Air Pollution Hotspot Detection and Identification of Their Source Trajectory

K B Gurumoorthy ¹, S P Vimal ², N Sathish Kumar ³ and M Kasiselvanathan ⁴

¹ Assistant Professor, Sri Ramakrishna Engineering College, Coimbatore-22, Tamilnadu, India.
² Associate Professor, Sri Ramakrishna Engineering College, Coimbatore-22, Tamilnadu, India.
³ Professor, Sri Ramakrishna Engineering College, Coimbatore-22, Tamilnadu, India.
⁴ Assistant Professor, Sri Ramakrishna Engineering College, Coimbatore-22, Tamilnadu, India.

Email: gurumoorthy.kb@srec.ac.in

Abstract. Millions of deaths everywhere the planet, thanks to anthropogenesis fine material (or PM2.5) is principally caused thanks to outside pollution. Coimbatore may be a centre of textile and cotton trade, producing, poultry farming, education, info technology and health care and it’s the second largest town once Chennai within the state of state. Thus, this paper predicts the accumulation of PM2.5 from wind (velocity and direction) and precipitation levels. It imbibes a machine learning (ML) algorithm supported six years of earth science and pollution information inferences. At present, pollution may be a world downside. Republic of India is additionally an enormous sufferer of this downside. Thus, it's necessary to spot the recent spots of pollutants and their transport specifically carbon monoxide gas (CO), sulphur-dioxide (SO2) and oxides of element (NO+NO2) victimization advanced information analysis techniques. Challenges concerned during this current statement is mining the datasets from completely different parameters and providing the ultimate output with moderate abstraction resolution on pollution info. Therefore, the study illustrates that the employment of applied mathematics models supported the ML algorithm is most relevant to predict PM2.5 accumulation from earth science information.

1. Introduction

The machine learning algorithmic rule or approaches to be selected area unit determined, supported the matter statement, here we have a tendency to study the surroundings and analyse what form of inputs and knowledge area unit about to be required. however generally, we will classify the machine learning algorithms in 2 groups: 1) Similarity algorithms and 2) Learning algorithms. the newest urban air quality info says that ninety-eight of cities in lower and middle financial gain countries with over one hundred,000 inhabitants don't meet air quality tips of the globe Health Organization (WHO). A recent study employing an international atmospherically chemistry model found that three.3 million annual premature deaths everywhere world area unit joined to outside pollution, that is anticipated to be doubled by the year of 2050, principally thanks to phylogeny fine PM two.5 is shown in figure 1.
With the excellence of business institution and fast increase of population, pollution is increasing day by day. the rise in industrial activities triggers will increase the employment of vehicles. Combustion of fuels within the completely different production method within the business and vehicles turn out heaps of dangerous gases within the surroundings. thanks to pollution, air is contaminating day by day which suggests the quality of air is decreasing that have negative effects on our health inflicting completely different severe diseases like respiratory organ diseases and bronchial asthma, aversions and internal organ diseases. pollution can even cause premature deaths as shown in figure (1). So, pollution should be controlled to confirm a more robust, inexperienced and pollution aggregation.

The machine learning is used as a preferred option in predicting pollution, since it is considers both meteorological factors and climatic parameters. Hence it is undoubtedly feasible approach to improve the prediction of PM 2.5 increase in coimbatore.

This methodology is done using appropriate choice of meteorological features for two important factors, namely easy availability of meteorological data in any urban area resulting in cheaper model and secondly proper meteorological parameters must be selected since general model may not fit for a particular city. Hence keeping these factors in mind the research focus is on connectivity between important meteorological factors such as wind velocity, wind direction, precipitation and PM 2.5 contamination in Coimbatore.

- The key sources of air contaminants in urban areas are alarming level of transport on-road dust, open waste burning and the industries like large cement plants as shown in figure 2.
- By the year of 2030, the contribution of emissions from domestic cooking and lighting my
decrease with entry of LPG and residential electrification in almost every section of society.

- The environmental standards need to be revisited enforced strictly in case of cement producing plants in order to curb the pollution and improve the urban air quality.

- One of the main concern is open waste burning in dumping area across the city which is key source of air pollution, this requires stricter regulation for addressing the issues.

In the present study, the figure 3 depicting the spatial spread of the particulate matter according to wind (velocity and direction) and precipitation criterion in Coimbatore is displayed.

![Figure 3. Topographic map of Coimbatore](image)

2. **Data Collection**

Air pollution is a manmade occurrence and is afflicted by many factors. To ensure accurate forecast, it is necessary to identify the right parameters influencing the air pollution. Therefore, the right data collection and data processing is very essential.

2.1. Site description

Coimbatore, additionally referred to as Kovai or typically corn as Covai, may be a district within the Tamil Nadu. It's settled near the Noyyal stream and within the foots of the Western Ghats. Coimbatore is the second largest town in Tamil Nadu when metropolis and placed as sixteenth biggest urban agglomeration in Republic of India as per the census 2011. Coimbatore town lies at 11°1′6″N 76°58′21″E in south of Republic of India at 412 metres (1349 ft) higher than water level. It covers a vicinity of 642.12 km² (247.92sq m). below the Koppen climate classification, Coimbatore town incorporates a moderate climate, with a wet seasoning lasting from Sep to Nov because of the north-east monsoon. That mean most temperature ranges from 35.9 °C (97°F) to 29.2 °C (85 °F) and therefore the minimum temperature varies from 24.5°C (76 °F) to 19.8 °C (68 °F) the very best temperature ever recorded is 42.6 °C. Coimbatore incorporates a population of 1,601,438. It's a serious centre for producing education and health care. Coimbatore is one amongst the quickest developing cities in Republic of India. It homes over 25,000 small, medium and enormous industries. Coimbatore is termed as "The Manchester of South India" because of its intensive textile trade, cater by the encircling fields of cotton. However, the town has major environmental problems like contamination, lack of correct waste treatment facilities and downfall of territorial water catchment areas.

2.2. Parameters

Coimbatore has multiple sectors that square measure to blame for its air quality like many little and
medium scale enterprises, construction site and a rise in passenger vehicle transportation and union trade. Pollution is caused by the poisonous substances gift within the atmosphere that square measure in the main made by human activities. however typically it may end up from natural phenomena like mud storms and firestorm that depletes the air standard.

The Anthropogenic pollution causative agents are

1. Deforestation, extensive use of fossil fuels for numerous functions like electricity and road transport, produces air pollutants like element and sulphur dioxide.

2. Discharge and outrush from industries, releases great amount of monoxide, organic compound and harmful chemicals into the air.

3. Agricultural activities, thanks to the utilization of pesticides, and fertilizers emit harmful chemicals into air.

The above three sources are mainly due to increase in the population size and extended utilization of fossil fuel like coal for the micro and medium manufacturing industries.

Another parameter is that the PM2.5, particulate (PM) is suspension of a combination of liquid and solid particles present in the air. These may be ultrafine, fine or coarse suspension. Coarse particles measuring 2.5 micrometers to 10 micrometers are comparatively heavier hence tend to settle. The major source of particles are fuel burning resulting in chemical reaction that buildup in air. The environmental disasters and forest fire also conjointly contribute to PM2.5 within the air. These particles cause air pollution.

Road transport accounts for a major portion of pollution in Coimbatore, inflicting serious pollution issues like increase monoxide and cause air pollution. The rise within the use of personal cars, traffic pollution is taken into account as a serious threat to scrub air. Traffic miasma contain harmful chemicals that dirty the atmosphere. Therefore, it's terribly essential to notice the contributions of traffic information as a parameter whereas assessing the pollution. Gases such as Nitrous Oxides (NOx), fine material (PM2.5), Volatile Organic Compounds (VOCs), carbon-monoxide (CO), and sulphur dioxide (SO2) are the foremost common air pollutants emitted by vehicles. Within the town, several VOCs area unit emitted from evolution sources, like vehicle exhaust, industrial effluents, chemical plants, crude oil refineries, bound construction and mining.

The temperature and humidness of the town contributes to its pollution level that ought to even be endlessly be monitored. humidness affects pollution which makes it additional harmful to our metabolism health. humidness may be an important explanation for the forceful temperature changes in our immediate surroundings. It plays an important role when deciding our daily weather and climate conditions. The pollution concentration within the Coimbatore city may be due to prevailing wind speed, a performance of blending depth and topographical size of the town. The typical wind speed varies more or less on topographical basis from place to position and time to time. The cyclic is additionally regulating the concentration of pollution. The concentration of particulate pollutants throughout winter is on top of that in different seasons.
2.3. Air Quality Measurements Monitoring and Instrumentation

There are two air quality index (AQI) monitoring centres in Coimbatore. They measure the real time AQI of the city. A typical air pollution monitoring system consists of sensor which collects the air pollution data. The acquired data are then processed by a microcontroller. After this the data is sent to a server where data are updated and recorded continuously. Thus, the real time air pollution data can be monitored. First there is data collection and then data transmission. Different parameters for the air pollution are collected in this way.

3. Data Preparation

In this segment, the methodologies for devising of the information area unit bestowed, so as to proceed with the classification as shown in figure 5. This includes purification steps to eliminate useless knowledge, amendment to perceptibly investigate and perceive the information, and fabrication of a medium concentration map of the PM2.5 with regard to the temperature and wetness.

First, the collected knowledge is refined. the information refinement method is that the most vital half, as during this method knowledge is chosen, and graded, marked, classified and systemized properly to make prediction. This method converts piles of data into usable answers. As a result we have a tendency to get the smoothened knowledge that is a lot of correct. The smoothing of the information is completed with the assistance of the machine learning rule. This any ends up in removal of noise from the information and permits vital pattern to square out.

Second, knowledge transformation is completed so as to represent the information in Cartesian coordinates. This mathematical formulation incorporates a lot of correct feature illustration of the information. this is often helpful for classification models that area unit enforced within the following stage. This associated with machine learning models that upgrade performance as a result of theirs never-ending relationship between parameters. This transformation ensures each validity and a lot of informative illustration of the first knowledge.

![Figure 4. Various data collection](image-url)
4. Classification Models

A ML algorithm is employed to isolate the info in numerous categories of PM2.5 accumulation. The supervised learning techniques (SLT) area unit used to make models on this segregation task. Here we have a tendency to use a machine learning technique referred to as NARX (Nonlinear autoregressive exogenous model). NARX may be a continual dynamic system consisting of numerous enveloped layers of networks with feedback connections. The NARX model trust on the linear ARX model, that is usually employed in time-series modeling. The shaping equation for the NARX model is

\[ y(t) = f(y(t-1), y(t-2), \ldots, y(t-ny), u(t-1), u(t-2), \ldots, u(t-nu)) \]

Here, consecutive price of the dependent signaling \( y(t) \) is subsided on preceding values of the signaling and former values of Associate in Nursing freelance (exogenous) sign. we will implement the NARX model shown in figure 6 by employing a feed forward neural network to indefinite the operate. The picture of the ensuing network is shown in figure 6, wherever a two-layered feed forward network is employed for the approximation. This application additionally permits for a vector ARX model, wherever the input and output will be multi-dimensional.
For additional machine learning analysis, Regression Analyses should be implemented for both sides which is done based on Neural Networks (NN). NN modeling is chosen owing to the fact that it is better in case of highly nonlinear modeling where no prior insight regarding the interdependence of parameters is assumed. In addition, any error in the parameters may influence the prediction of weather condition. Hence it is noted that error related to parameters may provide insight into prediction confidence to determine weather condition. Also examining the trend of the collected date over time provide us with information about use of time series forecasting.

5. Matlab and ThingSpeak

Here we tend to use Matlab for implementing machine learning (ML). There are a unit completely different benefits of mistreatment Matlab for metric capacity unit like it extract options from signals and pictures mistreatment entrenched manual and automatic ways. It helps to match approaches like logistical declension, classification trees, support vector machines, chorus ways, and deep learning. From Applying Auto-ML and alternative model rarefaction and depletion techniques, we will produce enhanced models. It's straight forward to Integrate machine learning models into endeavor systems, clusters, and clouds, also targeting models to time interval embedded hardware. Thing Speak is one such application. Matlab also can accomplish mechanized code generation for embedded detector partition.

ThingSpeak is an ASCII text file net of Things (IoT) application and API to store and retrieve information from things or systems mistreatment the protocol and MQTT protocol over the net or via an area space Network. ThingSpeak permits the creation of detector work applications, location chase applications, and a social network of things with standing updates. It's integrated with Matlab, thus it's terribly straightforward for time period observation of the pollution information for hotspot detection. Further, assembling pollution information for a amount of your time we will discover the hotspot and so we will with success find the supply flight.

6. Result

As our proposed method is a software/algorithim approach based on climatologically and contamination data survey to forecast the concentrations of air pollution particle from breeze(speed and direction) and condensation levels to find the hotspot areas and further identify the changes and their source trajectory. The data being measured as shown in figure 7.

![Figure 7. Air quality index measurement](image)

AQI tracks 5 major air pollutants: atmosphere gas, monoxide, pollutant, gas, mobile particles, or aerosols as shown in figure 8. If the AQI is beneath fifty it means the air standard is sweet. At this low AQI pitch, someone will pay time outdoors and pollution can cause little menace to the health.
the AQI variety will increase, therefore will the chance to human health.

![Air quality data](image1)

**Figure 8.** Air quality data

Ground level off ozone, PM 2.5 and airborne crumb are the air pollutants that facade the greatest risk to human health. They are also the prime innards in smog, a sort of air contamination that diminishes panorama.

![Air quality measured in air quality index values](image2)

**Figure 9.** Air quality measured in air quality index values

Our probing has ameliorated the competency of the ML algorithm implemented based on percolate that subsist poise in both the meteorological and air contamination particulars as well as identifying as shown in figure 9 the passover meteorological data. This work has tendered a novel perspective for air pollution prophecy for Coimbatore based on both stationary and non-stationary contaminated sources using ML algorithm and demographic techiques.

![Temperature and wind speed](image3)

**Figure 10.** Temperature and wind speed
The supreme prophecy times were considered, which accomplished a consequential fidelity within 3 days. The pertinent data such as air pollution measurement were poised from ground stations on regular intervals. In the end, using the inclusive algorithm, the important parameters contributing to air pollution, data collected daily, weekly and monthly, topography, wind direction, maximum temperature, and the appraise of pollutants for the two selected imminent neighbour air pollution stations were bracketed. The temperature variations and wind speed data collected is shown in figure 10.

![Figure 10. Temperature and Wind Speed Data](image)

Figure 10. Temperature and Wind Speed Data

Appealing more factors such as Coimbatore congested traffic, size and few meteorological criterion worsening pollution in provenance are considered vital factors in casting and projection of air pollution peculiarly in commercial cities as shown in figure 11.

![Figure 11. Area Plot to Compare Traffic Data Speed](image)

Figure 11. Area Plot to Compare Traffic Data Speed

This work impart a perception into the foremost constraint concerning PM2.5 prognosis from barometric data and ML algorithm. The sorting and degeneracy shows that engrossment sizeable than 20 mg/m³ seem to be persuade more by subsidiary criterion than the barometric strand used. For example in figure 12, daily temperature shown and other strands like solar radiation, and provocation do not oscillate much all along the year, which may alter if analyzed during random hours of the day, engender various contamination level in the city.

![Figure 12. Histogram of Temperature Variation](image)

Figure 12. Histogram of Temperature Variation

![Figure 13. Location of the Hotspot](image)

Figure 13. Location of the Hotspot
By as well as the Coimbatore traffic information, distance from road shown in figure 11 and a few additional meteoric parameters associated with pollution sources causes some delusion within the eventual pollution divination. However, by victimisation error correction models, we will find the hotspot as shown within the figure 13. Some restricted places wherever no pollution information was offered, could slightly have an effect on the ultimate pollution prediction accuracy that couldn't be avoided thanks to some information scariness. Figure 13 shows the concentration of pollution, that’s some places square measure additional impure than alternative places.

![Air Quality health](image)

**Fig 14.** Result of air quality health

The final result of air peculiarity health is shown in the figure 14. The results of air peculiarity health can vary as the air pollution changes in the city.

7. **Conclusion**

This work tender a machine learning perspective to envision PM2.5 buildup from meteorologic information within the Coimbatore town. normal levels of fine particulate (PM2.5) square measure classified by victimization totally various machine learning replica. This segregation is staged on the records of day-to-day climatologic entries of wind velocity, wind direction, and precipitation accumulation from air quality scrutiny stations. The ML algorithm and therefore the multivariate perusal implies that the consider criterion will envision PM2.5 accumulation up to 20μg/m3 and therefore the fidelity of the forecast is additionally higher in constrains of sturdy winds and high cloudburst. The scarcely higher correlation throughout the time of year ensures that the model will indicate PM2.5 accumulation higher for a lot of farthest weather. Additionally, to the current, unearthing trends over bout of your time with the employment of your time series algorithms might enhance the augury and would build a semipermanent forecast of PM2.5 accumulation attainable. more it's attainable to predict the supply flight.

Here, our planned work provides a lot of reliable and a lot of economical various to forecast PM2.5 levels, because it solely needs meteorologic information acquisition. additionally, correct meteorologic technology is much cheaper collated to air quality sensors which will surpass the value over. Lastly, this work is predicated on the essential meteorologic criterion and PM2.5 that have a simple impact on pollution. Thus, on the whole that our replica has sensible forecast potency of Coimbatore town of such a posh topography, we tend to squabble that it can be with success imbibed in alternative tropical whereabouts.

Also, this grind provides Associate in Nursing cognizance into the most shortcomings concerning PM2.5 prediction from earth science knowledge and machine learning. The stratification and
declension indicate that concentrations but 20μg/m³ appear to be regulate additionally by further framework than the earth science strands employed in this study. Different weather conditions and weird gripping events inflicting lofty pollution levels like festivities, agristal fires, accidents, seasonal unevenness, or natural calamities might conjointly make a case for changes in PM2.5 engrossment prodigious 20μg/m³. In view of the various machine learning models employed in air calibre divination, like Neural Network, decay we tend to appertain and appraise most of those classifiers during this work. Various perspective to boost the fidelity of our model would encompass acting a conjecture supported Associate in Nursing ensemble of various algorithms of information process and modelling.

Therefore, the summary of the study as follows.

[1] Choosing the simplest applied mathematics replica and its furtherance for air-pollution conjecture.

[2] Choosing the simplest rarefaction technique for air-pollution and earth science knowledge so as to envisage the incomprehensible knowledge and percolate the noise of the information.

[3] Diagnosis of the foremost regulating parameters in pollution conjecture and notice to seek out to search out the hotspots therein space therefore find the supply mechanical phenomenon.

References

[1] Jan Kleine Deters, Rasa Zalakeviciute, Mario Gonzalez and Yves Rybarczyk 2017 Modeling PM2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters Journal of Electrical and Computer Engineering, 5, pp.1-14.

[2] X. Xi, Z. Wei, R. Xiaoguang et al 2015 A comprehensive evaluation of air pollution prediction improvement by a machine learning method in the 10th IEEE International Conference on Service Operations and Logistics, and Informatics, pp. 176–181, Hammamet, Tunisia.

[3] Y. Li, Q. Chen, H. Zhao, L. Wang, and R. Tao 2015 Variations in pm10, pm2.5 and pm1.0 in an urban area of the sichuan basin and their relation to meteorological factors Atmosphere, 6, no. 1, pp. 150–163.

[4] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayesb, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. Saffell, and R. Jones 2013 The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks Atmospheric Environment, 70, pp. 186–203.

[5] Djebbri, N and Rouainia, M. 2017 Artificial neural networks based air pollution monitoring in industrial sites in the International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23.

[6] Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin,L. and Wang, J 2015 Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2015.02.030.

[7] X. Ni, H. Huang, and W. Du 2017 Relevance analysis and short-term prediction of PM2.5
concentrations in Beijing based on multi-source data *Atmospheric Environment*, 150, pp. 146–161.

[8] J. Chen, H. Chen, Z. Wu, D. Hu, and J.Z. Pan 2017 Forecasting smog-related health hazard based on social media and physical sensor *Information Systems*, 64, pp. 281–291.

[9] J. Zhang and W. Ding 2017 Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong *International Journal of Environmental Research and Public Health*, 14 no. 2, p. 114.

[10] K. P. Singh, S. Gupta, and P. Rai 2013 Identifying pollution sources and predicting urban air quality using ensemble learning methods *Atmospheric Environment*, 80, pp. 426–437.

[11] C. Brokamp, R. Jandarov, M. B. Rao, G. LeMasters, and P. Ryan 2017 Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches *Atmospheric Environment*, 151, pp. 1–11.

[12] W. Sun and J. Sun 2017 Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm *Journal of Environmental Management*, 188, pp. 144–152.

[13] P. Saide, G. Carmichael, S. Spak et al 2011 Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model *Atmospheric Environment*, 45, Issue no. 16, pp. 2769–2780.

[14] J. K. Wendt, E. Symanski, T. H. Stock, W. Chan, and X. L. Du 2014 Association of short-term increases in ambient air pollution and timing of initial asthma diagnosis among medicaid-enrolled children in a metropolitan area *Environmental Research*, 131, Issue no. 0, pp. 50–58.