Extensive Evaluation of Neural Network Based PM_{10} Prediction Models Using Openair Package

A. Suleiman[1], M.R. Tight[2] and A.D. Quinn[2]

[1, 2] School of Civil Engineering, University of Birmingham, Birmingham, West Midlands, B15 2TT, United Kingdom

Abstract

The results of air quality models have huge implications on decisions related to nation's budget, public health, environment and transport policies, therefore it is imperative that they are evaluated properly to achieve the desired outcome. The performance of air quality prediction models are often evaluated using statistical performance metrics. In this paper, Bivariate polar plots, Polar annulus plots, Time variation plots, Conditional quantile plots, Taylor's diagrams and Scatter plots in addition to the model performance statistics were used to evaluate the performance of Artificial neural network based PM_{10} prediction models. The models were developed using three variants of multilayer feed forward neural networks (i.e. Multilayer Perceptron Network (MLPN), Radial Basis Function Neural Network (RBFNN) and Generalised Regression Neural Network (GRNN)). The data for the modelling consisting of air pollutants, traffic and meteorological variables was collected from a street canyon in central London (Marylebone Road). The models test results showed that RBFN model was the most accurate of the three ANN variants considered having its fraction of predictions within the factor of two (FAC2), Coefficient of correlation (R), Coefficient of efficiency (CoE), Root Mean Squared Error (RMSE) and Mean Bias (MB) of 0.99, 0.94, 0.69, 7.13 and -0.69 respectively. When the result was further analysed with the graphical functions mentioned above revealed that the extreme cases of high and low pollutant concentrations were slightly underestimated by the models though improved significantly with the use of historic average of the PM_{10} as an input variable. The models also performed better in spring and summer, weekdays and daylight hours, but slightly under predicted the PM_{10} in autumn, winter, and during low traffic flow. The study demonstrated that ANN based models can be used to predict PM_{10} concentrations in street canyons with greater degree of accuracy and an in-depth analysis using openair package tools boosted our confidence in recommending the use of the ANN based models for street canyon air quality studies.

Keywords: Neural network, Air quality, Performance evaluation, Openair package.

1.0 Introduction

Transport is one of the major vehicles for socioeconomic development within and outside our immediate environment. However, it is the major source of air pollution which affects negatively the inhabitants of our cities and their respective environments. Diseases such as Asthma, Bronchitis, Pneumonia, and respiratory infections are common to the residents located near major roads (Brauer et al., 2002, Kim et al., 2004, McConnell et al., 2006, Lindgren et al., 2009, Heinrich et al., 2005). Brunekreef et al. (2009) discovered that long-term exposure to particulate and gaseous pollutants emitted by traffic have strong link to respiratory mortality. The effects of air pollution can be effectively reduced through provisions of adequate and effective air quality control and mitigation measures which are designed and tested using air quality models. The environmental regulatory agencies have to supplement air quality measurements with models that are able to predict pollutants concentrations and determine the cause of the air quality problems. The use of such models provides opportunity for using historic data to study the past scenarios of air pollution episodes and to forecast the likely pollution events for the future. The air quality models currently in practice require sufficient knowledge of the atmospheric processes and high computational and operational cost. Considering the growing need for the cheaper and reliable air quality models, and also the accumulation of air quality data over the years, artificial intelligent techniques such as artificial neural networks(ANN) can be used to build air quality models with comparable prediction accuracy at a lesser computational cost (Gardner and Dorling, 2000). ANNs have been successfully applied in many air quality studies (N Sharma, 2005, Amirsasha Bnanankhah, 2012, Singh et al., 2012, Kukkonen et al., 2003, Kumar and Goyal, 2013, Russo et al., 2013, Yan Chan and Jian, 2013, Zhang et al., 2013). ANNs can handle air quality variables which are complex and nonlinearly related (Esplin, 1995) and produce models that can perform extremely well in predicting future occurrences. ANN is a machine learning tool that is designed to mimics biological neural system. It acquires knowledge through training and produces outputs based on the knowledge of the relationship between the input variables within the training data. Although ANN models possesses the qualities required for air quality modelling, it is imperative they are evaluated extensively using different performance criteria and compare their performance with the existing air quality models in order to ensure their acceptability by the air quality modellers and regulatory agencies.

Models performance are subject to a human bias judgement which are often minimised through the use of quantitative tools to estimate suitable numerical metrics that will give overall performance of the model. The quantitative tools are often preceded by using graphical methods that represents the sensitivity of the model in the form of graphs, charts, or surfaces, which can be used together with the results of mathematical and statistical methods for better representation (Fei et al., 2005). The statistical tools often used for evaluating models include systematic and non -systematic root mean square error ($RMSE_S$ and $RMSE_U$), mean bias error (MBE), index of agreement (IA), bootstrap estimates of confidence and significance(Willmott et al., 1985). Nagendra and Khare (2006) used systematic and unsystematic root mean square error ($RMSE_S$ and $RMSE_U$), mean bias error (MBE), mean square error, coefficient of determination, linear best fit constant (*a*) and gradient (*b*), mean of the observed and predicted concentration ($\bar{O}$ and $\bar{P}$) and their standard deviations (δO and δP) respectively, and descriptive statistics 'd' values. Yap and Karri (2013) used the mean relative error (MRE), root mean square error (RMSE) and the correlation coefficient (R^2) in comparing ANN model and dynamic model for predicting engine power and tailpipe emission for a 4kw gasoline scooter. Kukkonen et al. (2003) evaluated artificial neural network, a linear model and a deterministic model using the index of agreement (IA), the squared correlation coefficient (R^2) and the fractional bias (FB). Detail analysis on statistical parameters for evaluating models can be found in Willmott et al. (1985), Willmott (1981) and Petersen (1997).

The aim of this paper is to compare the performance of PM_{10} prediction models developed using three multilayer feed-forward neural networks (i.e. MLP, RBF and GRNN). The models will be evaluated using model evaluation functions provided in an open source air quality data analysis tool "*openair package"* (Carslaw and Ropkins, 2012) of R statistical software (Development CoreTeam, 2013). The package has three functions dedicated for model evaluation which include: Model statistics, Taylor's diagram and Conditional Quantile diagram functions. In addition, Bivariate polar plots, Time variation plots, Scatter plots and Polar annulus plots functions provided in the package can also be used to explore the models performance with respect to the original relationship between the observed data and wind directions, wind speeds and traffic volume and other meteorological variables. The functions allowed the model predictions to be evaluated at a predefined time units e.g. hourly daily, weekday, season or annually. The rest of the paper is divided into three sections. Section two describes theoretically the three ANN variants and the evaluation methods used in this study. In section three, the results of the models evaluation are presented and discussed and section four is the conclusion of the study.

1.1 Study area

The data for this study was collected from two air quality monitoring sites (i.e. kerb and background sites) for a period between 2000 and 2007. The kerb site is located approximately 1.5m away from Marylebone road in central London. The road is housed in a regular street canyon with an aspect ratio (H/W) of approximately 0.8. It consists of three lanes in each direction with an annual average daily traffic of more than 80,000 vehicles and the outside lane of each direction has been dedicated to only buses and taxis (Jones and Harrison, 2006).The data consists of hourly concentrations of gaseous and particulate pollutants, traffic volume, vehicle composition, vehicular speeds and meteorological variables measured simultaneously at the site. The background concentrations of gaseous and particle pollutants were collected from an urban background site located within the south east corner of Russell Square Gardens in central London (London Bloomsbury site). These sites are part of UK's Automatic Urban and Rural Network (AURN).

1.2 Data

The traffic data at Marylebone road were collected using induction loops buried on each lane with an estimated accuracy of 99% for counting and classification. The instruments used for monitoring PM_{10} at the sites include two similar Tapered Element Oscillating Microbalance (TEOM) Model 1400AB with different sampling heads design (Aurelie and Harrison, 2005). The TEOM consists of a filter, tapered hollow glass tube and PM_{10} impactor inlet for measuring PM_{10} mass (Aurelie and Harrison, 2005). The traffic data collected was aggregated into Heavy duty (HDV) and Light duty (LDV) vehicles for all the six lanes, and their corresponding hourly averages were summed up to obtain the average hourly volume for heavy duty, light duty and the total vehicles flow on the road. The emissions for the hourly Light duty vehicles(LDV) and Heavy duty vehicles(HDV) were estimated using emission factors derived by Jones and Harrison (2006). The total hourly input variables prepared for the ANN models include: Background PM_{10}, 24 hours rolling mean of PM_{10}, Carbon monoxide (CO), Nitric oxide (NO), Sulphur dioxide (SO_2), Oxides of Nitrogen (NO_x), Nitrogen dioxide (NO_2), Wind direction, Wind speed, Temperature, Solar Radiation, HDV emission strength, HDV volume, LDV emission strength, LDV

volume, total traffic volume, Barometric Pressure, Relative Humidity, average traffic speed, and Rainfall. The target or response variable is the hourly roadside PM_{10} concentrations.

2.0 Methods

In this section, the three variants of artificial neural networks (i.e. MLP, RBF, and GRNN), the model development procedure and the evaluation methods used in this paper are briefly described. The modelling exercise considered two separate data compositions. 1. The data containing historic average (24 hours rolling mean) of PM_{10} observations and 2. The data without historic average of PM_{10}. In each case, the data collected for a period between 2000 and 2005 was allocated for model training and from 2006 to 2007 was allocated for testing the models. During training, the training data for each case was fed into an ANN toolbox in MATLAB R2014a software and the models were trained several times using various number of hidden neurons until a model with a lowest mean squared error (MSE) was obtained. The trained models were then tested using the testing data sets and the results were evaluated using model evaluation functions of *openair* package (Carslaw and Ropkins, 2012) of *R* statistical software(Development CoreTeam, 2013).

2.1 Artificial neural networks

Artificial Neural Network (ANN) models are Machine learning methods designed to mimics the behaviour of the part of human brain that are capable of learning and storing information about their environment (Bishop, 1995). The ANN architecture consists of layers of neurons that are sequentially connected from input layer, hidden layer to output layer respectively, and the data provided to the network is processed in the order in which they are arranged. The output of each layer serves as an input to the next layer and the output of the output layer is the overall output of the model (Božnar, 2011). A Neuron is made up of three elements, the connecting links characterised by its weight, a summation unit U_k which combines the weighted input signals, and an activation function $f(.)$ for translating the input signals into output signals. Mathematically a neuron say k, can be represented as follows:

$$U_k = \sum_{j=1}^{m} w_{kj} x_j \qquad j = 1,2,3, \ldots, \tag{1}$$

$$y_k = f(U_k + b_k) \tag{2}$$

where $x_1, \ldots, x_m$ *are the input signals,* $w_{k1}, .., w_{km}$ *are synaptic weights of neuron* k, U_k *is the linear combiner ,* b_k *is the bias,* $f(.)$ *is an activation function and* y_k *is the output signal of the neuron* k.

The type of ANN described above with one or more hidden layers is sometimes called multilayer perceptron network (MLPN). During MLPN training, an appropriate training algorithm is invoked to train it on the training data. The most commonly used algorithm is supervised back propagation algorithm. The ANN training, using this algorithm involved two major processes (i.e. forward and backward passes). In forward pass, the predictor variables received by the input neurons are passed through connecting links of various weights with which the predictor variables are weighted. The weighted predictor variables are then summed up by a linear combiner and transmitted forward to the hidden layer neurons. The outputs of the hidden layer estimated by its activation functions are then passed onto the output layer where the final output of the neural network will be estimated. The activation functions in the hidden layer are mostly differentiable nonlinear functions e.g. sigmoid function. They are used in transforming the input space into a higher dimensional space, thereby translating them into output signals. The Sigmoid transfer function shown in Equation 3 was used for developing Multilayer perceptron networks (MLPN) in this research.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

The network outputs are then compared with the target response, and their difference is taken as the network error, which will then be propagated backward to update various weights within the neural network (i.e. backward pass). The iteration continues until minimum error is obtained (Keng, 2010). The main difference between MLPN and RBFN is in the manner in which the outputs of the hidden

units are computed. RBFN network computes the Euclidean norm distance between the input vector and the centre of the RBF units as shown in Equation 4, whereas in MLPN, the activation functions in each hidden layer neuron computes the inner product of the input vectors and their respective synaptic weight vectors. Another important difference is that, in MLPN, the input-output mapping is carried out to give global approximations while RBFN networks construct the local approximations to nonlinear input-output mapping. This is an indication that RBFN require much more parameters than MLPN to achieve the same level of accuracy (Haykin, 2005).

$$y(X) = \sum_{j=1}^{m} w_j \, G\left(\|X - t_j\|\right) + \, b \qquad (4)$$

Where; y is the network output, X is the input vector,$t_j(\mathrm{j} = 1, 2, \ldots, \mathrm{m})$ is the centre of the RBFs, w_j is the weight vector, $G\left(\|X - t_j\|\right) = exp\left(-\frac{\|X-t_j\|^2}{2\delta^2}\right)$ and b is the bias.

The generalised regression neural network (GRNN) is a multilayer feed forward neural network which consists of four layers namely; input layer, the hidden layer, summation layer and an output layer. The outputs of the input neurons are fed into the hidden layer where the Euclidean distance between the centres of the neurons and the input vectors are estimated. And then apply RBF kernel function (e.g. Gaussian function) to estimate the output of the hidden layer neurons which are then passed to the summation layer. The summation layer contains two neurons one each for summing the numerator and the denominator components of the Equation 5. The numerator summation unit sums up the weight values and multiplied by the actual target value y of each hidden neuron while the denominator summation unit sums up the weights values coming out of each neuron. The results of the summation units are then passed to the output layer where the numerator values are divided by their corresponding denominator values for each case, and the result is the final output of the network. GRNN uses the concepts of conditional probability to arrive at the optimal estimation of continuous variables. The joint probability density functions (*pdf*) of the predictor variables X and the target variable y estimated using nonparametric estimation is used to determine the conditional *pdf* and the expected value. GRNN operates in one pass to achieve the desired estimation of the target variable not in an iterative manner as in MLPN.

$$E[y|X] = \frac{\int_{-\infty}^{\infty} yf(X,y)\,dy}{\int_{-\infty}^{\infty} f(X,y)\,dy} \qquad (5)$$

Where the density $f(X,y)$ is usually estimated from a sample of observations x and y.

2.2 Models evaluation functions

This section briefly describes the graphical and statistical models evaluation functions used in this study. The graphical functions include Scatter plot, Conditional quantile plot, Time variation plot, Taylor's diagram, Bivariate polar plot, and Polar annulus plot. The model statistics function include fraction of predictions within the factor of two (FAC2), Coefficient of correlation (R), Coefficient of Efficiency (CoE), Root Mean Squared Error (RMSE), Mean Bias (MB), Normalised Mean Bias (MB), Mean Gross Error (MGE) and Normalised Mean Gross Error (NMGE).

2.2.1 Scatter Plots

The Scatter plot is a simple visual technique which plots the prediction and observation pairs to show how they relate to one another. Visually, the model's over or under prediction can be easily detected and quantified. In *openair package*, the scatter plot function estimates the linear relationship between the observed and predicted values, the coefficients of determination, slopes and intercepts of the linear equations, and 95% confidence intervals of the fit.

2.2.2 Conditional Quantile Plots

The Conditional quantile plot is a simple way of looking at a model performance against real world observations for a continuous measurements (Wilks, 2011). The conditional quantile plot function in an *openair* divides both prediction values and observations into bin pairs of equal length and estimates the median, 25/75th and 10/90th percentile of each bin, the estimates are then plotted to show how predicted and observed values agree with one another. This plot differs from quantile – quantile plot in that for a particular interval, it does not use the distribution of the observations and prediction separately, but it uses the corresponding values of the observations in predictions. This plot is particularly important in revealing how well the distribution of predictions tally with that of the observations especially at lower and upper part of the distribution(Carslaw and Ropkins, 2012).

2.2.3 Time variation plots

Time variation plots are useful tools in describing how pollutant concentrations vary with time (i.e. hour of the day, day of the week, weekly and monthly etc.). In air pollution studies, these plots could be used to reveal information about the likely sources of the emission. In *openair package*, the time variation function produces four time scale variation plots: combined hour of the day and day of the week, mean hour of the day, day of the week and monthly variation plots. It can also show the 95% confidence interval in the mean values. The uncertainty intervals are calculated using bootstrap resampling that would provide a better estimate than assuming normality especially for small data. There is also an option for normalising the data which is useful when plotting data with different units. The normalisation is done by dividing the concentration or any other variable by their mean values. These plots can also be useful in comparing the model predictions and the observations to observe how the model predictions agree with the observations on a time scale (Carslaw and Ropkins, 2012).

2.2.4 Taylor's Diagram

Tylor's diagram is a useful tool for comparing graphically the performance of various models. It shows three model performance metrics; the correlation coefficient, standard deviation and centred RMSE. Taylor (2001) showed that it was possible to relate these statistics through the use of law of cosines on a 2D graph. Taylor's diagram should not be used alone because the values of the metrics used do not depend on the mean bias. They only measure unsystematic errors; therefore, a model might systematically over or under predict but still has the same scatter as the observation yielding a perfect match for standard deviation (Chang and Hanna, 2004).

2.2.5 Bivariate Polar Plots

Bivariate polar plots of pollutants concentrations have been used to discriminate against various types and characteristics of emissions sources (Carslaw et al., 2006). These plots describe the joint variation of pollutant concentrations, wind speeds and wind direction on a continuous surface using polar coordinates. Given a set of pollutant concentrations wind speed and wind directions, suitable wind speed-direction bins are obtained, and a mean concentration for each bin is estimated. The wind components *u* and *v* are estimated using Equation 6 and the pollutant concentrations formed the surface of the polar plot.

$$u = \bar{u}\sin\left(\frac{2\pi}{\theta}\right), \qquad v = \bar{u}\cos\left(\frac{2\pi}{\theta}\right) \tag{6}$$

The generalised additive model (GAM) is then applied to fit the surface in order to extract the real source characteristics rather than the noise (Carslaw and Beevers, 2013). A detail description of this plot can be found in Carslaw and Beevers (2013).

2.2.6 Polar annulus plot

The polar annulus plot describes the temporal variation of pollutants with respect to wind directions. It provides a system in which to visualise the trend, seasonal, diurnal and day of the week variations.in this paper we used polar annulus plots to display diurnal variations.

2.2.7 Coefficient of correlation R

Coefficient of correlation R is a measure of co-linearity between observed and modelled values often reported as the coefficient of determination (R^2) which indicates how much of the total variation in the observation is explained by the model prediction. The values of R range between minus one and one. A perfect model is required to have an R value of one while a value near zero indicates little or no relationship between the modelled and observed variables. R is estimated using Equation 7.

$$R = \frac{\sum_{i=1}^{N}(M_i - \bar{M}) \times (O_i - \bar{O})}{\left[\sqrt{\sum_{i=1}^{N}(M_i - \bar{M})^2}\right]\left[\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2}\right]} \tag{7}$$

Where $\bar{M}$ and $\bar{O}$ represents the mean modelled and observed values respectively.

Although this metric may appear attractive, its magnitudes are not consistently related to the predictive ability of the model especially when the prediction accuracy is the degree to which the model predictions match the magnitude of the observations (Willmott, 1982). It is a usual practice to report the statistical significance associated to this metric to aid in interpreting the correlation coefficients; however, Willmott and Wicks (1980) demonstrated that statistical significance of R may be misleading since they are not related to the sizes and differences between observed and modelled values. Willmott (1982) discourage the use of R as part of the model performance measures due to ill-defined relationship and inconsistencies between the values of R and the model performance. However, several model evaluation tools include this metric as part of the array of model performance measures and is still being reported in recent researches involving model evaluation (Bennett et al., 2013, Thunis et al., 2012, ASTM, 2010, Carslaw and Ropkins, 2012).

2.2.8 Root Mean square error (RMSE)

Root mean square error (RMSE) is a measure of the average error produced by a model and is among the best measures of overall model performance which can be easily interpreted since they carry the same unit as the modelled and observed values. Although it is sensitive to extreme values, it reveals the actual size of the error produced by the model unlike R that is affected by the higher and low standard deviations of both observed and modelled values. However it does not reveal the types or sources of the error which will assist greatly in refining the models (Willmott, 1982, Willmott, 1981). RMSE value varies between 0 and ∞, with 0 being RMSE for an ideal model. It is estimated using Equation 8.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2} \qquad (8)$$

2.2.9 Fraction of predictions within a factor or two (FAC2)

The Fraction of predictions within a factor or two (FAC2) measures the fraction of the model prediction that satisfies the condition in Equation 9. A model which satisfies the condition would have a value of FAC2 equal to 1. Chang and Hanna (2004) described FAC2 as a robust performance measure since it is not affected by outliers.

$$FAC2 = 0.5 \leq \frac{M_i}{O_i} \leq 2.0 \qquad (9)$$

2.2.10 Mean Bias (MB)/ Normalised mean bias(NMB)

The Mean Bias (MB) is the measure of the model under or over prediction estimated as the difference between the mean observed M_i and the mean predicted values O_i . It is estimated using Equation 10. MB values range from -∞ to + ∞ with zero being MB value for an ideal model. Although MB is being used as model performance measure its major weakness is that it does not provide more diagnostic value than the mean values of observed and predicted. Willmott (1982) suggested that the mean values of observed and predicted should be reported instead of MB since they are more familiar to the researchers and contain little more information than MB. NMB (i.e. Equation 11) is a normalised version of MB, and it is often used when comparing different pollutants concentration scales.

$$MB = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) \qquad (10)$$

$$NMB = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{M_i - O_i}{\sum_{i=1}^{N} O_i}\right) \qquad (11)$$

2.2.11 Mean Gross Error (MGE)/ Normalised mean gross error (NMGE)

Mean Gross Error (MGE) is a measure of model error regardless of whether it is under or over prediction and it has the same unit as model and observed values. Normalised mean gross error (NMGE) has the same interpretation as MGE with added advantage when comparing pollutants of different unit scales. They are estimated using Equation 12 and 13.

$$MGE = \frac{1}{N}\sum_{i=1}^{N}|(M_i - O_i)| \qquad (12)$$

$$NMGE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|(M_i - O_i)|}{\sum_{i=1}^{N} O_i}\right) \qquad (13)$$

2.2.12 Coefficient of Efficiency (CoE)

Coefficient of efficiency CoE is the measure of model efficiency that is robust and easy to interpret (Legates and McCabe, 2012). This measure has interpretation for zero and negative values. A perfect model has COE value of one. Zero values of COE indicate that the model's prediction accuracy is not more than the observed mean values of the data and negative COE values indicates that the model's prediction accuracy is worse than the observed mean. It can be estimated using equation 14.

$$COE = 1 - \frac{\sum_{i=1}^{N}|(M_i - O_i)|}{\sum_{i=1}^{N}|(O_i - \bar{O}_{i_i})|} \qquad (14)$$

3.0 Results and Discussions

This section presents the discussion on the test results of the three variants of ANN (i.e. MLPN, RBFN and GRNN) models developed in this study. The models were tested for predicting the PM_{10} concentrations collected at the Marylebone road site between 2006 and 2007 and their performance statistics are shown in Table 1. The MLPN model with 24 hour rolling mean of PM_{10} in the input variables (MLPN-Hav) and the two RBFN models showed excellent performance with 99 percent of their prediction within the factor of two (FAC2) of the observed PM_{10} and the root mean squared errors (RMSE) of 7 to 7.5 µg/m^3. They also have their coefficient of efficiency (CoE) values approximately 0.7 which means that the models prediction is approximately 70 percent more accurate than the mean value of the observed PM_{10}. The Normalised Mean Gross Error (NMGE) values shows that the gross errors for the three models are approximately 12 percent. However the two GRNN and MLPN models are the least performing models with quite lower performance statistics than the models mentioned above. The improvement in the performance of the models is evident due to the use of historic average of PM_{10} as one of the predictors. All the models have slightly under predicted the observed PM_{10} concentrations with normalised mean bias error (NMB) ranging between 2 and 4 percent. The coefficient of correlation (R) values for the models ranges between 0.9 and 0.94 which indicates that the models prediction correlates well with the observed PM_{10} concentrations.

Table 1: Model Performance Statistics

The prefix *Hav* indicates the models that includes the historic average (24hour rolling mean) of the observed PM_{10} in the input variables

Model	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
RBFN_Hav	9080	0.99	-0.69	4.74	-0.02	0.12	7.13	0.94	0.69
MLPN_Hav	9080	0.99	-0.72	4.76	-0.02	0.12	7.49	0.93	0.69
RBFN	9080	0.99	-0.66	4.96	-0.02	0.12	7.48	0.93	0.68
GRNN_Hav	9080	0.98	-1.92	6.20	-0.05	0.16	8.98	0.91	0.60
GRNN	9080	0.98	-1.77	6.77	-0.04	0.17	9.61	0.90	0.56
MLPN	9080	0.95	-0.86	7.02	-0.02	0.18	10.30	0.88	0.55

To further analyse the relative variation of the standard deviation, coefficient of correlation, and centred root mean squared error the Taylor's diagram became handy and was plotted for the four UK seasons as shown in Figure 1. Taking all the seasons into consideration, the MLP model shows various degree of overestimation indicated by the black dashed lines in the diagram, but in winter it has shown nearly the same standard deviation with the observed concentrations. Its correlation coefficient is lower than the remaining models while its cantered root mean squared error is consistently higher than that of the remaining models in all the seasons. The two GRNN models showed consistent underestimation much further away from the dashed lines throughout the seasons. The two RBFN and the MLPN_Hav models are the best models having consistently being closer to the dashed lines with higher correlation coefficient and lower centred root mean squared error. They showed slight under estimation in spring, autumn and winter, but in summer, the MLPN_Hav showed slight overestimation while the two RBFN models have the same standard deviation with the observed concentrations. Generally the models performance did not show significant seasonal variations.

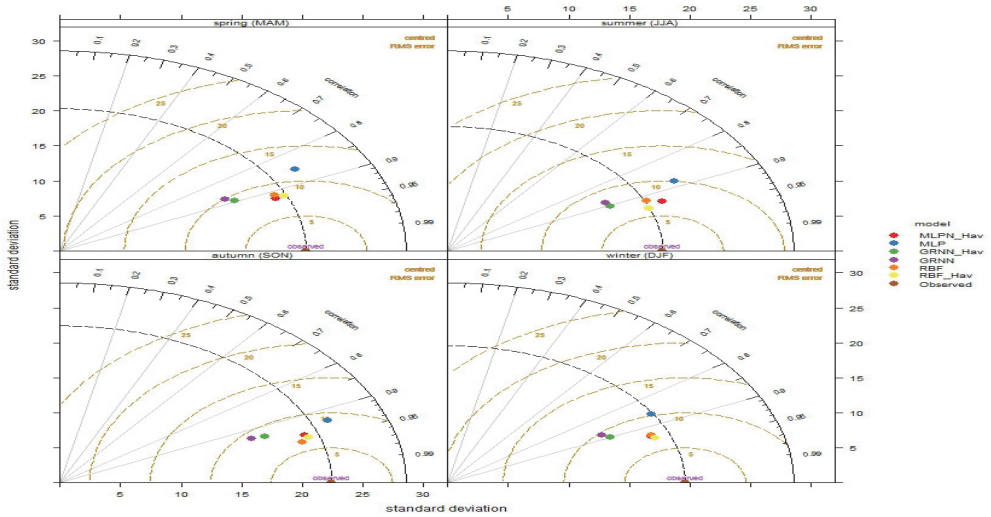

Figure 1: Taylor's Diagram showing models performance

3.1 Common Air Quality Statistics

The models have slightly under predicted the PM_{10} concentrations as indicated by the performance metrics above, however, it was not clear to what extent this will affect the credibility of the model predictions. Therefore, it is important to check the models accuracy in predicting the common air quality statistics as shown in Figure 2. The mean and median of the observed and modelled concentrations were approximately the same. The largest error occur in the estimation of maximum and minimum concentration however, all the values fall within the same air quality index band with the observed values. The RBFN and MLPN models performed well with slight underestimation of the days with PM_{10} concentrations greater than 50 $\mu g/m^3$ which is the daily limit that should not be exceeded more than 35 days a year according to EU directive 2008/50/EC on ambient air quality. The maximum daily values were either overestimated or underestimated by 1 index band but are all within the same description (i.e. High). The MLPN and RBFN models have their statistics much closer to the observed than the GRNN models and their predictions mostly fall within the same index band with the observed concentrations.

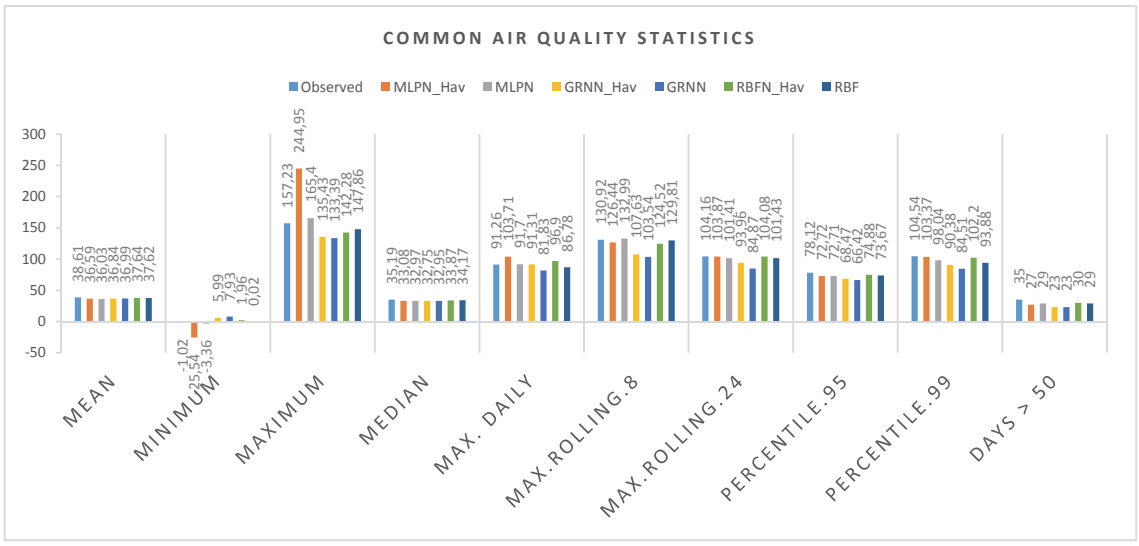

Figure 2: Air Quality Metrics for PM_{10} in 2007

3.2 Analysis of the spread of the modelled concentrations

The scatter plots in Figure 2 shows that the MLP and GRNN predictions have larger spread around the lines at the centre than the remaining models. This shows the extent of their over and under predictions. The bulk of the observed PM_{10} concentrations lies within 0 to 80 µg/m^3 and the models performed well in capturing these range of values. The GRNN models have more points under the continuous line at the centre which shows that they have underestimated most of the points and much more severe towards the higher values. The RBF models have much narrower prediction spread as such, they can be judged as the most accurate methods among the three methods under consideration.

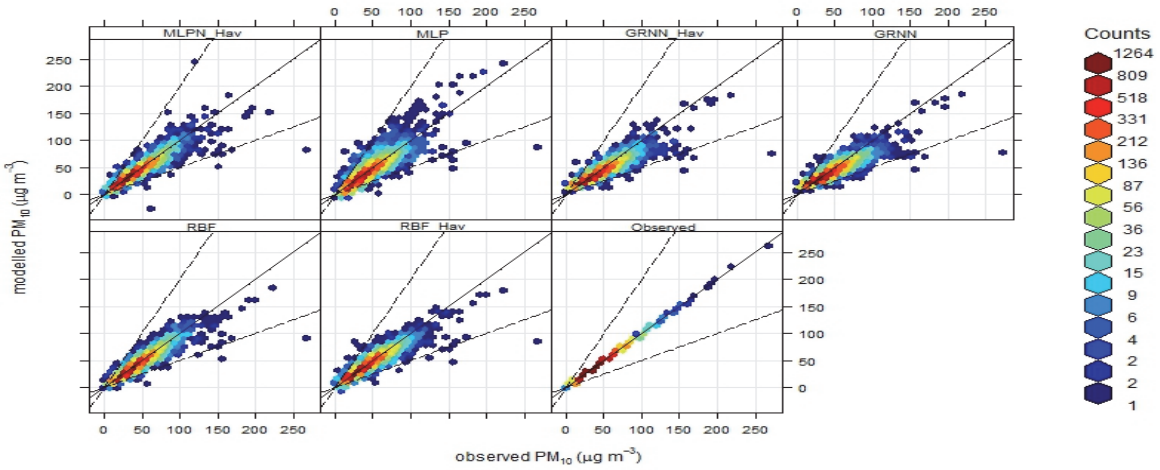

Figure 3: Correlation between modelled and observed PM_{10}

In Figure 3, the two dashed lines indicates the boundary of the factor of two (FAC2) and the continuous line at the centre shows the 1:1 line for the PM_{10} observations and the key shows the frequency of the values separated by the colour scale.

Figures 4 and 5 are the conditional quantile plots showing the overall and seasonal performance of the models respectively. The plots show how the models prediction agree with the observed values and more importantly how the minimum and maximum values are captured by the models as they are the main concern for regulatory purposes. The two RBFN models and MLPN_Hav performed well in predicting the lower and higher values. The shaded portion of the plot indicates the quantile intervals of the predictions and the models have narrow spread which is an indication of their degree of agreement with the PM_{10} observations. The MLP model performed poorly and its predictions barely captures more than 70 µg/m^3 accurately. Although the performance statistics showed that the MLPN and GRNN models performed poorer than the remaining models, it was not clear where the models failed. However looking at Figure 4 it is glaringly clear that the models failed in capturing extreme cases of PM_{10} concentrations.

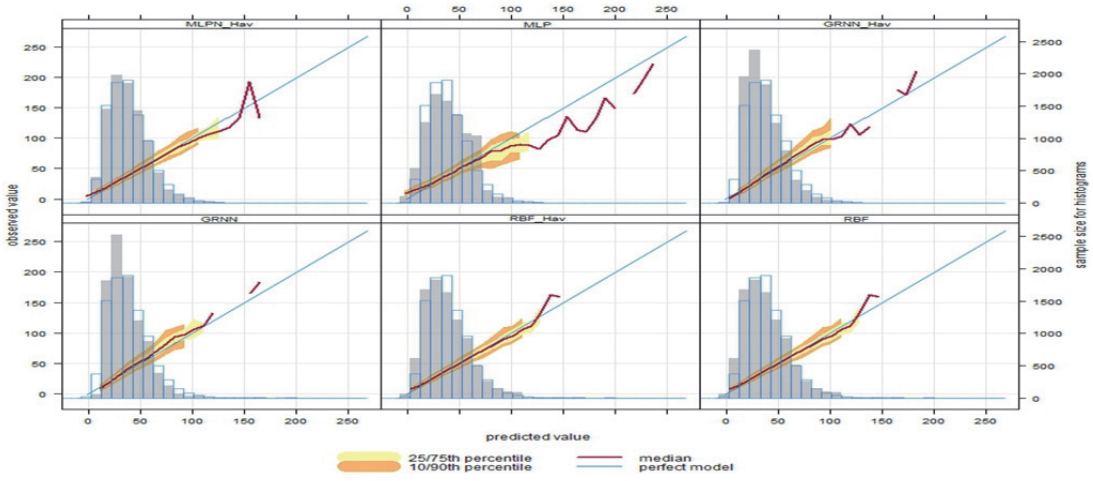

Figure 4: Conditional quantile plots comparing modelled and observed PM_{10}

It is a well-established fact that the air pollutants are sensitive to seasonal variations, therefore, it important to examine the models response to the seasonal changes in the pollutant concentrations. Figure 5 shows that the highest concentration was observed in autumn with about 160 µg/m^3 followed by summer and winter with about 150 and 140 µg/m^3 respectively. During spring, the MLPN_Hav and the two RBFN models performed well with slight under predictions in the region of higher concentrations. The remaining models barely predicted more than 100µg/m^3 accurately and their values have larger spread indicated by the quantile intervals. In summer, the MLPN model is the least performing model while GRNN and RBFN performed fairly well in predicting concentrations below 100 µg/m^3. The RBFN_Hav was the best model in this season because it captured the higher concentrations accurately. In autumn, the MLP and the two GRNN models did not predicted the concentrations above 100 µg/m^3 accurately, while the MLPN_Hav and the two RBF models were slightly better in that respect. In winter, the models were more accurate in capturing concentrations below 100 µg/m^3 with RBFN having more accurate predictions towards the region of higher concentrations. In general, the models prediction reflects the seasonal changes existed in the observed concentrations especially in terms of data coverage but mostly with slight underestimation of higher concentrations.

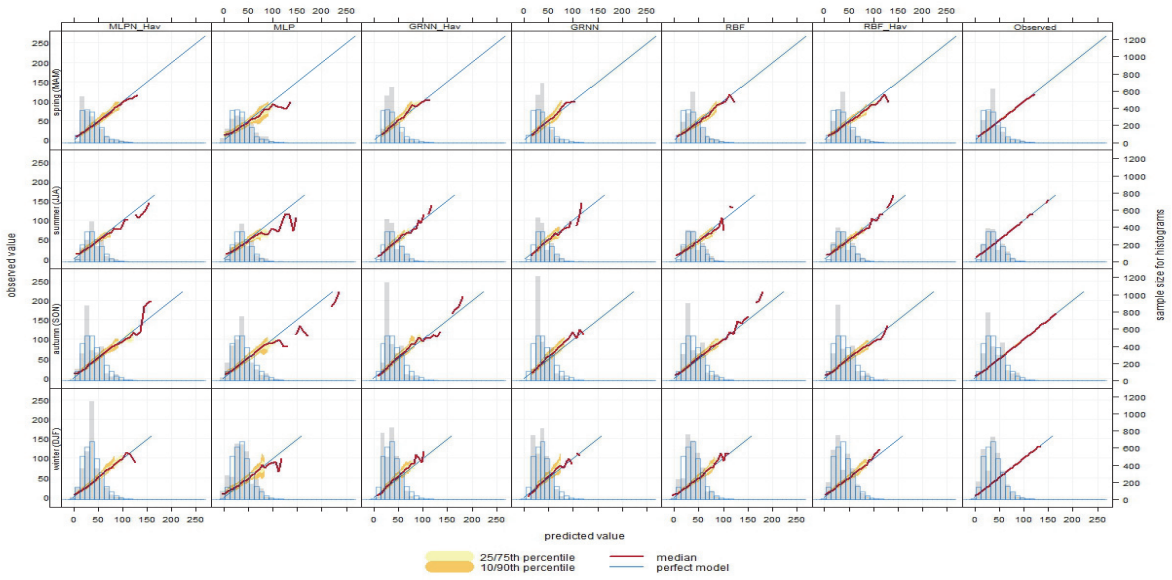

Figure 5: Conditional quantile plots showing the seasonal models performance.

3.3 Analysis of Temporal and Spatial variations

Temporal variation is an important characteristics of roadside pollutant concentrations as they tend to vary with the traffic flow on the road. The Polar annulus plots shown in Figure 6 depicts the temporal variations captured by the ANN models. The observed PM_{10} concentrations were lower between 12:00AM and 6:00AM and much higher from 8:00AM to 6:00PM. The higher concentrations were shown to be more associated with the winds coming from Southwest, South and Southeast. The models have accurately reflected these properties in their predictions most especially the RBFN and MLPN model. However the GRNN models were shown to have under predicted the higher concentrations though adequately captured the temporal variations.

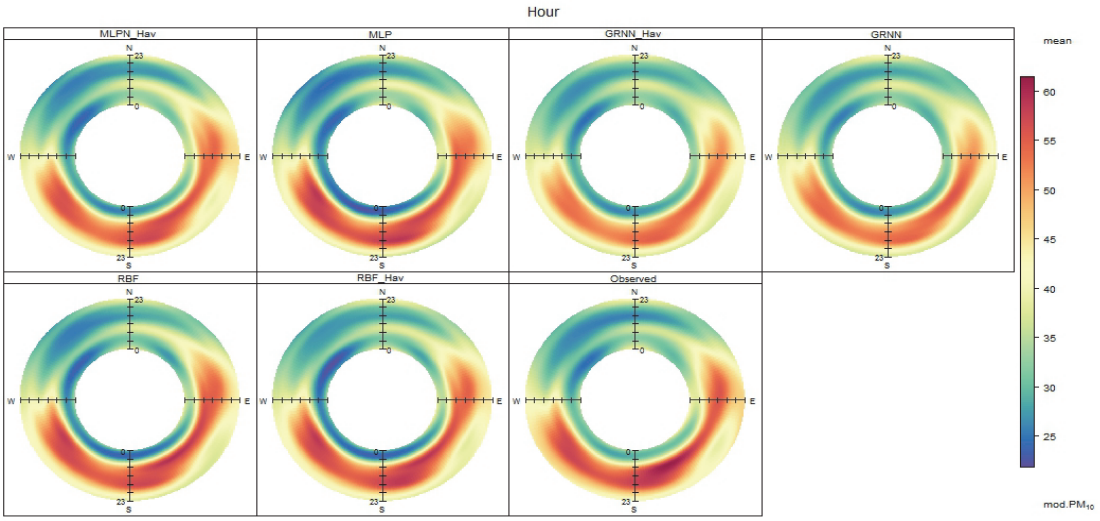

Figure 6: Polar annulus showing the models performance

Figure 6 shows the Polar annulus Plots comparing the agreement hourly variations between the model prediction and the observed data. The graduated marks on North, South, East and West indicates hours of the day. The plots also show the variation of the particle concentrations with wind direction and wind speed. The colour scale shows the PM_{10} concentration levels from low (blue) to high (red) by wind directions.

The reference (rooftop) wind directions and wind speeds plays a vital role in dispersing pollutants within and outside the street canyon (Marylebone road), therefore, polar plots shown in Figure 7 are important tools that describes the dispersion of the pollutants by wind sectors and at various wind speeds. The 7th panel in Figure 7 shows that the observed PM_{10} concentrations were higher along the axis of the road (75^0 - 225^0) and more associated with the East, Southeast, South and Southwest winds which were the dominant wind directions and speeds at the site. The high concentration associated with the East occurred mostly at higher wind speeds while those of Southwest occurred at both low and high wind speeds. The models performed extremely well in capturing the wind characteristics at the sites with various degrees of prediction accuracies. For example the RBF and MLP models have done well in identifying the regions on the plots with higher concentrations with little underestimation while the GRNN models generally underestimates the higher concentrations associated with lower to average wind speeds.

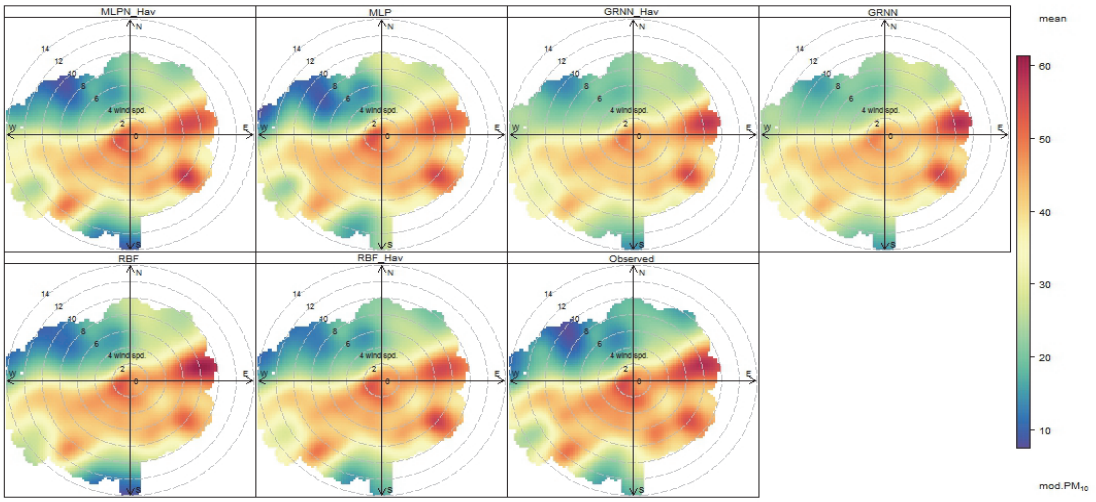

Figure 7: Polar Plots comparing modelled and observed PM_{10}

In Figure 7,the surface of the plot shows PM_{10} concentrations by wind directions and the colour code express the amount of PM_{10} concentrations from lower to higher concentrations and the dashed circles indicates wind speed intervals

The performance of the models in accounting for the daily variations in the PM_{10} concentration is shown in Figure 8. The two GRNN and MLPN models have shown consistent slight underestimations throughout the week, while the MLPN_Hav and the two RBFN models have shown better agreement with the observed daily PM_{10} concentrations. Also the models captured the general pattern of the daily concentrations.

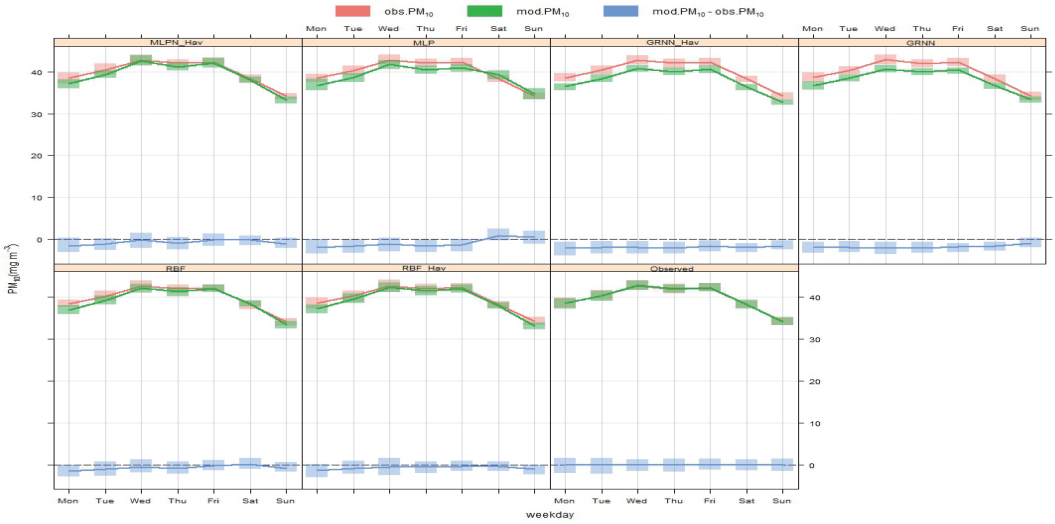

Figure 8: Time variation Plot comparing daily variation of the modelled and observed PM_{10}

The RBFN-Hav model was found to be the most accurate in most of the analysis carried out above, and hence it was further investigated to examine its hourly, daily, weekly and monthly variations as shown in Figure 9. Its hourly predictions were predominantly accurate with slight overestimation during the daylight hours and also slight underestimation during dark hours. The daily predictions were also accurate with an average underestimation of approximately 1 $\mu g/m^3$. The predictions were mostly accurate during spring and summer and mostly under predicted autumn and winter concentrations.

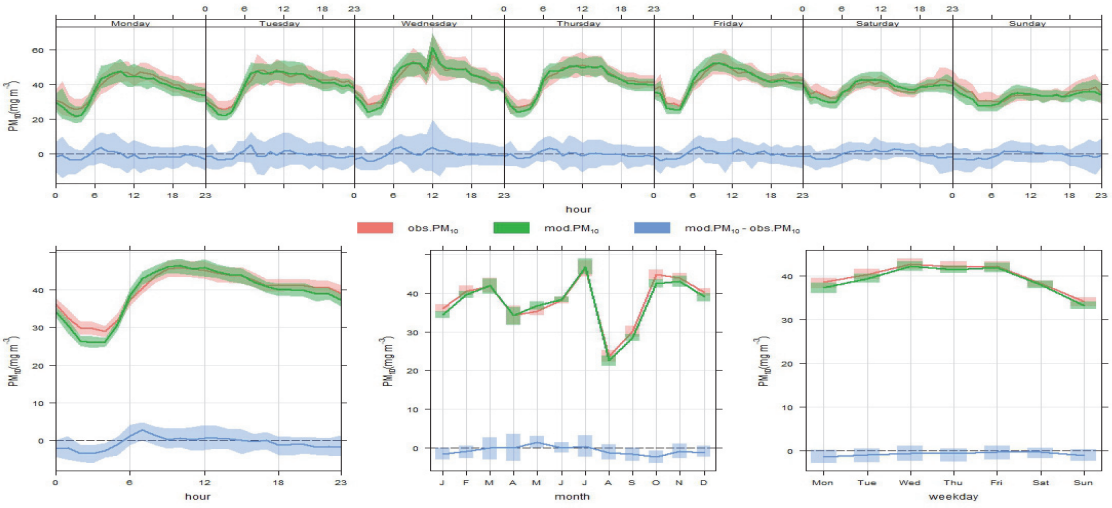

Figure 9: Time variation Plot comparing daily variation of the RBFN_Hav modelled and observed PM_{10}

4.0 Conclusions

Three variants of Neural networks namely GRNN, MLPN, and RBFN were used to develop ANN based PM_{10} prediction models in this study. The performance of the models were compared and evaluated extensively using "*openair package*" a freely available air quality data analysis tool based on *R statistical software*. The package offered great flexibility for analysing the performance of the models not only with statistical performance metrics but with graphically coded functions where the models performance can be visually examined under different circumstances.

Generally, the models performed excellently in capturing the overall nature of the PM_{10} observation at the sites. Their predictions mostly conformed to the seasonal, monthly, weekly, daily and hourly variations of the PM_{10} observations, however, the models differ in their prediction accuracies. The RBF

networks were found to be the best performing models, and the model with 24 hour rolling mean of the PM_{10} observations as part of their predictors were also better than those without the 24 hour rolling mean. The models slightly under predicted the PM_{10} observations between 6:00PM and 6:00AM but much more accurate prediction during daylight hours. The models were also better in predicting the PM_{10} concentrations during spring and summer and also slightly under predicted the PM_{10} concentrations during autumn and winter seasons. They also captured the relationships between the observation and the reference winds at the site. The RBF networks were the best in most the analysis carried out in this study. However, our experience has shown that MLP networks were easier and quicker to train when the training data is big and also they require less computer memory. But if the training data is not big RBFN is quicker and also much more accurate. The GRNN models were found to be the least performing models, but slightly better than MLPN when considering the predictors without 24 hour rolling mean of the PM_{10} concentrations. Therefore it can be concluded that the ANN based models can be used to predict PM_{10} concentrations in street canyons with greater degree of accuracy and in-depth analysis using openair package tools boosted our confidence in recommending ANN based models for street canyon air quality studies.

References

AMIRSASHA BNANANKHAH, F. N. 2012. Artificial Neural Networks: A Non-Linear Tool for Air Quality Modeling and Monitoring. *International Conference on Applied Life Sciences,* ICALS2012.

ASTM 2010. Standard guide for Statistical Evaluation of Atmospheric Dispersion Model Performance (D6589). *ASTM International.* West Conshohocken,PA: ASTM.

AURELIE, C. & HARRISON, R. M. 2005. Comparison between SMPS, nano-SMPS and Epiphaniometer data at an urban Background site (bloomsbury) and a Roadside site (marylebone road).

BENNETT, N. D., CROKE, B. F. W., GUARISO, G., GUILLAUME, J. H. A., HAMILTON, S. H., JAKEMAN, A. J., MARSILI-LIBELLI, S., NEWHAM, L. T. H., NORTON, J. P., PERRIN, C., PIERCE, S. A., ROBSON, B., SEPPELT, R., VOINOV, A. A., FATH, B. D. & ANDREASSIAN, V. 2013. Characterising performance of environmental models. *Environmental Modelling & Software,* 40**,** 1-20.

BISHOP, C. M. 1995. Neural networks for pattern recognition.

BOŽNAR, P. M. A. M. Z. 2011. Artificial Neural Networks - a Useful Tool in Air Pollution and Meteorological Modelling, Advanced Air Pollution, Dr. Farhad Nejadkoorki (Ed.). *InTech*.

BRAUER, M., HOEK, G., VAN VLIET, P., MELIEFSTE, K., FISCHER, P. H., WIJGA, A., KOOPMAN, L. P., NEIJENS, H. J., GERRITSEN, J., KERKHOF, M., HEINRICH, J., BELLANDER, T. & BRUNEKREEF, B. 2002. Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. *Am J Respir Crit Care Med,* 166**,** 1092-8.

BRUNEKREEF, B., BEELEN, R., HOEK, G., SCHOUTEN, L., BAUSCH-GOLDBOHM, S., FISCHER, P., ARMSTRONG, B., HUGHES, E., JERRETT, M. & VAN DEN BRANDT, P. 2009. Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study. *Res Rep Health Eff Inst,* 139**,** 5-71.

CARSLAW, D. C. & BEEVERS, S. D. 2013. Characterising and understanding emission sources using bivariate polar plots and k-means clustering. *Environmental Modelling & Software,* 40**,** 325-329.

CARSLAW, D. C., BEEVERS, S. D., ROPKINS, K. & BELL, M. C. 2006. Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. *Atmospheric Environment,* 40**,** 5424-5434.

CARSLAW, D. C. & ROPKINS, K. 2012. openair — An R package for air quality data analysis. *Environmental Modelling & Software,* 27-28**,** 52-61.

CHANG, J. C. & HANNA, S. R. 2004. Air quality model performance evaluation. *Meteorology and Atmospheric Physics,* 87**,** 167-196.

DEVELOPMENT CORETEAM, R. 2013. Development Core. R: a language and environment for statistical computing. Vienna, Austria.

ESPLIN, G. J. 1995. Approximate explicit solution to the general line source problem. *Atmospheric Environment,* 29**,** 1459-1463.

FEI, L., PING, M. & MING, Y. A validation methodology for AI simulation models. Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, 18-21 Aug. 2005 2005. 4083-4088 Vol. 7.

GARDNER, M. W. & DORLING, S. R. 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment,* 34**,** 21-34.

HAYKIN, S. 2005. Neural Networks - A Comprehensive Foundation Second Edition. *Pearson (Education), Pearson Printice Hall Publication*.

HEINRICH, J., TOPP, R., GEHRING, U. & THEFELD, W. 2005. Traffic at residential address, respiratory health, and atopy in adults: the National German Health Survey 1998. *Environmental Research,* 98**,** 240-249.

JONES, A. & HARRISON, R. 2006. Estimation of the emission factors of particle number and mass fractions from traffic at a site where mean vehicle speeds vary over short distances. *Atmospheric Environment,* 40**,** 7125-7137.

KENG, B. A. 2010. Neural Network Modelling of Present and Future Urban PM10 Concentrations based on Measurement Results. *Doctor Engineer (Dr.-Ing.) Approved paper; Institute of Combustion and Power Plant Technology, University of Stuttgart, Department of Air Pollution*.

KIM, J. J., SMORODINSKY, S., LIPSETT, M., SINGER, B. C., HODGSON, A. T. & OSTRO, B. 2004. Traffic-related air pollution near busy roads: the East Bay Children's Respiratory Health Study. *Am J Respir Crit Care Med,* 170**,** 520-6.

KUKKONEN, J., PARTANEN, L., KARPPINEN, A., RUUSKANEN, J., JUNNINEN, H., KOLEHMAINEN, M., NISKA, H., DORLING, S., CHATTERTON, T., FOXALL, R. & CAWLEY, G. 2003. Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment,* 37**,** 4539-4550.

KUMAR, A. & GOYAL, P. 2013. Forecasting of Air Quality Index in Delhi Using Neural Network Based on Principal Component Analysis. *Pure and Applied Geophysics,* 170**,** 711-722.

LEGATES, D. R. & MCCABE, G. J. 2012. A refined index of model performance: a rejoinder. *International Journal of Climatology*.

LINDGREN, A., STROH, E., NIHLEN, U., MONTNEMERY, P., AXMON, A. & JAKOBSSON, K. 2009. Traffic exposure associated with allergic asthma and allergic rhinitis in adults. A cross-sectional study in southern Sweden. *International Journal of Health Geographics,* 8.

MCCONNELL, R., BERHANE, K., YAO, L., JERRETT, M., LURMANN, F., GILLILAND, F., KUNZLI, N., GAUDERMAN, J., AVOL, E., THOMAS, D. & PETERS, J. 2006. Traffic, susceptibility, and childhood asthma. *Environ Health Perspect,* 114**,** 766-72.

N SHARMA, K. K. C. C. V. C. R. 2005. vehecular pollution modelling using artificial neural network technique: A review. *journal of scientific and industrial reasearch,* 64**,** 637-647.

NAGENDRA, S. M. S. & KHARE, M. 2006. Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecological Modelling,* 190**,** 99-115.

PETERSEN, R. L. 1997. A wind tunnel evaluation of methods for estimating surface roughness length at industrial facilities. *Atmospheric Environment,* 31**,** 45-57.

RUSSO, A., RAISCHEL, F. & LIND, P. G. 2013. Air quality prediction using optimal neural networks with stochastic variables. *Atmospheric Environment,* 79**,** 822-830.

SINGH, K. P., GUPTA, S., KUMAR, A. & SHUKLA, S. P. 2012. Linear and nonlinear modeling approaches for urban air quality prediction. *Science of The Total Environment,* 426**,** 244-255.

TAYLOR, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres (1984–2012),* 106**,** 7183-7192.

THUNIS, P., GEORGIEVA, E. & PEDERZOLI, A. 2012. A tool to evaluate air quality model performances in regulatory applications. *Environmental Modelling & Software,* 38**,** 220-230.

WILKS, D. S. 2011. Chapter 7 - Statistical Forecasting. *In:* DANIEL, S. W. (ed.) *International Geophysics.* Academic Press.

WILLMOTT, C. J. 1981. On the validation of models *Physical geography,* 2**,** 184 - 194.

WILLMOTT, C. J. 1982. Some comments on the evaluation of models. *Bulletin American Meteorological Society*.

WILLMOTT, C. J., ACKLESON, S. G., DAVIS, R. E., FEDDEMA, J. J., KLINK, K. M., LEGATES, D. R., O'DONNELL, J. & ROWE, C. M. 1985. Statistics for the evaluation and comparison of models. *Journal of geophysical research,* 90**,** 8995-9005.

WILLMOTT, C. J. & WICKS, D. E. 1980. An empirical method for the spatial interpolation of monthly precipitation within California. *Physical Geography,* 1**,** 59-73.

YAN CHAN, K. & JIAN, L. 2013. Identification of significant factors for air pollution levels using a neural network based knowledge discovery system. *Neurocomputing,* 99**,** 564-569.

YAP, W. K. & KARRI, V. 2013. Comparative analysis of artificial neural networks and dynamic models as virtual sensors. *Applied Soft Computing,* 13**,** 181-188.

ZHANG, H., LIU, Y., SHI, R. & YAO, Q. 2013. Evaluation of PM10 forecasting based on the artificial neural network model and intake fraction in an urban area: A case study in Taiyuan City, China. *Journal of the Air & Waste Management Association,* 63**,** 755-763.