method, capable of improving accuracy by taking exogenous variables into account, has also been proposed as a spatial interpolation method in addition to popular techniques such as Ordinary Kriging (OK) and Inverse Distance Weighted (IDW). Experimental results based on the multi-site air pollution data of Beijing suggest that the proposed class of hybrid networks have been effective in predicting the pollution of unknown locations with the best root mean squared error (RMSE) being 18.475 for Holt-Winters + IDW combination. Moreover, the proposed novel MLP Ensemble method for spatial inter-

polation has also been empirically shown to perform in the  $\pm 15\%$  range in terms of RMSE when used in combination with any temporal method discussed.

**Keywords** Spatio-Temporal Modeling · Hybrid Networks · Stacked Ensemble · Deep Learning · Air Pollution

## 1 Introduction

Air pollution has become a major concern in recent times. Rapid industrialization and towering technological growth have increased the levels of particulate matter in the atmosphere like never before. High socio-economic activities have made cities the hotbed of pollution in recent times. To monitor the levels of pollution, governments set up numerous air quality monitoring stations in major industrial districts all throughout the city. The hourly data collected is then analyzed and used by the concerned government agencies to frame new policies accordingly. Due to increased consciousness in the implications of air pollution and its long term effects on society, a new field of research termed Urban computing<sup>1</sup> has now become important in the recent decades, involving the application of wireless networks, sensors, computational resources and finally data to solve the issues arising in densely populated areas.

Data retrieval and processing is one of the major engineering challenges when it comes to air quality analysis and poses changes in both the spatial and temporal domains. In the majority of the cities worldwide, the number of air quality monitoring stations are comparatively lower than required to create an accurate geo-spatial profile of the pollution levels in the city. High real estate prices and expensive maintenance costs can be attributed to this problem as the focal points of pollution of a city are most often in the industrial and busy districts, thus producing a major challenge in the effective spatial distribution of sensor stations. Regarding the temporal

---

Pritthijit Nath  
Dept. of Computer Science and Engineering  
Jadavpur University, Kolkata, India  
E-mail: prithijit.nath@ieee.org

Pratik Saha  
Dept. of Computer Science  
SRM University, Kattankulathur, Chennai, India  
E-mail: pratiksaha198@gmail.com

Asif Iqbal Middy  
Dept. of Computer Science and Engineering  
Jadavpur University, Kolkata, India  
E-mail: asifm.rs@jadavpuruniversity.in

Sarbani Roy (Corresponding Author)  
Dept. of Computer Science and Engineering  
Jadavpur University, Kolkata, India  
E-mail: sarbani.roy@jadavpuruniversity.in

aspect of the problem, instrument failure turns out to be one of the inevitable challenges agencies face during air quality analysis. Due to instrument failure, historical data of certain time periods of various stations go missing and has to be imputed using several statistical techniques. In order to counter instrument failure, stations have to maintain backup monitoring devices, which further adds to the cost of maintenance of such air quality stations.

An extensive amount of research has been conducted to analyze the spatio-temporal variations of pollution in the recent decades. Sun et al.<sup>2</sup> investigated the pollution patterns of 5 year PM<sub>2.5</sub> data belonging to Jiangsu province in eastern China using Kriging and Hybrid Single Particle Lagrangian Integrated Trajectory (HYPLIT) models. Wang et al.<sup>3</sup> used the Box-Jenkins approach to predict pollution values of Long Beach in Los Angeles. Li et al.<sup>4</sup> proposed a hybrid CNN-LSTM model to forecast particulate matter concentration for the next 24 hours in Beijing. Although being crucial works in the field of pollution forecasting, the majority of the works in the field of air quality research are concentrated in the temporal aspect. The number of studies conducted to understand the spatio-temporal aspects have been relatively fewer.

Taking motivation from the recent advancements in spatial as well as the temporal aspects of the problem, this paper proposes a novel class of Hybrid Spatio-Temporal Pollution Forecasting Networks to counter the challenges mentioned before. The networks leverage the power of temporal forecasting by statistical and deep learning methods coupled with the effectiveness of geo-spatial and deep learning based interpolation techniques. Multi-site air-quality pollution data of Beijing<sup>5,6</sup> is used to demonstrate the performance of the proposed networks in this study.

The dataset is first segregated on the basis of different stations. The spatial coordinates of the stations are geocoded using Google Maps Geocoding API<sup>7</sup>. Stations are divided randomly into two sets - spatial train and spatial test. The proposed networks are trained on the stations mentioned in the spatial train set and the spatial interpolation evaluation is conducted on the stations present in the spatial test set. Within the spatial train set, the last one year is taken as the prediction horizon to evaluate the prediction accuracy of the temporal forecasting models. Various statistical approaches such as AutoRegressive (AR)<sup>8</sup>, AutoRegressive Integrated Moving Average (ARIMA)<sup>9</sup> and Holt-Winters<sup>10</sup> are used, whereas in deep learning models such as Stacked<sup>11</sup> and Bi-Directional LSTMs<sup>12</sup> are used to model temporal data. Geo-statistical methods such as Ordinary Kriging, Inverse Distance Weighted (IDW) along with a novel Multi-Layer Perceptron<sup>13</sup> Ensemble based approach, are used to perform spatial interpolation on the data as predicted by the temporal models to give us the required final predictions.

To summarize, the main contributions presented in this paper are as follows:

- A novel class of hybrid networks are proposed to perform spatio-temporal forecasts based on historical real world data to predict the pollution values of unknown locations.
- Display of a novel stacked ensemble spatial interpolation model, empirically shown to be capable of performing equivalently when compared to widely used methods such as Ordinary Kriging and IDW.
- Demonstration through extensive evaluation on a real world dataset that the proposed networks perform efficiently on the data.

The rest of the paper is organised as follows: In Section 2, relevant studies conducted in this field are discussed. Section 3 consists of the problem formulation along with a detailed description of the various methods used in the proposed class of networks. The results of the study are laid out in great detail in Section 4, along with a comprehensive discussion on the performance of the models. The conclusions drawn from the findings are ultimately presented in Section 5.

## 2 Related Works

This section deals with the recent research studies conducted in the field of air quality analysis to model pollution data using different novel spatio-temporal models .

Lindström et al.<sup>24</sup> developed a model using spatial and spatio-temporal covariates. They combined data from several different monitoring stations along with Geographic Information System (GIS) covariates to produce NO<sub>x</sub> predictions in the area around Los Angeles. Zheng et al.<sup>25</sup> used a semi-supervised learning approach to model the pollution data. A spatial classifier based on an artificial neural network (ANN) was used to model the spatial correlation between air qualities of different locations. To predict the temporal data, a linear chain conditional random field (CRF) was used to model the temporal dependency of air quality in a location. A previous study by Nath et al.<sup>26</sup> resampled the daily pollution data into monthly to better forecast the long term trends of temporal data in the city of Kolkata. Das et al.<sup>27</sup> performed a comparative review of high granular and short term time-series forecasting of PM<sub>2.5</sub> pollutant.

Self Organizing Maps (SOM) were used by Chang et al.<sup>28</sup> to extract and distinguish the spatio-temporal features of PM<sub>2.5</sub> concentrations of Taiwan. Liu et al.<sup>29</sup> proposed Spatio-Temporal Extreme Learning Machine (STELM) method to construct a prediction model based on air quality and meteorological data of Beijing. Abirami et al.<sup>30</sup> presented a hybrid spatio-temporal deep learning framework based on CNN-LSTM to model PM<sub>2.5</sub> concentrations based on data collected from different IoT based air quality sensors.

Author	Year	Method	Description
Li et.al <sup>14</sup>	2018	Sparse Denoising Autoencoding neural network	A novel spatio-temporal-aware architecture was proposed which makes full use of characteristics to improve accuracy.
Xu et.al <sup>15</sup>	2019	Spatio-Temporal Interpolated Echo State Network	Used cubic spline method for series smoothing later feeding into an echo state network computed by elastic-net algorithm.
He et.al <sup>16</sup>	2020	Spatio-Temporal Neural Network (STNN)	Used STNN with an encoder-decoder architecture and region-based spatial dependencies and external factors to improve accuracy.
Das et al. <sup>17</sup>	2017	Fuzzy Bayesian Network with incorporated Residual Correction mechanism (FBNRC)	Proposed a hybrid probabilistic model which reduces uncertainty during parameter learning process and compensates for the unknown factors during prediction of variables.
Zhou et.al <sup>18</sup>	2019	Knowledge Graph and deep Spatio-Temporal Convolutional Neural Network	Used features from a Graph convolutional network as the input to a spatio-temporal convolutional neural network to accurately forecast traffic congestion.
Soh et al. <sup>19</sup>	2018	Adaptive Deep Learning based Prediction Model	Used a combination of artificial neural network (ANN), a convolutional neural network (CNN), and a long-short-term memory (LSTM) to extract spatial-temporal relations for better performance.
Bui et al. <sup>20</sup>	2020	Spatio-Temporal Prediction using Multimodal Approach	Multimodal fusion of critical factors were used to predict future air quality levels thereby reducing the mean absolute errors of PM <sub>2.5</sub> prediction.
Fan et al. <sup>21</sup>	2017	Deep RNN based Prediction Framework	Used missing value processing algorithms and (DRNN) for prediction thereby outperforming baseline models.
Szpiro et al. <sup>22</sup>	2008	Bayesian hierarchical spatio-temporal model	Used spatial regression structure in a spatio-temporal model implemented using an optimized Markov Chain Monte Carlo (MCMC) sampling algorithm to estimate concentration of air pollutants.
Sahu et al. <sup>23</sup>	2006	Spatio-Temporal Modelling of Fine Particulate Matter	A fully Bayesian model was implemented, using MCMC techniques, enabling inference with regard to process unknowns as well as predictions in time and space.

Table 1: Summary of spatio-temporal models proposed by researchers in recent decades

A novel deep convolution neural network model named Multi-Scale Spatial Temporal Network (MSSTN) was proposed by Wu et al.<sup>31</sup> to better discover multi-scale spatial temporal patterns. MSSTN was used to make hourly PM<sub>2.5</sub> concentration predictions jointly for a number of cities based on the training done on Urban Air Pollution Datasets in North China.

Kloog et al.<sup>32</sup> proposed the use of recent advances in Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data processing algorithms to use high resolution Aerosol Optical Depth (AOD) data which is then used to train generalized additive mixed models with spatial smoothing to generate grid predictions.

An Extended Spatio-Temporal Granger Causality model was proposed by Zhu et al.<sup>33</sup> to analyze the causality relations present in air quality data using an efficient solution to overcome the computational complexity to process massive volumes of data along with an adaptation of the grid-based algorithm used to other non-grid based applications. Other related works are summarized in Table 1.

The merits of the class of networks proposed in this paper can be enumerated in the following ways. First, widely used temporal methods such as AR, ARIMA, Holt-Winters are used to produce temporal predictions. Since, pollution data follows repeated patterns, statistical and LSTM based methods perform efficiently compared to other temporal forecasting methods. Second, in the multi-site air-quality pollution data of Beijing<sup>5,6</sup>, inter-station distances being closer in the order of magnitude of 10<sup>1</sup> kilometers cause geo-statistical methods such as Ordinary Kriging and IDW methods to produce projections very close to actual levels if measured accordingly. Third, a novel Multi-Layer Perceptron (MLP) Ensemble based approach for spatial interpolation is also proposed which could handle data containing exogenous variables to further increase the accuracy of the model. The MLP Ensemble of various station data is shown to perform equivalent to existing geo-statistical methods mentioned before.

### 3 Problem Formulation and Proposed Approach

This section deals with the formulation of the problem statement followed by a detailed discussion of the approach along with the steps undertaken to build the model for this study.

#### 3.1 Problem Statement

Given a set of  $n$  known air quality monitoring stations  $S = \{S_1, S_2, \dots, S_n\}$  having pollution values  $P_i = \{v_{i1}, v_{i2}, \dots, v_{it}\}$  for the  $i$  th station over a duration of  $t$  timesteps in the past, our intention is to find the pollution data  $P_j = \{w_{j_{t+1}}, w_{j_{t+2}}, \dots, w_{j_{t+t'}}\}$  over the next  $t'$  timesteps for each  $j$  th location belonging to the set of  $m$  unknown locations  $S' = \{S'_1, S'_2, \dots, S'_m\}$ . Here,  $v$  and  $w$  denote actual and predicted pollution values respectively

#### 3.2 Proposed Approach

The approach undertaken in this study can be divided into two folds. First, the pollution data for the  $i$  th station,  $P_i = \{w_{i_{t+1}}, w_{i_{t+2}}, \dots, w_{i_{t+t'}}\}$  for station  $S_i \in S = \{S_1, S_2, \dots, S_n\}$  over the next  $t'$  timesteps are found out using one-step ahead prediction. These predictions are then taken as input for the spatial interpolation methods to calculate the pollution values  $P_j = \{w_{j_{t+1}}, w_{j_{t+2}}, \dots, w_{j_{t+t'}}\}$  for each unknown location  $S_j \in S' = \{S'_1, S'_2, \dots, S'_m\}$ .

---

#### Algorithm 1: ONESTEPAHEADPREDICTIONS

---

**Input:**  $P_{in} = \{v_{i1}, v_{i2}, \dots, v_{it}\}, t', \Omega$   
**Output:**  $P_{out} = \{w_{i_{t+1}}, w_{i_{t+2}}, \dots, w_{i_{t+t'}}\}$

- 1  $tmpPredictions \leftarrow \{\}$
- 2  $input \leftarrow \{P_{in}\}$
- 3 **for**  $i \leftarrow 1$  **to**  $t'$  **do**
- 4      $model \leftarrow \Omega(input)$
- 5      $modelFit \leftarrow fit(model)$
- 6      $nextPred \leftarrow predictNext(modelFit)$
- 7      $input \leftarrow append(input, nextPred)$
- 8      $tmpPredictions \leftarrow append(tmpPredictions, nextPred)$
- 9 **return**  $tmpPredictions$

---

##### 3.2.1 Temporal Modelling

For temporal modelling purposes, different widely used statistical methods such as AR<sup>8</sup>, ARIMA<sup>9</sup>, Holt-Winters<sup>10</sup> and deep learning methods such as Stacked<sup>11</sup> and Bi-Directional LSTMs<sup>12</sup> are used. The approach for carrying out one-step ahead predictions using a temporal model  $\Omega$  is laid out in Algorithm 1.

---

#### Algorithm 2: INVERSEDISTANCEWEIGHTED

---

**Input:**  $P_{out} = \{P_{out_1}, P_{out_2}, \dots, P_{out_n}\}, S, S'_j$   
**Output:**  $P_j = \{w_{j_{t+1}}, w_{j_{t+2}}, \dots, w_{j_{t+t'}}\}$

- 1  $distances \leftarrow generateDistanceMatrix(S, S'_j)$
- 2  $weights \leftarrow 1/(distances)^2$
- 3  $normalWeights \leftarrow weights/sum(weights)$
- 4  $sptPredictions \leftarrow P_{out} \cdot normalWeights$
- 5 **return**  $sptPredictions$

---



---

#### Algorithm 3: ORDINARYKRIGING

---

**Input:**  $P_{out} = \{P_{out_1}, P_{out_2}, \dots, P_{out_n}\}, S, S'_j$   
**Output:**  $P_j = \{P_{j_{t+1}}, P_{j_{t+2}}, \dots, P_{j_{t+t'}}\}$

- 1  $sptPredictions \leftarrow \{\}$
- 2  $variogram \leftarrow \text{"linear"}$
- 3 **for**  $i \leftarrow 1$  **to**  $t'$  **do**
- 4      $row \leftarrow \{\}$
- 5     **for**  $k \leftarrow 1$  **to**  $n$  **do**
- 6          $cell \leftarrow (P_{out_{k_i}}, S_{k_{lat}}, S_{k_{lng}})$
- 7          $row \leftarrow append(row, cell)$
- 8      $model \leftarrow OrdinaryKriging(row, variogram)$
- 9      $testCell \leftarrow (S'_{j_{lat}}, S'_{j_{lng}})$
- 10      $value \leftarrow executeGrid(model, testCell)$
- 11      $sptPredictions \leftarrow append(sptPredictions, value)$
- 12 **return**  $sptPredictions$

---



---

#### Algorithm 4: STACKEDENSEMBLE

---

**Input:**  $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_t\}, S, C = \{C_1, C_2, \dots, C_t\}$   
**Output:**  $\Psi$

- 1  $modelPred \leftarrow \{\}$
- 2  $actualVals \leftarrow \{\}$
- 3 **for**  $i \leftarrow 1$  **to**  $n$  **do**
- 4      $X, y \leftarrow ModelData(C, S_i, S)$
- 5      $preds \leftarrow predict(\Phi_i, X)$
- 6      $modelPred \leftarrow append(modelPred, preds)$
- 7      $actualValues \leftarrow append(actualValues, y)$
- 8  $model = LinearRegressor()$
- 9  $\Psi = fit(model, modelPred, actualValues)$
- 10 **return**  $\Psi$

---

The model  $\Omega$  is re-trained in each run on new data. The new data is created by appending the predicted data in the  $i - 1$  th run to the existing data  $P_{in}$  provided as initial input. For each run, the resulting prediction  $w_{i_k} \forall k \in \{t + 1, t + 2, \dots, t + t'\}$  is appended to the result set  $P_{out}$  produced as the required output. This output so produced serves the basis for the spatial interpolation process in the following subsections.

##### 3.2.2 Geo-Statistical Spatial Interpolation

Spatial interpolation is carried out in two ways in the study. One way is using geo-statistical methods such as Inverse

**Algorithm 5: CREATECELLS**

**Input:**  $P_{in} = \{P_{in_1}, P_{in_2}, \dots, P_{in_t}\}, S = \{S_1, S_2, \dots, S_n\}, t$   
**Output:**  $C = \{C_1, C_2, \dots, C_t\}$

```

1 values ← {}
2 for i ← 1 to t do
3   row ← {}
4   for j ← 1 to n do
5     cell ← (Pinj, Slat, Slong)
6     row ← append(row, cell)
7   values ← append(values, row)
8 return values

```

**Algorithm 6: MODELDATA**

**Input:**  $C = \{C_1, C_2, \dots, C_t\}, S_f, S = \{S_1, S_2, \dots, S_n\}$   
**Output:**  $X, y$

```

1 X ← {}, y ← {}
2 for i ← 1 to t do
3   fixCell ← getCell(Ci, i, Sf)
4   fixPollVal ← getPollutionValue(fixCell)
5   for j ← 1 to n do
6     distance ← getDistance(fixCell, Cj)
7     pollValue ← getPollutionValue(Cj)
8     cell ← (fixPollVal, distance)
9     X ← append(X, cell)
10    y ← append(y, pollValue)
11 return X, y

```

Distance Weighted (IDW) and Ordinary Kriging, while the other way is by using the proposed multi-site MLP Ensemble. Methods such as Ordinary Kriging (OK) and Inverse Distance Weighted (IDW) have been used due to the wide variety of applications in spatial interpolation purposes. Recent studies such as by Kostopoulou<sup>34</sup> and Apriyono et al.<sup>35</sup> use methods such as OK and IDW to aid in interpolation of spatial features when the datapoints are sparsely distributed. The approaches for geostatistical spatial interpolation using IDW and Ordinary Kriging as discussed are laid out in Algorithms 2 and 3.

In case of the IDW approach, a distance matrix is calculated first, based on the spatial co-ordinates of  $S$  and  $S'_j$ . A power parameter, in our case, a value of 2 is chosen, thereby implying that the square of the inverse of the distances will be used as the weights. The weights are further normalized by dividing with the sum. The normalized weights so calculated are then multiplied using the scalar product operation, with the temporal predictions  $P_{out}$  from different stations to ultimately produce the final results  $P_j$ .

Ordinary Kriging uses a different approach to solve the spatial interpolation problem. A variogram is used to model the spatial dependence between different pairs of points situated at different distances. Rows of values based on the the pollution value  $P_{out_{k_i}}$  and the spatial coordinates of stations

$S_k \forall k \in \{1, 2, 3, \dots, n\}$  are created which are used to fit the variogram. The output for each  $i$  th timestamp is calculated by executing the grid based on the spatial values of the unknown station  $S'_j$ .

**3.2.3 Spatial Interpolation using MLP Ensemble**

The multi-site ensemble proposed in this study is based on a deep learning approach to model spatial relations in the data and use the spatial relations so found out to help in producing required predictions. The novel ensemble also aids in the investigation of the closeness in terms of efficacy when compared to tried and tested approaches such as OK and IDW. In this study, a stacked ensemble of various station based MLP deep learning models are used for modeling the data to produce the final predictions.

The stacked ensemble learning approach is illustrated in Algorithm 4. In the ensembling stage, the processed datasets containing  $X$  and  $y$  output from the data modeling stage is used to train a MLP based deep learning model  $\Phi_i$  for each fixed station  $S_f$ . As illustrated in Fig.1, all intermediate models  $\Phi_i \forall i \in \{1, 2, 3, \dots, n\}$  are then used to produce predictions which are used to train a linear regressor based meta-learner  $\Psi$ , to produce the final predictions.

The data modeling stage involves the raw data to be first converted using Algorithm 5 into tuples, which are also described as cells in this study. After the creation of data cells collected column wise over all rows, the data cells are then used to create the independent  $X$  and dependent  $y$  sets.

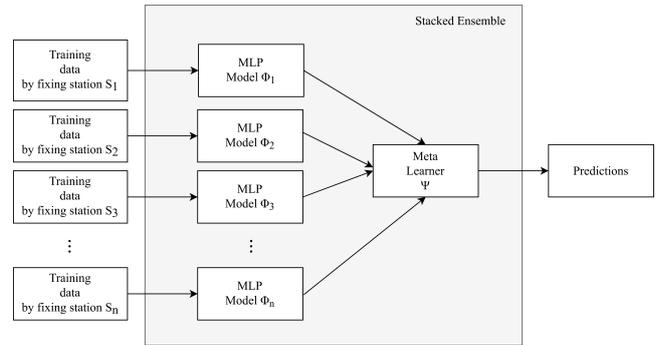


Fig. 1: Representation of the stacked ensembling process

As shown in Algorithm 6, a station  $S_f$  is first fixed  $\forall f \in \{1, 2, 3, \dots, n\}$ . The pollution value at each timestep  $i$ , is used as the  $y$  value of the row, while the tuple containing the pollution value of the fixed station  $S_f$  and distance of the fixed station  $S_f$  w.r.t  $S_j \forall j \in \{1, 2, 3, \dots, i, \dots, n\}$  is taken as the  $X$  value. If additional values are present, then such corresponding relative differences between  $S_f$  and  $S_i$  can be appended to  $X$ , thereby making the model input extremely flexible for a wide number of use-cases. The  $X$  and  $y$  sets

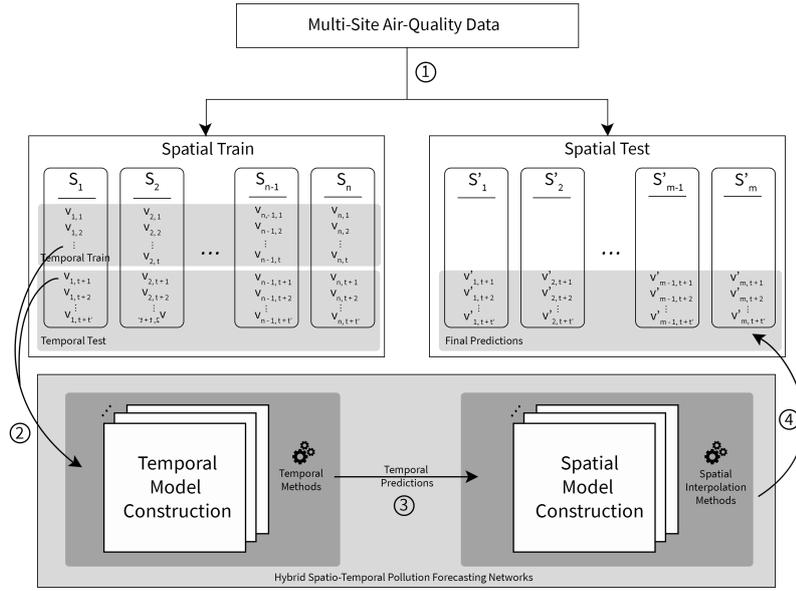


Fig. 2: General overview of the steps undertaken to build the class of proposed hybrid networks

so created based on the fixed station  $S_f$  are then fed into corresponding models for training to create the final stacked MLP ensemble model.

### 3.3 Model Building

The steps followed to build the model are illustrated in Fig. 2. In step 1, the multi-site air quality data is split into two components namely, Spatial Train and Spatial Test. Spatial Train consists of known stations denoted by  $S = \{S_1, S_2, \dots, S_n\}$  while Spatial Test consists of unknown stations denoted by  $S' = \{S'_1, S'_2, \dots, S'_m\}$ . In the following step, the time series pollution values belonging to Spatial Train are subdivided further into Temporal Train and Temporal Test (also known as the Prediction Horizon) for the training and evaluation of temporal methods such as AR<sup>8</sup>, ARIMA<sup>9</sup>, Holt-Winters<sup>10</sup>, Stacked<sup>11</sup> and Bi-directional LSTMs<sup>12</sup>.

Step 3 involves the carrying forward of the predictions  $P_{i_{out}} \forall i \in \{1, 2, 3, \dots, n\}$ , produced by the temporal methods mentioned earlier. The predictions  $P_{i_{out}}$  are then fed into various spatial-interpolation methods such as IDW, Ordinary Kriging and the proposed stacked MLP ensemble to produce the final predictions  $P_j \forall j \in \{1, 2, 3, \dots, m\}$  needed. The resulting final predictions  $P_j$  from the spatial-interpolation models are then used to carry out the final evaluation w.r.t the ground truth values present in Spatial Test as pointed by the arrow for step 4. The extensive analysis of different combinations of temporal and spatial interpolation methods are reported later in the following section of this study.

## 4 Evaluation and Results

In this section, the performance of the proposed class of networks is presented along with the discussion on the evaluation metrics and the parameters.

### 4.1 Experimental Setup

All calculations have been performed on a test bench consisting of a 6C/12T Ryzen 5 3600 CPU clocked at 3.6 GHz coupled with a 16 GB 3000 MHz DDR4 RAM and a 1 TB NVMe SSD. In case of deep learning based models, an Nvidia RTX 2070 Super GPU is used as a hardware accelerator in order to increase parallelism in matrix related calculations.

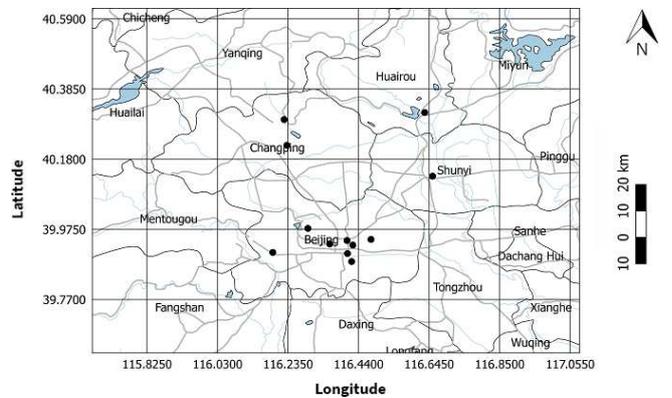


Fig. 3: Map showing all 12 pollution monitoring sites (indicated by black markers) in Beijing

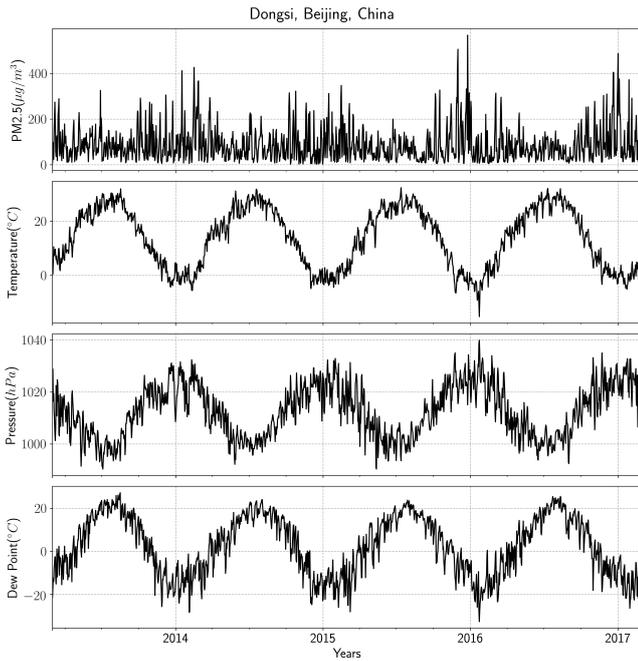


Fig. 4: PM<sub>2.5</sub>, Temperature, Pressure and Dew Point plots for pollution monitoring station located at Dongsi, Beijing

The code for the development of this study has been written in Python 3.8<sup>36</sup>, due to the rich ecosystem consisting of several highly efficient libraries such as NumPy<sup>37</sup>, TensorFlow<sup>38</sup>, StatsModels<sup>39</sup> and Sci-kit Learn<sup>40</sup> to reduce overall complexity of the code without any trade off with regards to performance. The operating system by choice has been Ubuntu 20.04 LTS due to increased support for the developed code in platforms like Linux.

For carrying out the performance analysis of the models, the dataset containing 12 stations is first split randomly in half to create sub-datasets for training and evaluation purposes respectively. Due to the presence of temporal methods, the last year of both datasets is taken as the prediction horizon. The results presented are the final predictions performed by the spatial interpolation methods, on the one step ahead predictions produced by the temporal models. The metrics are evaluated by comparing the spatially interpolated predictions w.r.t. the actual last year values for all unknown stations in the test set.

#### 4.2 Data Description

The dataset used for this study is the Beijing multi-site air quality pollution data<sup>5</sup> donated to the UCI Machine Learning Repository<sup>41</sup> by Zhang et al.<sup>6</sup>. The dataset contains multi-site pollution data from 12 different government controlled air quality monitoring stations scattered all throughout the capital city of Beijing. 6 pollutants such as PM<sub>2.5</sub>, PM<sub>10</sub>,

SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub> are reported with hourly frequency in the dataset, out of which this study concerns only with the PM<sub>2.5</sub> data. In addition to pollution data, other meteorological factors such as temperature, pressure, dew point, precipitation, wind speed and wind direction are also present in the dataset, of which all are currently out of scope for the current version of the study. As the geographical locations of the air quality monitoring stations are absent, the required coordinates for each location have been externally found out using the Google Maps Geocoding API<sup>7</sup>. Missing values (if present) in the data for each station are imputed using Mean Before After and Window based Last Observation Carried Forward (LOCF) techniques depending on the time period of the values missing. The data ranges from 1st March, 2013 to 28th February, 2017 thus, accumulating to a total of 1461 rows for each station when resampled into daily frequency by taking the mean.

#### 4.3 Evaluation Metrics

The performance of different models of the proposed class of networks is based on Root Mean Squared Error (RMSE) and Mean Average Error (MAE) in contrast to the Test Set Mean. RMSE causes the effect of each error to be proportional to the size of the squared error, thereby causing larger errors to disproportionately have a larger effect on the RMSE. The Test Set Mean is the mean of all the values present in the testing set, thus helping in providing a perspective of the errors made during prediction.

In addition, relative improvements in the performance of all temporal models having the same spatial interpolation method is also studied to analyze how much the models outperform when compared with the worst performing ones.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}, \text{MAE} = \frac{\sum_{t=1}^T |\hat{y}_t - y_t|}{T} \quad (1)$$

where  $\hat{y}_t$  is the prediction made by the model and  $y_t$  is the actual value at instant  $t$ . Here  $T$  denotes the count of the number of time-series samples.

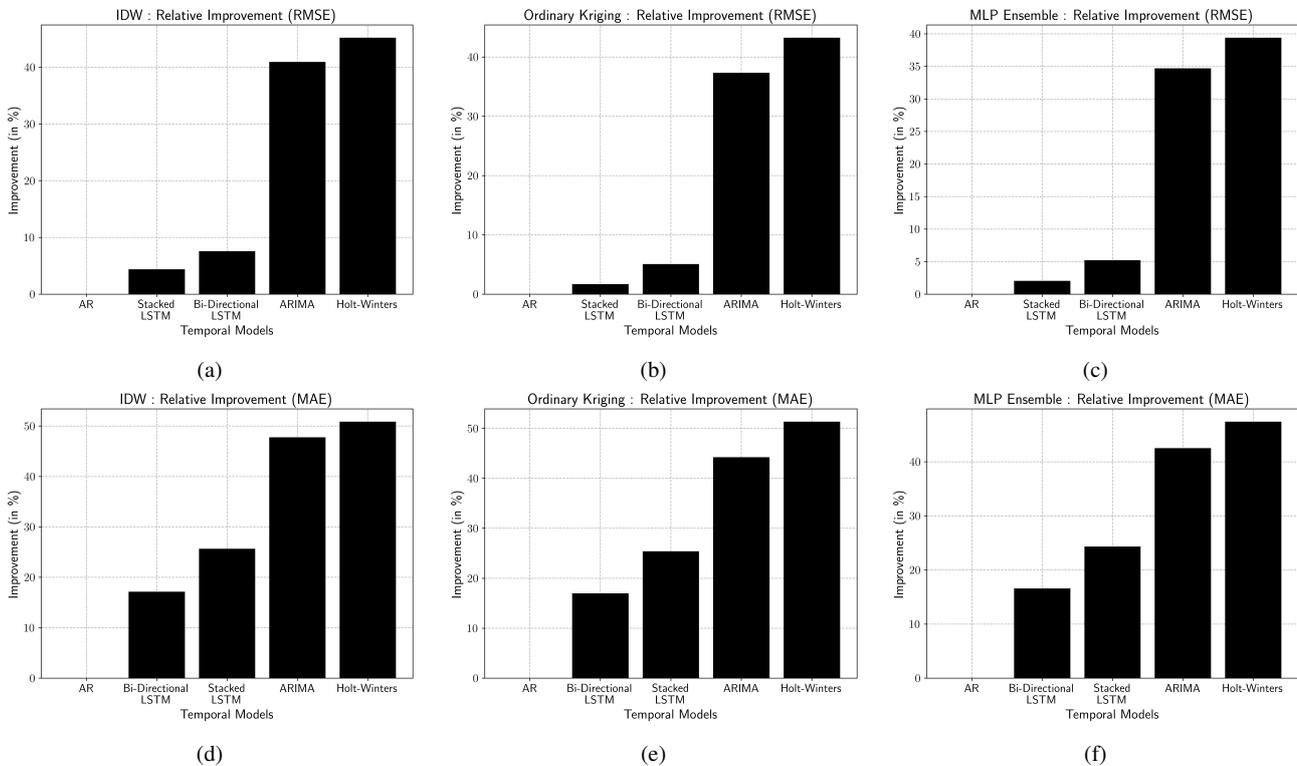
Due to the high training time for deep learning models, one epoch was used to retrain the model during each one step ahead forecast.

#### 4.4 Results

The performance metrics of all possible combinations of temporal methods and spatial interpolation methods used in the class of hybrid networks proposed are presented in Table 2. Out of all workable combinations, from Fig.5, it is clearly visible that combinations having Holt-Winters<sup>10</sup> as

Temporal Method	Spatial Interpolation Method	RMSE	MAE	Test Set Mean
AR <sup>8</sup>	Inverse Distance Weighted	33.728	25.333	75.630
	Ordinary Kriging	34.136	25.35	
	MLP Ensemble	34.920	25.898	
ARIMA <sup>9</sup>	Inverse Distance Weighted	19.919	13.228	
	Ordinary Kriging	21.391	14.135	
	MLP Ensemble	22.802	14.885	
Holt Winters <sup>10</sup>	Inverse Distance Weighted	<b>18.475</b>	<b>12.447</b>	
	Ordinary Kriging	<b>19.368</b>	<b>12.332</b>	
	MLP Ensemble	<b>21.158</b>	<b>13.621</b>	
Stacked LSTM <sup>11</sup>	Inverse Distance Weighted	32.251	18.831	
	Ordinary Kriging	33.554	18.907	
	MLP Ensemble	34.206	19.595	
Bi-Directional LSTM <sup>12</sup>	Inverse Distance Weighted	31.168	20.992	
	Ordinary Kriging	32.405	21.046	
	MLP Ensemble	33.096	21.604	

Table 2: Performance metrics of the proposed hybrid models

Fig. 5: Relative improvement graphs in RMSE (a-c) and MAE (d-f) scores for various temporal models over Auto Regressive<sup>8</sup> (grouped by different spatial interpolation methods)

the temporal method showed the best performance with a relative improvement  $\approx 40\%$  in RMSE, when compared w.r.t the hybrid combination with different spatial interpolation methods.

The ability for hybrid pairs containing Holt-Winters<sup>10</sup> to outperform all others stems from the advantage of using triple exponential smoothing for the level, trend and seasonal components of the data. Since exponential smoothing mod-

els are based on the description of trend and seasonality, it performs better than hybrid pairs using AR<sup>8</sup> and ARIMA<sup>9</sup> as those are built to describe the correlations in the data.

From Table 2, it can be seen that the worst performing combinations out of all, are those hybrid pairs involving AR<sup>8</sup> as the temporal method. This can be attributed to the presence of only the auto-regressive component whereas methods like

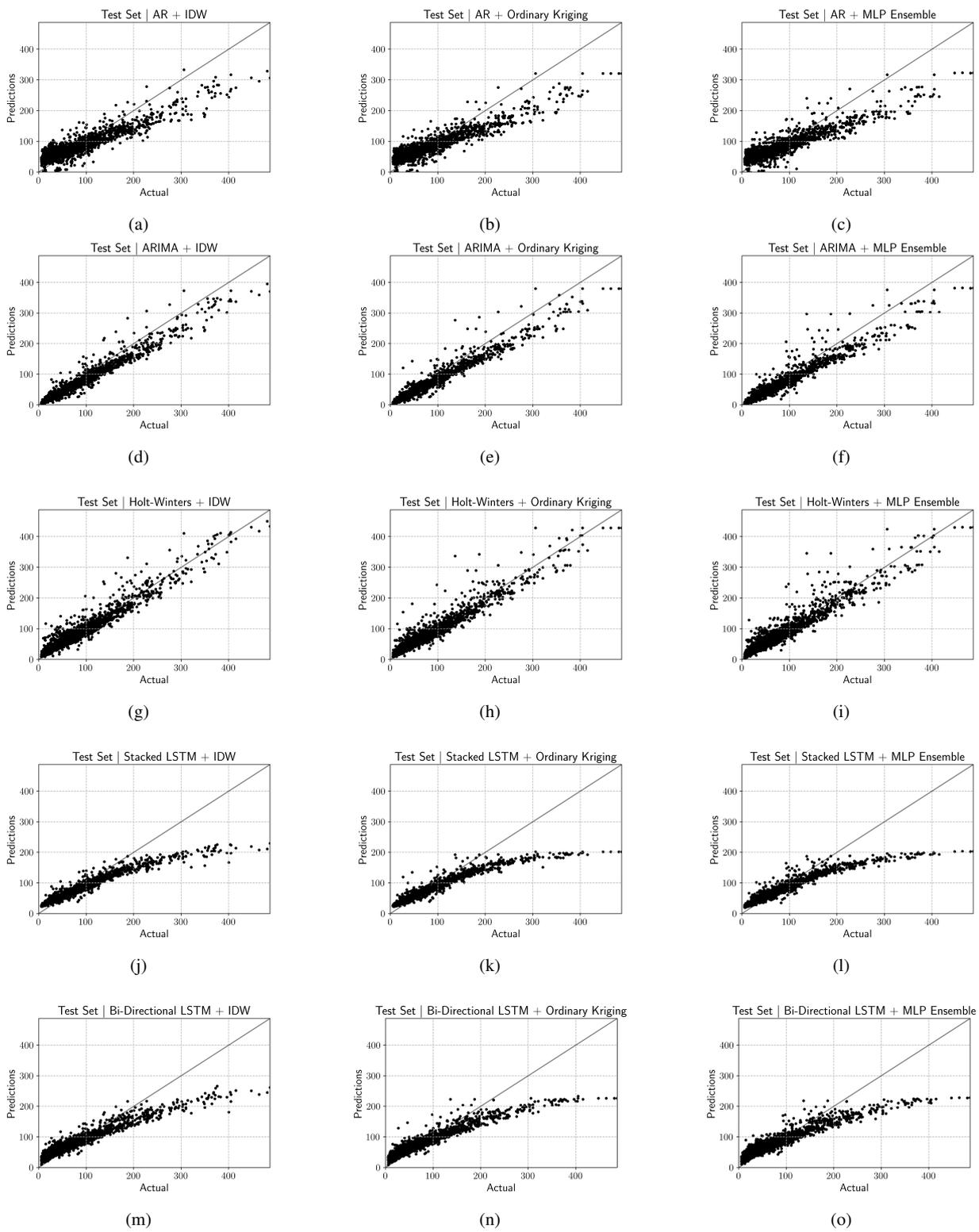


Fig. 6: Actual vs Predicted correlation plots for one-step ahead prediction by AR (a-c), ARIMA (d-f), Holt-Winters (g-i), Stacked LSTM (j-l) and Bi-Directional LSTM (m-o) based hybrid forecasting models

ARIMA<sup>9</sup> take into account both the auto-regressive as well as the moving-average component of the time series.

Hybrid pairs involving deep learning methods such as Stacked LSTM<sup>11</sup> and Bi-Directional LSTM<sup>12</sup> performed relatively worse compared to methods involving ARIMA<sup>9</sup> and Holt-Winters<sup>10</sup> as can be inferred from Table 2. The relatively poor performance of deep learning methods can be attributed to many different causes. One of them might be the lack of adequate data for the methods to properly model data while another cause can be attributed to noisy components present. One interesting observation to note from Fig.6 is that deep learning based temporal methods were unable to model high values such as peaks and hence plateaued around the value of 200, while models such as AR<sup>8</sup>, ARIMA<sup>9</sup> and Holt-Winters<sup>10</sup> were successful in producing predictions close to the actual ones both in the higher as well as in the lower ends of the spectrum.

Another interesting observation that can be concluded from Table 2 is that all spatial interpolation methods including the proposed novel MLP Ensemble, showed equivalent performances with each other. This also can be attributed to the fact that the geo-spatial surface in Beijing is flat and not uneven, thereby enabling the spatial interpolation methods to model the surface accurately.

#### 4.5 Discussion

Out of all the different combinations tested on the Beijing multi-site hourly pollution data<sup>5</sup>, hybrid pairs containing Holt-Winters<sup>10</sup> gave the overall best performance. It was also shown that the novel MLP based multi-site ensemble performed equivalently to other widely used spatial interpolation methods such as Ordinary Kriging and Inverse Distance Weighted (IDW). The biggest contributing performance of the multi-site ensemble method lies in its flexibility to account for more than one feature to accurately model the geo-spatial surface. If the amount of data increases to the order of magnitude  $\geq 10^6$ , deep learning based temporal methods will be expected to outshine the statistical ones.

In the case of Beijing, the range of pollution values was quite high compared to other major cities in the world, thus leading to higher error values in terms of RMSE for temporal methods. Cities, where the range of pollution values are comparatively lower, will allow better performance for temporal methods, thereby leading the proposed class of networks to produce much better overall predictions for unknown locations. In addition, the performance can be also attributed to the type of surface being operated on. If there are undulations present in the surface, spatial interpolation methods may not produce good and accurate results.

**As pollution monitoring sensors are prone to instrument failures, in the likely scenario of a real-world deployment, the**

**additional data due to failure in the sensor will be absent from the feed. As prediction on absent input data is not possible for the proposed class of networks, the missing data will have to be imputed through LOCF technique for the model to produce the required predictions for the future timesteps.**

## 5 Conclusion

In this study, a novel class of hybrid spatio-temporal networks have been proposed to aid in predicting pollution values of unknown locations when historical time-series data of other locations are known. At this stage, exogenous variables in temporal data forecasting were not taken into account. As a possible future work, statistical methods such as ARIMAX can be used which takes into usage the contributions of exogenous variables in producing predictions. The data input for LSTMs can be extended to take into account a higher channel depth to include exogenous variables such as meteorological factors and any other location specific features.

It is our hope that through the use of the class of hybrid networks proposed in this paper, government agencies will be able to efficiently monitor pollution in their cities using the existing infrastructure of air quality monitoring stations present. As the errors produced are low enough to cause little to no change in AQI categories<sup>42</sup>, the proposed class of networks are highly effective in producing accurate air quality level forecasts for places where the government is unable to set up infrastructure for real time air pollution monitoring stations. Even in case of instrument failure, it is our hope that these networks will come in handy to compensate for the missing pollution readings till the instrument undergoes maintenance and put back to proper working condition.

## Acknowledgements

The research work of Asif Iqbal Middy is supported by UGC-NET Junior Research Fellowship (UGC-Ref. No.: 3684/(NET-JULY 2018)) provided by the University Grants Commission, Government of India. This research work is also supported by the project entitled "Participatory and Real-time Pollution Monitoring System For Smart City, funded by Higher Education, Science & Technology and Biotechnology, Department of Science & Technology, Government of West Bengal, India".

## Data Availability

Manuscript has associated data in a private data repository. Data will be made available on reasonable request.

## Compliance with Ethical Standards

### Conflict of Interest

The authors declare that they have no conflict of interest.

### Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## References

- Paulos, E., Anderson, K., Townsend, A.: Ubicomp in the urban frontier (2004)
- Sun, X., Luo, X.S., Xu, J., Zhao, Z., Chen, Y., Wu, L., Chen, Q., Zhang, D.: Spatio-temporal variations and factors of a provincial pm 2.5 pollution in eastern china during 2013–2017 by geostatistics. *Scientific reports* **9**(1), 1–10 (2019)
- Wang, W., Guo, Y.: Air pollution pm<sub>2.5</sub> data analysis in los angeles long beach with seasonal arima model. In: 2009 international conference on energy and environment technology, vol. 3, pp. 7–10. IEEE (2009)
- Li, T., Hua, M., Wu, X.: A hybrid cnn-lstm model for forecasting particulate matter (pm<sub>2.5</sub>). *IEEE Access* **8**, 26,933–26,940 (2020)
- Chen, S.: Beijing multi-site air-quality data data set. UCI Machine Learning Repository (2017). URL <https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data>. Accessed: 2020-08-26
- Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., Chen, S.: Cautionary tales on air-quality improvement in beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **473**, 20170,457 (2017). DOI 10.1098/rspa.2017.0457
- Google maps geocoding api. URL <https://developers.google.com/maps/documentation/geocoding/overview>. Accessed: 2020-08-26
- Walker, G.T.: On periodicity in series of related terms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **131**(818), 518–532 (1931)
- Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: *Time series analysis: forecasting and control*. John Wiley & Sons (2015)
- Winters, P.R.: Forecasting sales by exponentially weighted moving averages. *Management science* **6**(3), 324–342 (1960)
- Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
- Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* **45**(11), 2673–2681 (1997)
- Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. Tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science (1985)
- Li, Y., Shen, X., Han, D., Sun, J., Shen, Y.: Spatio-temporal-aware sparse denoising autoencoder neural network for air quality prediction. In: 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 96–100. IEEE (2018)
- Xu, M., Yang, Y., Han, M., Qiu, T., Lin, H.: Spatio-temporal interpolated echo state network for meteorological series prediction. *IEEE transactions on neural networks and learning systems* **30**(6), 1621–1634 (2018)
- He, Z., Chow, C.Y., Zhang, J.D.: Stnn: A spatio-temporal neural network for traffic predictions. *IEEE Transactions on Intelligent Transportation Systems* (2020)
- Das, M., Ghosh, S.K.: Spatio-temporal prediction under scarcity of influencing variables: a hybrid probabilistic graph-based approach. In: 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), pp. 1–6. IEEE (2017)
- Zhou, G., Chen, F.: Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network. In: 2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT), pp. 105–108. IEEE (2019)
- Soh, P.W., Chang, J.W., Huang, J.W.: Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *Ieee Access* **6**, 38,186–38,199 (2018)
- Bui, T.C., Kim, J., Kang, T., Lee, D., Choi, J., Yang, I., Jung, K., Cha, S.K.: Star: Spatio-temporal prediction of air quality using a multimodal approach. In: *Proceedings of SAI Intelligent Systems Conference*, pp. 389–406. Springer (2020)
- Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., Lin, S.: A spatiotemporal prediction framework for air pollution based on deep rnn. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**, 15 (2017)
- Szpiro, A., Sampson, P., Sheppard, L., Wilton, D., Larson, T., Lumley, T., Adar, S., Kaufman, J.: Spatio-temporal modeling to predict intra-urban variation in air pollution levels. *Epidemiology* **19**(6), S322 (2008)
- Sahu, S.K., Gelfand, A.E., Holland, D.M.: Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics* **11**(1), 61–86 (2006)
- Lindström, J., Szpiro, A.A., Sampson, P.D., Oron, A.P., Richards, M., Larson, T.V., Sheppard, L.: A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates. *Environmental and ecological statistics* **21**(3), 411–433 (2014)
- Zheng, Y., Liu, F., Hsieh, H.P.: U-air: When urban air quality inference meets big data. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1436–1444 (2013)
- Nath, P., Saha, P., Mridha, A.I., Roy, S.: Long-term time-series pollution forecast using statistical and deep learning methods. *Neural Computing and Applications* pp. 1–20 (2021)
- Das, R., Mridha, A.I., Roy, S.: High granular and short term time series forecasting of pm<sub>2.5</sub> air pollutant - a comparative review. *Artificial Intelligence Review* (2021). DOI 10.1007/s10462-021-09991-1. URL <https://doi.org/10.1007/s10462-021-09991-1>
- Chang, F.J., Chang, L.C., Kang, C.C., Wang, Y.S., Huang, A.: Explore spatio-temporal pm<sub>2.5</sub> features in northern taiwan using machine learning techniques. *Science of the Total Environment* **736**, 139,656 (2020)
- Liu, B., Yan, S., Li, J., Li, Y.: Forecasting pm<sub>2.5</sub> concentration using spatio-temporal extreme learning machine. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 950–953. IEEE (2016)
- Abirami, S., Chitra, P., Madhumitha, R., Kesavan, S.R.: Hybrid spatio-temporal deep learning framework for particulate matter (pm<sub>2.5</sub>) concentration forecasting. In: 2020 International Conference on Innovative Trends in Information Technology (ICITIT), pp. 1–6. IEEE (2020)
- Wu, Z., Wang, Y., Zhang, L.: Msstn: Multi-scale spatial temporal network for air pollution prediction. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 1547–1556. IEEE (2019)
- Kloog, I., Chudnovsky, A.A., Just, A.C., Nordio, F., Koutrakis, P., Coull, B.A., Lyapustin, A., Wang, Y., Schwartz, J.: A new hybrid spatio-temporal model for estimating daily multi-year pm<sub>2.5</sub> concentrations across northeastern usa using high resolution aerosol

- 1 optical depth data. *Atmospheric Environment* **95**, 581–590 (2014)
- 2 33. Zhu, J.Y., Sun, C., Li, V.O.: An extended spatio-temporal granger  
3 causality model for air quality estimation with heterogeneous urban  
4 big data. *IEEE Transactions on Big Data* **3**(3), 307–319 (2017)
- 5 34. Kostopoulou, E.: Applicability of ordinary kriging modeling tech-  
6 niques for filling satellite data gaps in support of coastal manage-  
7 ment. *Modeling Earth Systems and Environment* **7**(2), 1145–1158  
8 (2021)
- 9 35. Apriyono, A., Santoso, P.B., et al.: Landslide susceptible areas iden-  
10 tification using idw and ordinary kriging interpolation techniques  
11 from hard soil depth at middle western central java, indonesia.  
12 *Natural Hazards* pp. 1–12 (2021)
- 13 36. Van Rossum, G., Drake Jr, F.L.: Python tutorial, vol. 620. Centrum  
14 voor Wiskunde en Informatica Amsterdam (1995)
- 15 37. Van Der Walt, S., Colbert, S.C., Varoquaux, G.: The numpy array:  
16 a structure for efficient numerical computation. *Computing in  
17 science & engineering* **13**(2), 22–30 (2011)
- 18 38. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro,  
19 C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: Tensorflow:  
20 Large-scale machine learning on heterogeneous distributed systems  
21 (2016)
- 22 39. Seabold, S., Perktold, J.: Statsmodels: Econometric and statistical  
23 modeling with python, vol. 57 (2010)
- 24 40. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion,  
25 B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.,  
26 Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M.,  
27 Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal  
28 of Machine Learning Research* **12**, 2825–2830 (2011)
- 29 41. Dua, D., Graff, C.: UCI machine learning repository (2017). URL  
30 <http://archive.ics.uci.edu/ml>. Accessed: 2020-08-26
- 31 42. Air Now, U.S. Environmental Protection Agency: Aqi ba-  
32 sics. URL <https://www.airnow.gov/aqi/aqi-basics/>. Ac-  
33 cessed: 2020-08-26
- 34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65