An Empirical Study of PM2.5 Forecasting Using Neural Network

Sachit Mahajan[1,2,3], Ling-Jyh Chen[2], Tzu-Chieh Tsai[3]
[1]Social Networks and Human Centered Computing Program, Taiwan International Graduate Program
[2]Institute of Information Science, Academia Sinica
[3]Department of Computer Science, National Chengchi University

***Abstract*—In the recent years, a lot of efforts have been made to regulate air pollutant levels in most of the developed and developing countries. Fine particulate matter (PM2.5) is considered to be one of the major reasons behind deteriorating public health and a lot of efforts are being made to keep a check on PM2.5 levels. Accurately forecasting PM2.5 level is a challenging task and has been highly dependent on model based approaches. In this paper, we explore new possibilities to hourly forecast PM2.5 . Choosing the right forecasting model becomes a very important aspect when it comes to improvement in prediction accuracy. We used Neural Network Autoregression (NNAR) method for the prediction task. The paper also provides a comparative analysis of prediction performance for additive version of Holt-Winters method, autoregressive integrated moving average (ARIMA) model and NNAR model. The experimentation and evaluation is done using real world measurement data from Airbox Project, which shows that our proposed method accurately does the prediction with significantly low error.**

***Keywords*—PM2.5, Forecast, Artificial Neural Network, ARIMA, Holt Winters.**

I. Introduction

Whenever we talk about Smart Cities, the first thing that comes in the mind is extensive usage of Information and Communication technology to provide the people a better quality of life. Good air quality can be linked directly to good quality of life. There are many ways in which we can make sure that the city's air quality is good; by continuously monitoring the air quality using smart sensors and by developing systems which can forecast the air quality. The air quality is always changing because of the variations in the emissions from different sources. So , one day it can be normal and the other day it can change drastically. This makes it even more important to come up with a method to precisely forecast the future air quality. Air pollution prediction and forecasting has always been a challenging task, especially when we talk about PM2.5(Fine Particulate Matter with diameter less than 2.5 micrometers). Recently a lot of research has been done to understand the adverse effect of PM2.5 on human health. And it has been found out that it can actually penetrate into the lungs and cause severe health problems like lung cancer, asthma and other respiratory diseases [1]. PM2.5 monitoring has been happening all over the world and there are many organizations such as Environment Protection Agencies (EPA), National Oceanic and Atmospheric Administration (NOAA) in the U.S. which regularly keep a check on the pollutants level. Many cities monitor the changes in the air quality on a regular basis. To get a complete understanding of variations in air quality it is important to have detailed knowledge about the strength of emission sources, which is not easy to get. Because of these reasons, more research has to be done in development of air quality forecasting models that use limited parameters and yet come up with accurate predictions. Most the air pollution modelling approaches depend on mathematical simulations. These methods rely on atmospheric dynamics which are captured by various generative models. The simulations are run to make predictions [2]. Sufficient and accurate data is needed to predict the air quality which in turn helps in controlling the air pollution in Air Quality Management (AQM). The approaches to predict particulate matter have been categorized into different ways: using empirical models for prediction, using fuzzy logic based models, using models based on simulation, data driven statistical models and statistical learning methods which are model driven [3].The other approach is to use air pollution modelling software. But the disadvantage of using them is that sometimes the results are not accurate as they don't consider all the factors effecting air pollution [4]. In contrast to the conventional approaches, our method is very data centric. We make use of the historical data obtained by the Airbox Project to do the experimentation.

The contributions of our paper is three-fold:

1) We present a prediction model for predicting PM2.5 for the next one hour with high accuracy.

2) We conduct a comparative analysis of our model with other forecasting models to compare the prediction accuracy.

3) We use real world data obtained from Airbox project and evaluate our model's predictions over other stations which highlights the performance of our model.

The rest of the paper is organized as follows. In Section II, we discuss about related works. In Section III, we discuss about the Airbox Deployment . In Section IV, we elaborate the Airbox data and also discuss about data visualization systems. In Section V, we discuss the methodology by explaining in detail the models used for the prediction task. In Section VI, we implement the models on the data and analyse the results. We also perform a comparative analysis of all the models used in order to come up with the best model. Then we evaluate our final model by performing prediction for 50 stations in the Taichung area, Taiwan. We compare the observed PM2.5 values and the predicted PM2.5 values. We also compare our results with the other state of the art models proposed by other researchers to show higher accuracy of our system. In Section VII we conclude the paper and give some directions about the future work.

II. Related Work

In the last few years, a lot of particulate matter forecasting models have been proposed in order to keep a check on air quality across different parts of the world. Satellite remote sensing techniques have been used to assess PM2.5 where surface PM2.5 monitoring sensors are not available [5]. Several authors have used Neural networks for predicting PM2.5. In [6], authors have proposed a feed forward neural network forecasting model. It forecasts hourly PM2.5 value. Another work [7] uses a Deep Hybrid Model for weather forecasting. It doesn't forecast PM2.5, but predicts other variables like wind, dew point, temperature. In [8], authors use a generative approach in which the weather system is simulated via numerical method.In one of the works [9], authors have proposed a model combining wavelet transformation and air mass trajectory to improve neural network forecasting accuracy. It forecasts daily PM2.5 concentration two days in advance.Some studies [10] have used a hidden markov model to forecast daily concentrations. The authors were able to perform the prediction for Concord and Sacremento, California. Other than the statistical models, research has been done in doing the prediction using chemical models. In [11], the authors proposed a chemical transport model which could provide space-time continuous estimate of pollutants including PM2.5. Their model used a hybrid approach which chemical transport model as well as regression model.

Prediction has also been done using time series analysis such as ARIMA models and classifiers based on ANN [12,13].Then there are some models which use Support Vector Machines to perform the prediction [14]. Although machine learning has been widely used in lot of tasks but it hasn't been exploited very widely when it comes to PM2.5 predict for small intervals of time. Most of the related work mentioned above predicts PM2.5 on a daily basis or on an hourly basis. In [15] , the authors tried to predict pollutants like NO2, PM10 and o3 concentrations for the next 30 minutes by using spatial and temporal factors. In [16], the authors did a comparative analysis of five linear models to predict PM10 concentration mean on a daily basis. We try to address the limitations of the existing works by using machine learning techniques in order to perform the hourly forecasting.

III. Airbox Deployment

Under the Airbox Project, Airbox devices have been deployed on a large scale throughout Taiwan for PM2.5 monitoring. This project aims to motivate people around the country to get themselves involved in PM2.5 sensing. The inspiration behind this project is the Location Aware Sensing System (LASS) community that engages local people in sensing PM2.5 and also motivates them to design PM2.5 sensing devices by themselves. The most important aspect of such kind of system is that the PM2.5 can be monitored at a very fine spatio-temporal granularity. Also the data can be easily accessed by anyone which makes the data analysis easy [17]. In Airbox Project, the sensing devices are developed by professional manufacturers which makes sure that the devices are reliable. Also during the deployment it was made sure that the places had a continuous power supply and internet connection.

The deployment of Airbox devices began in Taipei City. On March 22, 2016, 150 devices were deployed in the city.

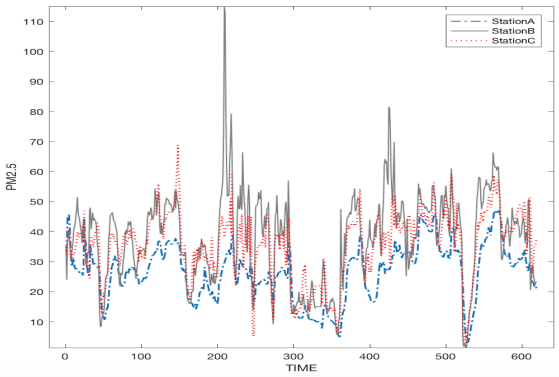

Fig. 1. Hourly representation of PM2.5 ($\mu g/m^3$) for 3 stations

Since then the project has been widely acknowledged by the everyone and spread over other major cities in Taiwan. 230 devices have been deployed in Taichung ; 242 devices have been deployed in Kaohsiung; 298 devices have been deployed in New Taipei and 220 devices have been deployed in Tainan. As of now there are more than 1500 devices installed in 20 cities in Taiwan and 24 countries all over the world.

IV. Data Set

The measurement data collected by the Airbox Devices for Taichung area during the time period January 18, 2017 (0:00 hours) and February 12, 2017 (18:00 hours) was used to perform the data analysis. As most of the devices donated were installed at the elementary schools with regular power source and internet connection, the measured data is expected to be reliable and of better quality. So, we only considered those datasets for doing the data analysis. Also, it was observed that most of the devices in the school follow a weekly pattern. Number of devices online is more during the weekdays than during the weekends. The reason might be that most of the schools switch off all the electronic equipments in order to save electricity during the weekends.

According to the manufacturers, the data sampling frequency for the Airbox devices is claimed to be every five minutes. However, it was found out that inter-sample time was around six minutes for 80% of the devices and for rest of them it was around 12 minutes. The standby time between collecting samples of an Airbox device is five minutes and it takes around one minute to perform the sampling. So it makes the inter-sampling time to be six minutes. If there is an error and the first measurement of data fails, then the inter-sampling time becomes 12 minutes. For our research, we converted the data to hourly data. The data sampled at 6 minutes interval doesn't provide a lot of information about the variations in the PM2.5 level. On the other hand , with hourly data one can see the fluctuations in the PM2.5 level over the time.

In Fig. 1 , we see the hourly PM2.5 data for Station A (Latitude 24.057 and Longitude 120.699), Station B (Latitude 24.145 and Longitude 120.693) and Station C (Latitude 24.104 and Longitude 120.727). It can be easily observed that there are huge fluctuations in the PM2.5 levels at different time periods for different stations. This makes it even more important and useful to have a forecasting model with high accuracy.

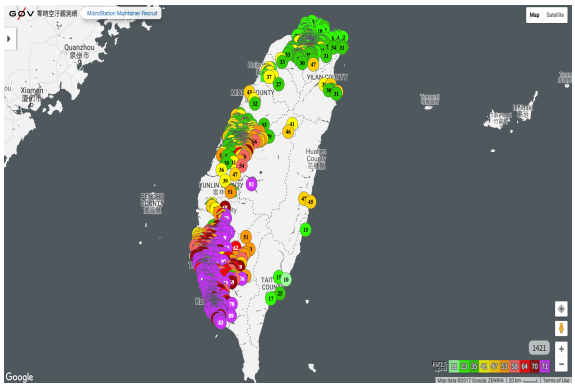

Fig. 2. GIS visulaization provided by g0v community

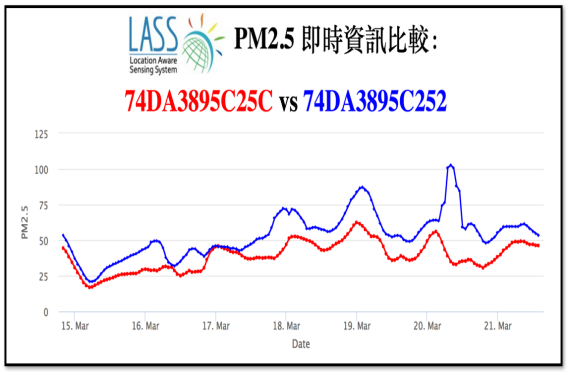

Fig. 4. Comparison of data from two Airbox Devices

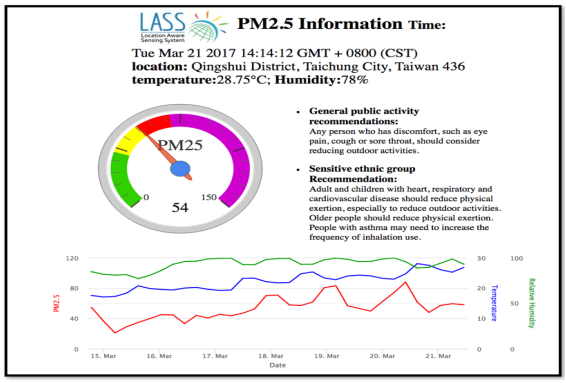

Fig. 3. Dashboard for visualization of Airbox Data

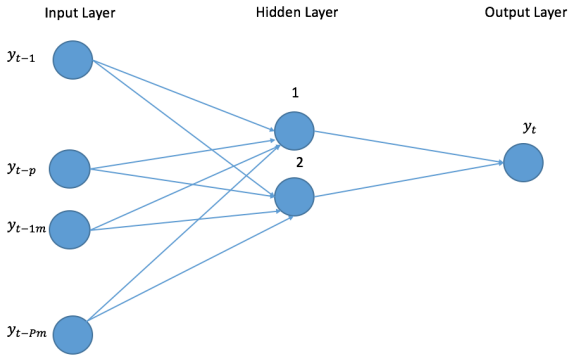

Fig. 5. Representation of NNAR (p, P, k)m model

A. Airbox Data Applications

The data from Airbox Devices has been used to develop data visualization systems. Fig. 2 shows a geographic information system (GIS) based visualization method which shows the data on a map by combining the measured data and the location data. There are visualization systems which focus on displaying information about every Airbox device in detail (Fig. 3).

In Fig. 3, it can be seen that the visualization platform provides the information about the specific Airbox Device and also gives detailed information about historic data.It also provides information about temperature and relative humidity levels. Another way to visualize Airbox data is to compare two devices. Fig. 4 shows a comparison analysis of different devices on the same day. So this makes it easy to look for data for any device and interpret the data.

V. Methodology

We applied different forecasting models and then came up with the best one after doing the comparative analysis. The models are discussed in detail in the following paragraphs.

A. Autoregressive Integrated Moving Average Model (ARIMA)

There has been an extensive usage of ARIMA models for time series prediction and forecasting purposes. The main reason for its wide use is that it is easy to understand and implement. Also, the robustness of this model is a plus as compared to other models for time series forecasting. ARIMA forecasting is done by first identifying the model, parameters estimation, diagnostic check performance and considering other models based on the result of diagnostic check. The appropriate model is found which is then used for the task of forecasting.

An ARIMA (p, d, q) model comprises of p, d and q which are integers, greater than or equal to zero and refer to the order of the autoregressive (AR), integrated (I) and moving average (MA) parts of the model [18]. If there is a time series X_t, where t denotes an integer index and X_t denotes real numbers which correspond of values at a given time t , then an ARIMA (p,d,q) model can be described by the following equation

$$(1-B)^d(1-\sum_{i=1}^{p}\phi_l B^l)X_t = (1+\sum_{i=1}^{p}\theta_l B^l)\varepsilon_t \qquad (1)$$

In the above equation, B is the backward shift operator, ϕ_l and θ_l denote the parameters of autoregressive and moving part respectively. ε_t is the error term. If in case d=0 then, the ARIMA model is an ARMA model.

B. Neural Network Autoregression Model (NNAR)

Artificial Neural Networks (ANN) have been widely used for time series forecasting. It has been possible to model complex relationships between input variables and output variables through ANNs. In Neural Network Autoregression Model, the

input to the model is in the form of lagged time series values and the output gives the predicted values for the time series.In [19] , authors proposed the NNAR (p, P, k)m model. p and P denote the lagged seasonal and non seasonal inputs and k is the number of nodes in the hidden layer and m donates seasonality. The detailed structure of the NNAR (p, P, k)m model is shown in the Fig. 5. For modelling time series, the model consists of two functions, a linear combination function and an activation function(sigmoid) . The linear combination function is given as

$$y_t = a_t + \sum_{i=1}^{n} w_{i,t} x_t \tag{2}$$

a_t and $w_{i,t}$ are obtained from the data and x_t are the lagged time series values. The activation function is given as

$$f(y_t) = 1/(1 + e^{-y_t}) \tag{3}$$

For this research, we consider feed-forward neural networks which is based on the Nonlinear Autoregressive Model for time series forecasting.An NNAR(p,k) model actually shows that there are p lagged inputs and k nodes in the hidden layer of the model.

C. Holt-Winters (HW) Forecasting Model

Holt-Winters method is a statistical forecasting method that is used for time series data. It uses historical trends to predict future values. We use Additive Holt-Winters method for prediction. In this method there are several parameters involved [20].The components include overall smoothing denoted by x_t, s_t is the seasonal factor and y_t gives the trend. The equation belows shows the relationship between the three parameters.

$$x_t = \alpha(n_t - s_{t-T}) + (1 - \alpha)(x_{t-1} + y_{t-1}) \tag{4}$$

$$y_t = \gamma(x_t - x_{t-l}) + (1 - \gamma)(y_{t-1}) \tag{5}$$

$$s_t = \beta(n_t - x_t) + (1 - \beta)(s_{t-T}) \tag{6}$$

$$n_{t+h} = x_t + hy_t + s_{t-T+h} \tag{7}$$

n is the number of observations. α, β and γ are assigned values for smoother forecasting. These values directly impact the prediction behavior. n_{t+h} in equation (7) gives the prediction value for the instant $t + h$. h can be varied as per the requirements.

VI. Results and Evaluation

Initially we considered the data from Station A (Latitude 24.057 and Longitude 120.699) in Taichung, Taiwan. Hourly PM2.5 data from January 18, 2017 (0:00 hours) to February 12, 2017 (18:00 hours) was considered. There were in total 620 observations. 600 of them were used for training and the rest for testing. Three models described in the Methodology section were applied to the data to see which one produces the best result with more precision.

The parameters to compare the results from three models were chosen to be Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) . The calculation were made by using the following equations:

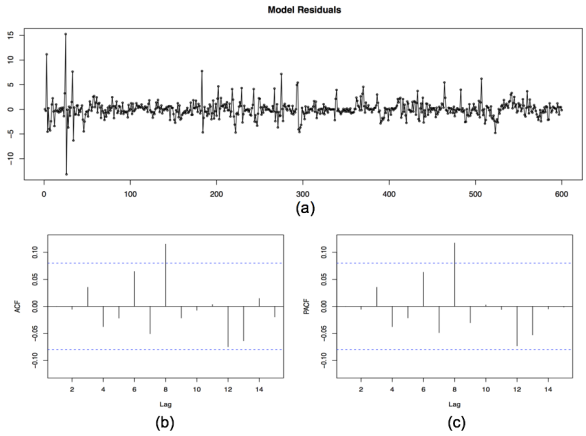

Fig. 6. Plots of (a) Fitted ARIMA Model Residuals (b) ACF and (c) PACF

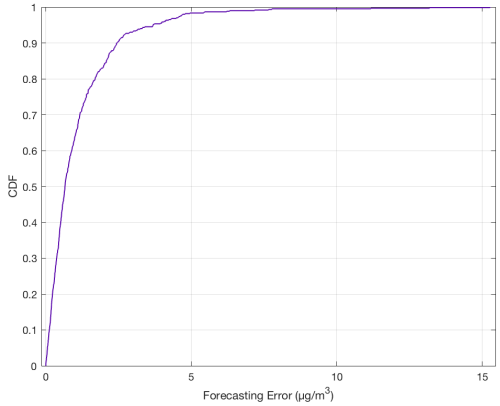

Fig. 7. The CDF plot for fitting with ARIMA model

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(x_{p,i} - x_{m,i})^2}{n}} \times 100 \tag{8}$$

In the above equation, x_p and x_m are predicted and measured values respectively.

$$MAE = 1/n \sum_{i=1}^{n} |y_m - \hat{y_m}| \tag{9}$$

In Equation (9), y_m is the actual value and $\hat{y_m}$ is the predicted value. The results obtained after the implementation of different models are discussed below.

A. Result with ARIMA

As we already discussed in Section IV , an ARIMA mode consists of three parameters (p, q and d). We used ARIMA (3,1,1) model here for the prediction. Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) can give an idea about whether the series is stationary or non - stationary. Fig. 6 shows the final model residuals, ACF and PACF. It can analysed that there is some unusual behavior during the initial part. Other than that , the remaining residuals give us an idea that the model has captured the pattern quite closely.

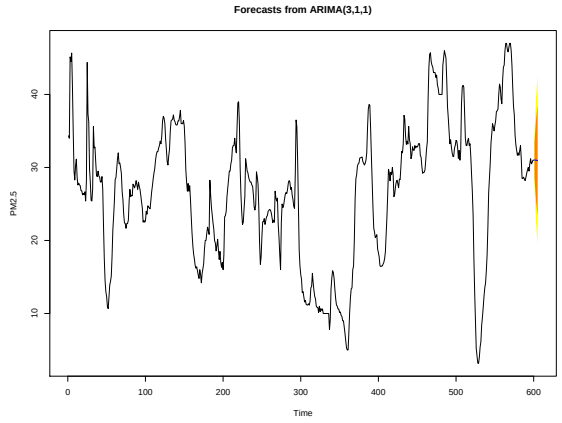

Fig. 8. PM2.5 ($\mu g/m^3$) forecasting with ARIMA model

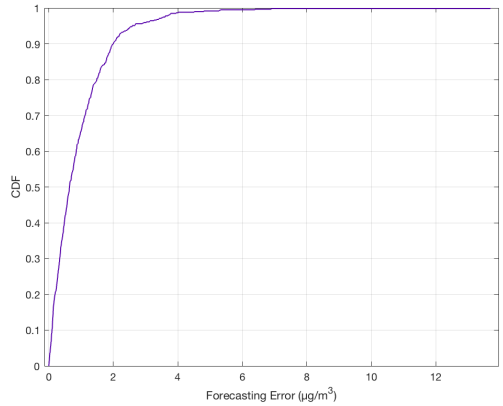

Fig. 9. The CDF plot for fitting with NNAR model

Fig. 7 shows the Cumulative Distribution Function (CDF) plot for fitting obtained with ARIMA model on the training data. It can be observed from the plot that around 90% of the observations have an error under $2.5\mu g/m^3$. We can infer from this plot that the fitting is very close. Fig. 8 shows the PM2.5 forecasting with ARIMA model for the next 5 hours.

B. Result with NNAR

We used an NNAR(9,5) model here to do the forecasting. The model uses 9 lagged inputs and has 5 nodes in the hidden layer. As we are using NNAR(9,5) model,so the model uses the previous nine values to predict the next value. In Fig. 9, it can be observed that 90% of the observations have an error below $2\mu g/m^3$. So we can say that the fitting observed with NNAR model is better as compared to what we achieved with ARIMA model. Forecasting plot for next 5 hours using NNAR model can be seen in Fig. 10 .

C. Result with Holt-Winters Model

Using Holt-Winters model, the value of α was found to be 1. This meant that the recent values had more weight for smoothing the level component. β value was calculated to be 0.25 which meant that the older values were weighted more to smooth the trend component.

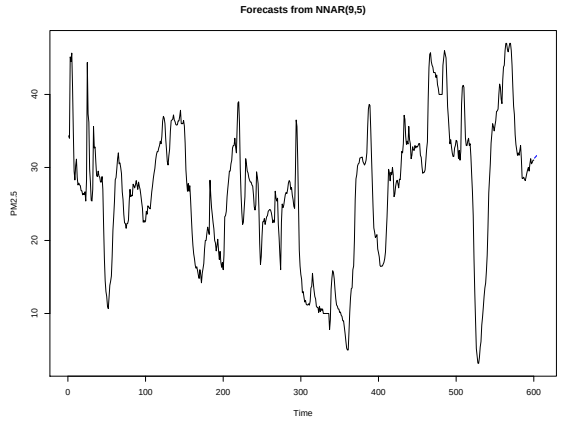

Fig. 10. PM2.5 ($\mu g/m^3$) forecasting with NNAR model

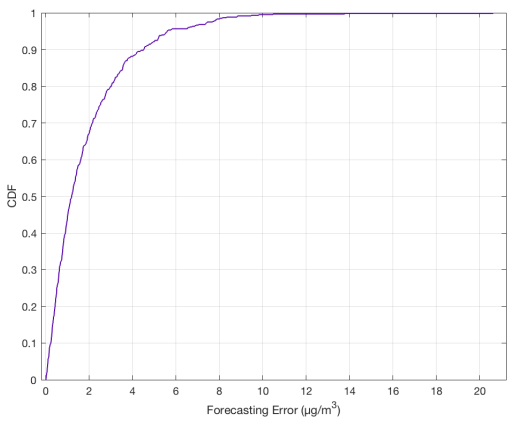

Fig. 11. The CDF plot for fitting with HW model

From the CDF plot in Fig. 11, it can be observed that the fitting with Holt Winters model is slightly less as compared to the other two models. 90% of the observations have a forecasting error below $4\mu g/m^3$ which is more than what we achieved with ARIMA and NNAR models. The next step shows forecasting done using HW method. The result is shown in Fig. 12.

So finally after testing three different models on the data set, we get an idea about the accuracy of all three models in forecasting hourly PM2.5 ($(\mu g/m^3)$). After doing the comparative analysis of three models using RMSE and MAE, it can be seen that NNAR model has the lowest values for both RMSE and MAE. The results can be seen in TABLE I. This implies that the accuracy of NNAR model is more as compared to other two models.

TABLE I. RMSE AND MAE CALCULATIONS FOR 3 MODELS

Comparative Analysis			
Models	Data	RMSE	MAE
ARIMA	Training	1.82	1.11
	Testing	1.98	1.75
NNAR	Training	1.40	0.93
	Testing	1.58	1.47
HW	Training	1.99	1.28
	Testing	1.75	1.63

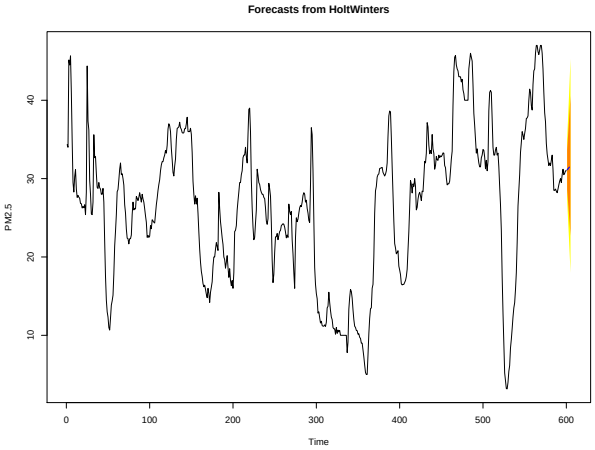

Fig. 12. PM2.5 ($\mu g/m^3$) forecasting using HW model

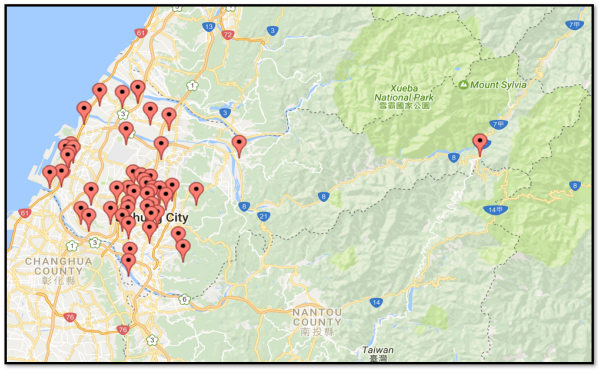

Fig. 13. Geographical location of 50 stations

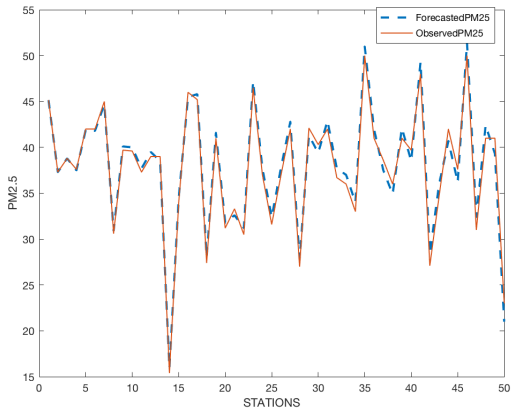

Fig. 14. Plot for Observed and Forecasted 1 hourly PM2.5 ($\mu g/m^3$) at 50 stations

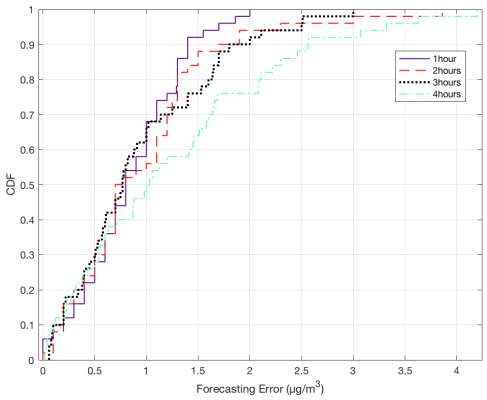

Fig. 15. The CDF for error between Observed and Forecasted PM2.5 at 50 stations for different time periods

D. Evaluation

To do the evaluation of our final model, we first tested the NNAR model on 50 stations in Taichung area of Taiwan. The geographical location of the fifty stations is shown in Fig. 13. We forecasted the PM2.5 value for next 1 hour and compared it with the original observed value. The result is shown in the Fig. 14. It can observed from the figure that the forecasted PM2.5 values are very close to the observed PM2.5 values. The difference is significantly low. At some stations, the observed and predicted values are exactly the same.

Additionally we tried forecasting the PM2.5 value for next four hours for 50 stations. We did it using a sliding window method. The forecasted observation was taken and used to train the model and simultaneously get a forecast value for the next hour. We calculated the CDF (Commulative Distribution Function) and plotted it in Fig. 15 for forecasting for different durations of time (1 hour, 2hours, 3 hours and 4 hours). CDF was calculated for fifty forecasts, one for each station. It can be observed that in case of 1 hour forecasting, for around 90% of the stations the error is less than 1.3 $\mu g/m^3$ and the maximum error between the observed and forecasted PM2.5 values is less than 2 $\mu g/m^3$. On the other hand, for forecasting after 2 hours ; around 88% of the stations showed error under 1.5$\mu g/m^3$. Forecasting for 3 hours showed that 90% of the stations had an error under 2 $\mu g/m^3$. Similarly, for forecast after 4 hours, it showed that almost 90% of the stations had an error under 2.5$\mu g/m^3$. So, we can see that the error in forecasting with NNAR model is significantly low in all the cases, lowest being in the case for next hour forecast.

In order to further check the performance of our model, we compared our results with other state of the art previous works. The first one which was proposed by [21] uses a semi-empirical model that predicts hourly PM2.5 concentration using satellite remote sensing and meteorological measurements. Our second comparison was with the model proposed by [22] which forecasts hourly roadside PM2.5 concentration using first-order and one variable grey differential equation model. TABLE II shows the comparison of the accuracies of these two models with our model.

TABLE II. COMPARISON OF PROPOSED METHOD WITH OTHER STATE OF THE ART WORKS (HOURLY FORECASTING)

Models	RMSE
Proposed Model	1.58
Tian et al. 2010	6.1
Pai et al. 2011	3.40

VII. Conclusion and Future Work

There is a big concern about the effect of high concentration of pollutants in the air and its adverse effect on human health. Many a times the pollutant concentration level increases rapidly. In order to alert people, it is important to do short term forecasting of pollutants. In this paper, we addressed the issue of hourly PM2.5 prediction. We proposed a model to accurately predict PM2.5 for the next hour. Using the Airbox data , we implemented different models in order to check the performance and then came up with the conclusion that NNAR based model has better prediction accuracy as compared to ARIMA based model and Holt Winters based Model. Our results were based on RMSE and MAE calculations. We then evaluated our model by testing it on 50 stations in Taichung area of Taiwan. The result showed significant prediction accuracy for all the stations. Also we made a comparison of our proposed model's performance with other state of the art works and it showed that our model's forecasting accuracy is better than other works.

This work can be extended to do forecasting for other cities as the PM2.5 level is different for different cities because of the pollution sources. Another future task can be forecasting for longer durations like 12 hours or daily forecasting. The results would be very beneficial for agencies monitoring air quality and environment policies.

References

[1] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian. The impact of PM2.5 on the human respiratory system. In *Journal of Thoracic Disease*, vol. 8, no. 1, pp. 6974, January 2016.

[2] G. Marchuk. Numerical Methods in Weather Prediction. Elsevier, 2012.

[3] Dong, G.H., Zhang, P., Sun, B., Zhang, L., Chen, X., Ma, N. Long term exposure to ambient air pollution and respiratory disease mortality in Shenyang, China: a 12 year population - based retrospective cohort study. In *Respiration*, 84(5), pp.360-368, 2012.

[4] Harsham, D.K., Bennett, M. A sensitivity study of validation of three regulatory dispersion models. In *American Journal of Environmental Sciences*, 4(1),pp.63-76, 2008.

[5] Walter Di Nicolantonio, Alessandra Cacciari and Claudio Tomasi. Particulate Matter at Surface: Northern Italy Monitoring Based on Satellite Remote Sensing, Meteorological Fields, and in-situ Samplings. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume: 2, Issue: 4,pp.284 - 292, December 2009.

[6] Patricio Perez and Ernesto Gramsch. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. In *Atmospheric Environment*, Volume 124, Part A, pp. 22-27, January 2016.

[7] Aditya Grover, Ashish Kapoor and Eric Horvitz. A Deep Hybrid Model for Weather Forecasting. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 379-386 , 2015.

[8] Lewis Fry Richardson. Weather Prediction by Numerical Process. Cambridge University Press, 2007.

[9] Xiao Feng, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin and Jingjie Wang. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. In *Atmospheric Environment*, Volume 107, pp. 118-128,2015.

[10] Wei Sun, Hao Zhang, Ahmet Palazoglu, Angadh Singh, Weidong Zhang and Shiwei Liu. Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. In *Science of The Total Environment*, Volume 443, pp.93-103, 2013.

[11] Qian Di, Petros Koutrakis and Joel Schwartz. A hybrid prediction model for PM2.5 mass and components using a chemical transport model and land use regression. In *Atmospheric Environment*, Volume 131, pp. 390-399, 2016.

[12] Cyril Voyant , Marc Muselli , Christophe Paoli and Marie Laure Nivet. Numerical Weather Prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. In *Energy*, 39(1):341-355, 2012.

[13] Ling Chen and Xu Lai. Comparison between ARIMA and ANN Models Used in Short-Term Wind Speed Forecasting. In *Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE*, pp. 1-4, 2011.

[14] Nicholas I. Sapankevych and Ravi Sankar. Time Series Prediction Using Support Vector Machines: A Survey. In *Computational Intelligence Magazine, IEEE*, vol.4, no.2, pp.24-38, May 2000.

[15] Arie Dipareza Syafei, Akimasa Fujiwara, and Junyi Zhang. Prediction Model of Air Pollutant Levels Using Linear Model with Component Analysis. In *International Journal of Environmental Science and Development*, Vol. 6, No. 7, July 2015.

[16] J.C.M. Pires, S.I.V. Sousa, M.C. Pereira, M.C.M. Alvim-Ferraz and F.G. Martins . Management of air quality monitoring using principal component and cluster analysisPart I: SO2 and PM10. In *Atmospheric Environment*, Volume 42, Issue 6, Pages 12491260, February 2008.

[17] LJ Chen, W.Hsu, M.Cheng and H.C.Lee. LASS: A Location-Aware Sensing System for Participatory PM2.5 Monitoring. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*, Pages 98-98, 2016.

[18] Charisios Christodoulos, Christos Michalakelis and Dimitris Varoutas . Forecasting with limited data: Combining ARIMA and diffusion models. In *Technological Forecasting and Social Change*, Volume 77, Issue 4, Pages 558-565, 2010.

[19] Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. OTexts, 2013.

[20] Kondratyev M., Tsybalova L.. Long-Term Forecasting of Influenza-Like Illnesses in Russia. In *International Journal of Pure and Applied Mathematics*, Vol. 89. No. 4. P. 619-642, 2013.

[21] Jie Tian and Dongmei Chen. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. In *Remote Sensing of Environment*, Volume 114, Issue 2, , Pages 221-229, 2010.

[22] Tzu-Yi Pai et al. . Using Seven Types of GM (1, 1) Model to Forecast Hourly Particulate Matter Concentration in Banciao City of Taiwan In *Water, Air, Soil Pollution*, Volume 217, Issue 1, pp 25-33, 2011.