

Analyzing the Dynamics of Particulate Matters (PM) using Nonlinear Dynamical Techniques and Predicting the Behavior Based on Robust Regression Models

Sharjil Saeed¹ Adnan Idris³, Lal Hussain^{1, 2*}, Imtiaz³ Ahmed Awan¹

¹Department of Computer Science & IT, The University of Azad Jammu and Kashmir, City Campus, 13100, Muzaffarabad,

^{2*}Correspondence Author, Quality Enhancement Cell, The University of Azad Jammu and Kashmir, City Campus, 13100, Muzaffarabad, Pakistan

Pakistan

³Department of Computer Science & IT, University of Poonch Rawalakot, Rawalakot, Azad Kashmir, Pakistan

Summary

Particulate matter (PM) concentrations are outcome of combination of complex natural and anthropogenic contributors. Considerable interest among environmental research community has been found to study the dynamics of particulate matter concentration using nonlinear time series analysis techniques. Due to the reconstruction work after the earthquake 8, 2005 and congestion of vehicles because of traffic jam, the indoor environment of houses located along the road sides is badly affected in the suburb of Muzaffarabad. For quantifying the nonlinear dynamics of the particulates (PM_{2.5} and PM₁₀), time series data was acquired using Haz-Dust EPAM-5000 monitor from the indoor air of the closed rooms located near the main roadsides of Muzaffarabad city. The data was then transferred to computer for analyzing nonlinear behavior using phase space reconstruction and largest Lyapunov exponent (LLE). The average mutual information (AMI) function was used to estimate the time delay. False nearest neighbors (FNN) approach was used to obtain optimum embedding dimension for phase space reconstruction and Wolf's algorithm was used to calculate LLE. Moreover, prediction in terms of root means squared error (RMSE), mean square error (MSE), mean absolute error (MAE), observed and estimated functional values and feasible points in terms of regression tree (RT), support vector regression (SVR) and gaussian process regression (GPR) models was made by extracting multifeatured extracted from indoor particulate matters. Based on nonlinearity, non-stationarity and complex temporal dynamics, the time domain, frequency domain, complexity, wavelet based and statistical features are extracted from indoor particulate matters PM_{10.0} and PM_{2.5}. The results clearly indicated that indoor particulate time series manifested chaotic behavior. Furthermore, to test whether indoor particulates time series data is measure of temporal dynamics or not, sensitivity of Poincaré plot descriptors, Sv (representing short term variability), Lv (representing long term variability) and complex correlation measure (CCM) was quantified. The findings demonstrated that relative changes in CCM are more sensitive to the changes in temporal dynamics as compared to standard descriptors Sv and Lv for both PM_{2.5} and PM₁₀. The prediction results reveal that Support vector linear regression and Gaussian Process regression models gives minimal differences

between observed and predicted values and optimal performance evaluation in terms of RMSE, R², MSE and MAE

Keywords:

Complex correlation measure, Lyapunov exponent, Particulate matter, Phase space reconstruction, Poincare plot, Gaussian process regression, support vector regression, root mean squared error

1. Introduction

Particulate Matters (PM) are atmospheric pollutants effecting health of humans across the globe irrespective of geographical locations [1-2]. The PM particles range in size from few nano-meters to tens of micrometer (μm) in diameter (PM_{1.0}, PM_{2.5} and PM₁₀). The hazardous effects of these particulates depend on the size, distribution and composition of particulates [3-5]. The ultra and fine particles (PM_{1.0} and PM_{2.5}) can have more impacts on human health as compared to coarse particles (PM₁₀) [6-8]. It was estimated by the World Bank (1993) that indoor air pollution is responsible for almost 50% of the total diseases resulting from poor household environments in developing countries [9-10]. In a properly maintained home, most of the airborne particulate comes from the outside [11]. However, some homes do have significant sources of indoor PM which come from cigarette smoking, cooking, malfunctioning combustion appliances, wood burning stoves [12-13]. Among all indoor air pollution, indoor PM is recognized as the one of the most important pollutants in terms of adverse effects on human health [14]. PM has a severe effect on public health via the respiratory system [15] such as asthma, bronchitis, obstructive disease, cancer [15] fever, cough, bronchial constriction [16]. The indoor pollutants are considered more hazardous for humans compared to their outdoor counterparts, because on average, people spend over 90% of their life time indoor [17].

The houses which are located along the road side are more prone to chemical and biological containments and humans residing in these houses might have hazardous effects of these pollutants on their health. In this context, the characterization of underlying dynamics of PM concentration has been an important endeavor for researchers and Government environmental control agencies. The nonlinearity is inherited in atmospheric and other related real systems [1-2] and it would be useful to study temporal dynamics of these systems using nonlinear time series analysis techniques. In order to quantify the nonlinear dynamics of PM_{2.5} and PM₁₀, time series data was collected from the houses located near roadside in Muzaffarabad city at six different sites. The chaotic behavior of indoor particulates (PM_{1.0} and PM_{2.5}) was investigated using phase space reconstruction [18] and largest Lyapunov exponent [19] Poincaré plot has found its use in diverse fields such as astronomy, geophysics, mathematical biology and medical science. In the context of biomedical science, it is mainly used for characterizing cardiac autonomic dysfunction by performing heart rate variability analysis [38,40-41]. To test whether indoor particulates time series data is measure of temporal dynamics or not, sensitivity of Poincaré plot descriptors Sv, Lv and CCM was quantified.

To extract the most relevant features is highly desired in order to predict and detect the information using regression and classification techniques. The features are extracted based on nature of data, its nonlinear dynamics, complexity, no-stationarity, morphology, spectral and temporal variations etc. Recently [42-43] extracted geometric features to detect the colon and predict the colon cancer. [44] extracted the texture features to detect the breast cancer, Hussain et al. extracted morphological and texture features to detect human faces [45], acoustic and mel-frequency cepstral coefficients features [46] to detection emotions in human speech signal and combination of texture, morphological, entropy, SIFT, EFDs features [47] to detect and predict the prostate cancer, time-domain, frequency domain and entropy features [48-49] to detect and predict heart rate variability and alcoholism using machine learning classification methods. The temporal and spectral variation have extensively been studied in varying applications such as coronary artery disease [50] to diagnose mortality risk in patients, cognitive heart failure [52-53], dilated cardiomyopathy [51]. Moreover, automated methods such as time domain [54] spectral features [55] fast Fourier transforms, entropy features and DWT [56-58] have been employed to detect the predict various pathologies including epileptic seizure detection.

In this study to predict the indoor particulate matters PM_{10.0} and PM_{2.5} multifeatured strategies was employed including time domain features such as standard deviation of whole time series (SDNN), standard deviation of

average interval of time series (SDANN), square mean squared of standard deviation (RMSSD), standard deviation of standard deviation (SDSD); frequency domain features such as total power (TP), ultra-low frequency (ULF), very low frequency (VLF), low frequency (LF), high frequency (HF); complexity and wavelet based features such as multiscale sample entropy [60] and fast sample entropy based on KD tree algorithm [59], wavelet entropy including Shannon, threshold, sure, log Energy and norm [58], and statistical features including smoothness, variance, kurtosis, and skewness.

2. Material and Methods

Study Area

Muzaffarabad is the capital city of Pakistani administered Kashmir commonly known as Azad Jammu & Kashmir (AJ&K). It is a beautiful cup shape valley; the land of velvet green plateaus, charming lakes and waterfalls. The geographical location of the city contains 73.47°E (Longitude) and 34.37°N (Latitude) at the conjunction of Neelum and Jhelum rivers. City is surrounded by arrogant mountains and weather conditions are almost same as Islamabad, Pakistan.

The city was the site of epicenter of the 2005 Kashmir earthquake, had a magnitude of 7.6 at Richter scale. The disaster destroyed fifty percent of the buildings in the city; due to reconstruction of the city affect the weather conditions which results in increase in the particulates especially indoor. Muzaffarabad district is adjacent to the Khyber Pakhtunkhwa province of Pakistan in the west, Indian held Kashmir in the east and Chinese area in the North.

Data Acquisition

Particulate Matter (PM) concentration was measured at six different sites of premises located near the roadside of Muzaffarabad (AJK), the capital city of Pakistan administered Kashmir. Data was collected by the Haz-Dust, environmental particulate air monitor (EPAM-5000), which is unique equipment in this area to measure the concentration. The measurement device used was a high sensitivity real-time particulate monitoring device that uses an infrared light source positioned at a 90° angle from a photodetector. The airborne particles scatter the infrared beam light so that the amount of light received by the photodetector is directly proportional to the aerosol concentration. It was calibrated using standard Arizona Road Dust (ARD) with NIOSH method 0600 (NIOSH, 1998) with a precision of ± 0.03 mg m⁻³ and an accuracy of $\pm 10\%$ for irrespirable dust.

At each site, EPAM-5000 was installed for consecutive six hours measurement in closed location to monitor the

concentration of indoor particulate matter concentrations. Sampling rate was one second per measurement that generated 21600 data points. After acquiring the data, device was returned to laboratory and data transferred to a personal computer with the help of manufacturer-provided software.

3. Non-Linear Techniques

Phase Space Reconstruction:

Phase space represents set of all possible states of a dynamical system on which modeling of that dynamical system relies [18-21]. The phase space and mathematical description of the experimentally and naturally occurring dynamics systems are often unknown [18,21]. The method of time delay embedding firmly rooted in the mathematical topology theory to Taken's theorem is one of the most frequently used methods of phase space reconstruction [18,20]. The method time delay embedding uses $x(t)$ to 'reconstruct' the attractor with a 'scatter' plot of the following lagged coordinate vectors: $x(t)$, $x(t-\tau)$, $x(t-2\tau)$, ..., $x(t-(m-1)\tau)$, where τ is the time delay and m is the embedding dimension.

The proper selection of reconstruction parameters τ and m is directly related with the accuracy of the invariables of the described characteristics of the reconstructed attractor.

Selection of Time Delay:

Methods for selecting a suitable time delay have centered on measures of autocorrelation and mutual information [23]. In this study, we have used mutual information function, which accounts for the nonlinear correlation in the time series to evaluate ' τ ' for the time series. The average mutual information for a given time series with consecutive values of x_i and $x_{i+\tau}$ is given by the relation

$$I(\tau) = (x + a)^n = \sum_{i,i+\tau} \rho(X_i, X_{i+\tau}) + \tau) \ln \frac{\rho(X_i, X_{i+\tau})}{\rho(X_i) \rho(X_{i+\tau})} \quad (1)$$

For non-monotonous decrease of $I(\tau)$, the location of first local minimum is considered the suitable value of τ [23]. For monotonous decrease of $I(\tau)$, either the decrease of average mutual information to $I(t)/I(0)=0.2$ or $I(t)/I(0)=1/e$ can be used as time delay [21]. Since $I(\tau)$, decreases non-monotonously with increasing τ , so we have used location of first change for calculating time delay (τ) in this study.

Embedding Dimension

During last three decades, various approaches based on Takens' embedding theorem have been proposed for computing the optimal embedding dimension for the study of chaos in time series [20-21,25]. In this study, we have

used false nearest neighbor (FNN) approach to examine whether points that are near neighbors in one dimension are also near neighbors in the next higher embedding dimension. If so, then image has been completely unfolded and dimension has been established to characterize the dynamics of the chaotic system [25].

To identify false nearest neighbor, following two criteria were used.

$$\delta_n > R_{tol} \quad (2)$$

$$\frac{D_n^{(m+1)}}{R_A} > R_{tol} \quad (3)$$

Where δ_n is the relative increase in Euclidean distance when dimension of reconstructed state space is increased from m to $m+1$, and it is computed as given in equation (4);

$$\delta_n = \sqrt{\frac{D_n^2(m+1, \tau) - D_n^2(m, \tau)}{D_n^2(m, \tau)}} = \frac{|x_{n+m\tau} - x_{n+\tau m}|}{D_n(m, \tau)} \quad (4)$$

The parameters R_{tol} and A_{tol} in equations (2) and (3) are constant thresholds and R_A is the standard deviation of a time series.

Lyapunov Exponent

The system's sensitivity on initial condition demonstrates the Lyapunov exponents [19]. Evidence of chaos is considered having positive Lyapunov exponents, while negative exponents suggest a mean returning behavior. Hence, positive Lyapunov exponents are directly proportional to degree of chaos in a system and inversely proportional to time scale of the system's predictability [27]. Entropy ' S ' is achieved by summing up of positive Lyapunov exponents (base e), its reciprocal is roughly the time over which meaningful prediction is possible [27]. Lyapunov exponents can be calculated from equation (5), as follows;

$$\lambda = \lim_{t \rightarrow \infty} \lim_{\Delta z_0 \rightarrow 0} \frac{1}{t} \ln \frac{|\Delta z(t)|}{|\Delta z_0|} \quad (5)$$

Sensitivity to Changes in Temporal Structure using Poincare Plot Descriptors

Poincaré plot is a computer technique used to analyze short and long-term recordings providing visual as well quantitative information of a data sequence. As a visual technique, Poincaré plot provides geometrical presentation of time series data in phase space revealing nonlinear patterns of the data sequence and a quantitative measure; it introduces various descriptors which quantify the information contained in a Poincaré plot [38,40-41]. Tulppo *et al*, fitted an ellipse to the shape of Poincaré plot and proposed two descriptors ' S_v ' and ' L_v ' to quantify the plot geometry. The descriptor ' S_v ' represents short terms variability measuring dispersion of points

perpendicular to the line of identity and 'L_v' represents long term variability measuring the dispersion of points along the line of identity [38-40]. The standard descriptors 'S_v' and 'L_v' evaluate gross representation of a data sequence and to evaluate point to point variation of the data sequence. Chandan *et al.* [41] proposed complex correlation measure (CCM). The CCM is a function of multiple correlation of a data sequence and is computed in windowed manner which embeds temporal information of the data.

Poincare plot has found its use in diverse fields such as astronomy, geophysics and biomedical science. In the context of biomedical science, it is mainly used for characterizing cardiac autonomic dysfunction, prognosis of mortality in post myocardial infarction, sudden cardiac death, and sudden infant mortality and to assess risk of ventricular arrhythmias using inter-beat interval time series [38,40-41].

Sensitivity is defined as the rate of change of the value due to the change in temporal structure of the signal. The change in temporal structure of the time series in a window is achieved by surrogating the signal (i.e., time series data points) in that window. The sensitivity of descriptors ΔS_v, ΔL_v and ΔCCM were calculated using following equations (6 to 8).

$$\Delta S_v = \frac{S_{vj} - S_{v0}}{S_{v0}} \quad (6)$$

$$\Delta L_v = \frac{L_{vj} - L_{v0}}{L_{v0}} \quad (7)$$

$$\Delta CCM = \frac{CCM_j - CCM_0}{CCM_0} \quad (8)$$

Where S_{v0}, L_{v0} and CCM₀ were the Poincare plot descriptors for the original data set without surrogating and 'j' represents the window number whose data was surrogated. Moreover, S_{vj}, L_{vj} and CCM_j represents the values of these descriptors after surrogating of jth window. The entire time series data was in to 20 windows, resulting in 20 values of ΔS_v, ΔL_v and ΔCCM.

Support Vector Regression (SVR)

SVR is most widely used prediction methods in many applications including wind speed and many others [61-70]. Chen *et al.* [71] used this method as an extension of SVM. The empirical error and model complexity can be best described using this method [72]. In the hyperplane, the SVM regression function f(.) is constructed in such a way as to maximize the margin. The reduced training data sets also called the support vectors are actively involved in optimizing the problem.

Consider a dataset in form of $(X_i, Y_i) \in R^N \times R$, the SVR dual optimization problem can be formulated as

$$\begin{aligned} & \text{Maximize } S(\gamma, \gamma^*) \\ & = \frac{-1}{2} \sum_{i,j=1}^m (\gamma_i, \gamma_i^*) (\gamma_j, \gamma_j^*) \langle \varphi(X_i), \varphi(X_j) \rangle \\ & - \varepsilon \sum_{i=1}^m (\gamma_i^*, \gamma_i) \gamma_i \\ & \text{Subject to } \sum_{i=1}^m (\gamma_i, \gamma_i^*) = 0 \\ & 0 \leq \gamma_i \leq C, \forall i = 1, \dots, m \\ & 0 \leq \gamma_i^* \leq C, \forall i = 1, \dots, m \end{aligned} \quad (9)$$

Where C denote the complexity penalization in term of γ, γ^* corresponds to the dual variables for the active constraints. For nonlinear problem, SVR use the mechanism of kernel trick by applying the kernel functions such as linear, Gaussian, Polynomial, Radial base function etc.

SVM Polynomial Kernel

$$K(x_i, y_i) = (x_i \cdot y_i + 1)^n \quad (10)$$

SVM Gaussian (RBF) kernel

$$K(x_i, y_i) = \exp\left(\frac{-1}{2} \frac{\|x_i - y_i\|^2}{\sigma^2}\right) \quad (11)$$

SVM Fine Gaussian (RBF) kernel

$$K(x_i, y_i) = \exp\left(\frac{-1}{2} \frac{\|x_i - y_i\| \|x_i - y_i\|}{\sigma^2}\right) \quad (12)$$

Where n is the order of polynomial kernel and σ is the width of RBF. The dual formulation for non-linear case is given by

$$\alpha^* = \max_{\alpha} \left(\sum_i \alpha_i + \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \right) \quad (13)$$

Subject to

$$0 \leq \alpha_i \leq C \text{ and } \sum_i \alpha_i y_i = 0$$

Decision Tree (DT)

DT is used in problems relevant to estimating the target classification patterns and or target variables. It is constructed using top down recursive partitioning approach. The internal space of the tree is divided into set of separate areas by assigning response value to each area. For regression problems, the response is determined on the mean target value which are associated with patterns located each area. The single tree has low generalizability and is vulnerable to overfitting. Thus, small change in learning pattern results in change in fundamental of tree structure. This problem is resolved by employing random forest (RF) which has the ability to predict on averaging the results obtained from all the relevant decision trees [73-74]. In CART, the Regression trees have a constant numerical value in the leaves and use the variance as a

measure of impurity. The split selection measure for a node T is:

$$Err(T) \stackrel{\text{def}}{=} E[(Y - E[Y|T])^2] \\ \Delta Err(T) = Err(T) - \sum_{j=1}^n P[q_j(X)|T] Err(T_j) \quad (14)$$

The impurity measure for the use of variance is justified by the fact that the best constant predictor in a node is the expected value of the predictor variable providing that the data points belong to the node, $E[Y|T]$; the variance in this case is thus measured as the square error of the best predictor.

Boosted regression trees (BRT)

To predict the quantitative (regression tree) or classified (classification tree), the decision trees are created, the classification and regression tree (CART) methods under BTR. In order to achieve the adequate predictive performance, the BTR predictor models through optimized and boosted binary separation combined the larger number of simpler models [75]. The lesser sensitivity to overfitting [76] and higher speed analysis even to high volume of data [77] is achieved using these methods. This performance is achieved by adjusting the boosted tree and shrinking rate options.

Gaussian Process Regression (GPR)

For continuous domain functions, the GPR are most suitable and flexible models to be used [78-80] covariance function $K(x^{(i)}, x^{(j)})$ and mean function $m(x)$ are used to define this model, thus the joint probability distribution of the function value $f = [f(x)^{(1)}, f(x)^{(2)}, \dots, f(x)^{(N)}]^T$ at any finite set of input points $X = [x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(N)}]^T \in R^{N \times D}$

Is a multivariate Gaussian distribution $P(f) = N(m, K(X, X'))$, where $m = [m(x)^{(1)}, m(x)^{(2)}, m(x)^{(3)}, \dots, m(x)^{(N)}]^T$ and notation $K(X, X')$ represents co-variance matrix with entries $K_{i,j} = K(x_i, x_i')$.

Thus, GPR can be considered as an infinite dimensional generalization of multivariate Gaussian distribution with prior probability distribution for unknown function f.

Performance Evaluation

The regression predictor performance was measured in term of root mean squared error (RMSE), co-efficient of determination (R2), mean square error (MSE) and mean absolute error (MAE).

Root mean squared error (RMSE)

The performance of the predictors was estimated quantitatively by measuring the accuracy.

The RMSE was computed using following method:

$$RMSE = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (16)$$

Where x_i and y_i denote the measured and predicted values of the i-th sample, and 'n' denote the total number of samples of the training dataset. The smallest value of RMSE denote the best descriptor.

Co-efficient of determination (R2)

R² can be computed using the following function:

$$R^2 = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{y})^2} \quad (17)$$

Here $\bar{y}$ denote the average values of all the samples.

Mean Square Error (MSE)

MSE is computed using following method.

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (18)$$

The MSE of an estimator measures the average of the squares of errors or deviations. MSE also denote the second moment of error that incorporate both variance and bias of an estimator.

Mean Absolute Error (MAE)

MAE is the measure of difference between two consecutive variables, for example variable y and x denote the predicted and observed values, then MAE can be calculated as:

$$MAE = \sum_{i=1}^n \frac{|y_i - x_i|}{n} \quad (19)$$

4. Results and Discussions

Results obtained from time series of PM_{2.5} and PM₁₀ are listed in Table 1. Each of values listed for PM_{2.5} and PM₁₀, at specific location, are average of 21600 data points. For PM_{2.5} minimum mean value (75.65 µg/m³) was recorded at Chehla Bridge, while the maximum mean value (357.04 µg/m³) at CMH Chowk. Measured values were then compared with standard deviation and 95% confidence interval of concentration of the particulates (PM_{2.5} and PM₁₀) at selected sites in the downtown city of Muzaffarabad. Same procedure was followed for PM₁₀ concentration. For PM₁₀ minimum mean value (57.00

$\mu\text{g}/\text{m}^3$) was recorded at Chattar Chowk, while maximum (409.05 $\mu\text{g}/\text{m}^3$) at CMH Chowk.

Table 1: Statistical Measures for all indoor sites, N=21600 * 6 sites

Types of PM	Location	Mean Value of PM($\mu\text{g}/\text{m}^3$)	Standard Deviation	95% CI
PM _{2.5}	Bank Road	190.84	90.24175	189.6356 - 192.0426
	Bus Stand	80.87	92.99781	79.62579 - 82.10634
	Chattar Chowk	91.08	97.28224	89.77925 - 92.37408
	Chehla Bridge	75.65	94.48731	74.3862 - 76.90648
	CMH Chowk	357.04	218.6989	354.121 - 359.9545
	Old Secretariat	82.93	83.63161	81.81737 - 84.04809
PM ₁₀	Bank Road	243.6	154.74	241.5359 - 245.6632
	Bus Stand	82.87	90.170	81.66525 - 84.07031
	Chattar Chowk	57.00	85.420	55.85636 - 58.13475
	Chehla Bridge	69.09	93.510	67.84566 - 70.33990
	CMH Chowk	409.05	258.00	405.6083 - 412.4900
	Old Secretariat	72.49	91.240	71.27476 - 73.70838

Phase Space Reconstruction:

The selection of time delay ' τ ' and optimum minimum embedding dimension ' m ' is vitally important for phase space reconstruction. The average mutual information, AMI, a generalization of auto-correlation function is used for determination of time delay ' τ '. In figure 1, the AMI is plotted against varying ' τ ', for optimal determination of its value for PM_{2.5} and PM₁₀. It is evident from the plot in figure 1, if AMI decreased monotonically with increasing ' τ ', so $I(t)/I(0)=1/e$ was used as the criteria for determination of optimal time delay.²⁸ The optimal time delay, ' τ ', for particulates PM_{2.5} and PM₁₀ using this criterion are 18 and 19 respectively.

The false nearest neighbors (FNN) approach is used to find the minimum embedding dimension ' m '. For any given embedding dimension ' m ', the proportion of the identified FNN to all the neighbors was computed for the given time delay ' τ '. The percentages of the FNN are plotted as function of the embedding dimension. A zero FNN percentage indicates the optimum, i.e. minimum,

embedding dimension. In Figure 2, the results of the FNN approach for determining the optimum embedding dimension of N = 21,600 data points of PM₁₀ and PM_{2.5} using various values for the threshold parameters R_{tol} and A_{tol} are shown. In this study, the value of the parameter R_{tol} was varied from 20 to 200 and $A_{tol} = 0.2 \times R_{tol}$ was used. The results summarized in Figure 1 and 2 showed that the optimum embedding dimension obtained by the FNN approach varies from 4 and 6 depending on the values of thresholds R_{tol} and A_{tol} . The higher values of optimal embedding showed that the concentration time series of PM_{2.5} and PM₁₀ have dominant degrees of freedom, indicating dominance of some complexity in their concentrations respectively. The complexity in PM concentrations may be due to the various influencing factors such as temporal variations, humidity, indoor temperature, number of person in the room, wind directions and speed, tobacco smoking, smoke produced from automobiles.

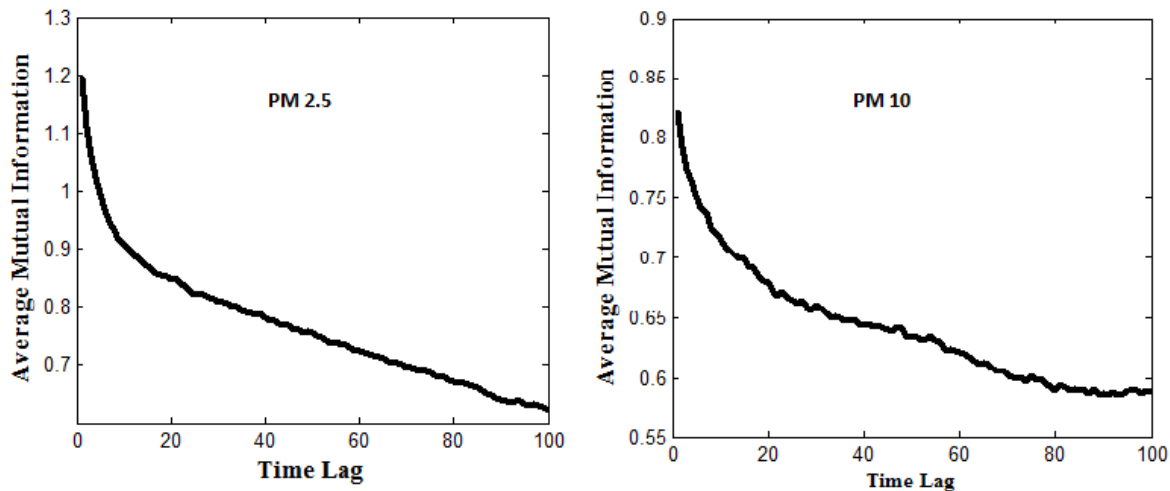
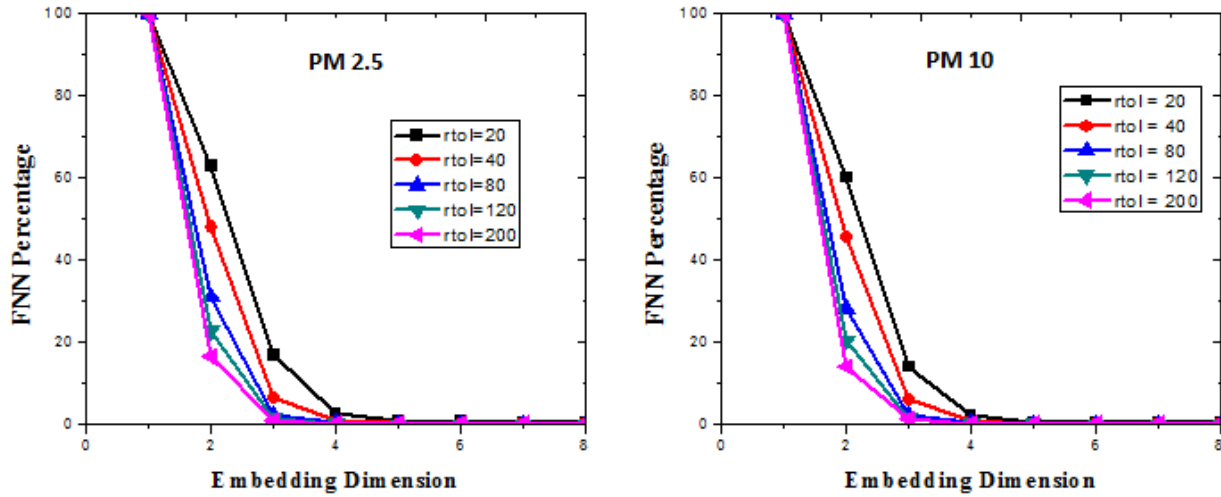


Figure 1: The optimal time delay τ , for particulates PM_{2.5} and PM₁₀ using AMI method.

Figure 2: False nearest neighbor curves with various thresholds for determining the embedding dimension of particulates $PM_{2.5}$ and PM_{10} Table 2: Embedding Dimension using FNN for particulates $PM_{2.5}$

Embedding dimension 'm'	$\theta = 0.2\varepsilon$				
	$\varepsilon = 20$	$\varepsilon = 40$	$\varepsilon = 80$	$\varepsilon = 120$	$\varepsilon = 200$
1	100	100	100	100	100
2	63	48	31	23	16
3	17	6	2	1	1
4	3	1	0	0	0
5	1	0	0	0	0
6	1	0	0	0	0
7	0	0	0	0	0

Table 3: Embedding Dimension using FNN for particulates PM_{10}

Embedding dimension 'm'	$\theta = 0.2\varepsilon$				
	$\varepsilon = 20$	$\varepsilon = 40$	$\varepsilon = 80$	$\varepsilon = 120$	$\varepsilon = 200$
1	100	100	100	100	100
2	60	45	28	20	14
3	14	6	2	1	1
4	2	1	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0

Table 4: LLE Values for all indoor sites, $N=21600$ for all indoor sites.

Particulates	Sites	Largest Lyapunov Exponent (λ)	Mean $\pm$ STD (λ)
$PM_{2.5}$	Bank Road	0.0078	0.0104 ± 0.0039
	Bus Stand	0.0063	
	Chattar Chowk	0.0112	
	Chehla Bridge	0.0089	
	CMH Chowk	0.0175	
	Old Secretariat	0.0105	
PM_{10}	Bank Road	0.0026	0.0055 ± 0.0027
	Bus Stand	0.0068	
	Chattar Chowk	0.0054	
	Chehla Bridge	0.0033	
	CMH Chowk	0.0101	
	Old Secretariat	0.0047	

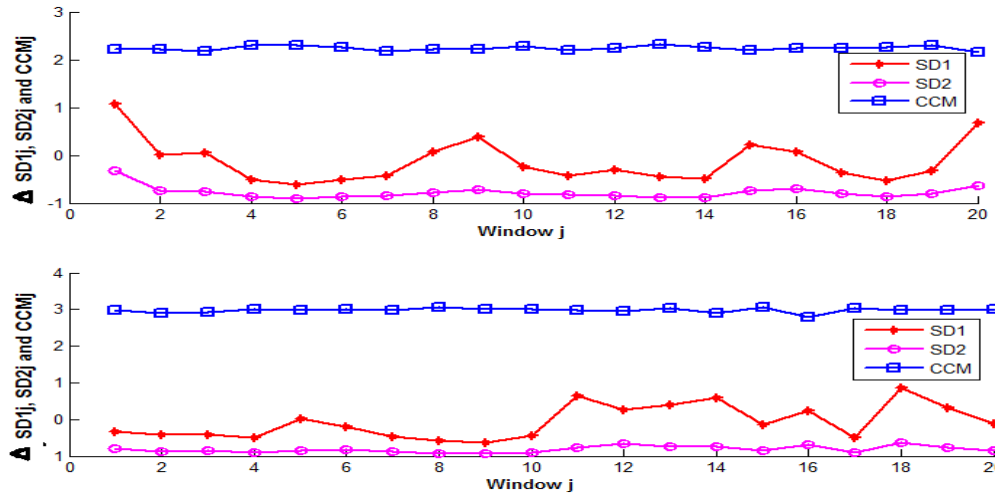


Figure 3: The sensitivity to the changes in temporal structure of three Poincare plot descriptors Sv, Lv and CCM (a) PM_{2.5} (b) PM₁₀.

Sensitivity to Changes in Temporal Structure using Poincare Plot Descriptors

The sensitivity to the changes in temporal structure of three Poincare plot descriptors Sv, Lv and CCM for PM_{2.5} and PM₁₀ are shown in Figure 3. The entire time series data was divided into 20 windows, resulting in 20 values of ΔS_v , ΔL_v and ΔCCM . It is evident from the figure that ΔCCM is much higher for both PM_{2.5} and PM₁₀ as compared to ΔS_v and ΔL_v . It manifests that Poincare plot descriptor CCM is much more sensitive to changes in temporal variations of the data sequence. The findings indicate that CCM may be used as an efficient feature for evaluating point to point variation of particulates in the atmosphere and to develop real time monitoring systems of particulates in the atmosphere.

Lyapunov Exponent

The Lyapunov exponent (λ) characterizes the nature of time evolution of close trajectories in the phase space and is considered as the key component of chaotic behavior. In the present study, Wolf's algorithm was used to calculate the largest Lyapunov exponent [19] using the same embedding parameters as before. In Table 4, the value of the largest Lyapunov exponent ' λ ' of PM_{2.5} and PM₁₀ at different sites along with their mean $\pm$ STD values are presented. The positive values of ' λ ' indicates an exponential divergence of the trajectories and hence a strong signature of chaos for both particulates.

Table 5: Regression Analysis and Prediction of Particulate Matter PM_{10.0} and PM_{2.5}

Method	RMSE	RS	MSE	MAE
Regression				
Linear	7.02	-11.46	49.33	5.14
Stepwise Linear	1.06	0.72	1.12	0.70
Tree				
Complex Tree	1.86	0.12	3.47	1.14
Medium Tree	1.99	0.00	3.96	1.62
Support Vector Regression (SVR)				
Linear SVR	0.52	0.93	0.27	0.45
Quadratic SVR	0.81	0.83	0.66	0.65
Cubic SVR	2.12	-0.13	4.48	1.16
Fine Gaussian	1.99	-0.00	3.97	1.63
Medium Gaussian	1.41	0.50	2.00	1.01
Coarse Gaussian	1.40	0.51	1.96	1.12
Ensemble				
Boosted Tree	2.00	-0.01	4.0	1.66
Bagged Tree	1.98	0.01	3.92	1.65
Gaussian Process Regression (GPR)				
Squared Exp. GPR	0.66	0.89	0.44	0.46
Matern 5/2 GPR	0.58	0.91	0.34	0.42
Exponential GPR	1.18	0.65	1.40	0.91
Rational Quadratic GPR	0.66	0.89	0.44	0.46

The linear regression is employed by minimizing the distance between fitted line and all other data points. Moreover, ordinary least square (OLS) regression is minimized through sum of square residuals. The best model fits on data is represented when there is smaller and unbiased difference between observed and predicted

values. Moreover, R-squared measure indicates that how close the data fitted to the regression line also known as coefficient of determination. The performance was measured in form of RMSE, R^2 , MAE and MSE. For predicting the Particulate Matters ($PM_{10.0}$ and $PM_{2.5}$), the robust Machine learning regression methods are employed such as Regression includes linear, interactive, & step-wise, ii) Support vector regress (SVR) includes linear, quadratic, cubic, fine Gaussian, coarse Gaussian, iii) Regression Tree; complex tree and simple tree, iv) Ensemble Regression; boosted tree, bagged tree v) Gaussian Process Regression (GPR); squared exponential GPR, Matern 5/2 GPR, exponential GPR, rational exponential GPR

The highest performance was obtained using SVR linear regression model with RMSE (0.52), R^2 (0.93), MSE

(0.27), and MAE (0.45) followed by Matern 5/2 GPR with RMSE (0.58), R^2 (0.91), MSE (0.34) and MAE (0.42); Squared exponential GPR & Rational quadratic GPR with RMSE (0.66), R^2 (0.89), MSE (0.44) and MAE (0.46); Quadrative SVR with RMSE (0.81), R^2 (0.83), MSE (0.66) and MAE (0.65). The performance in term of RMSE for rest of the regression models was obtained as Stepwise linear regression (RMSE 1.06) followed by Exponential GPR (RMSE 1.81), Coarse Gaussian (RMSE 1.40), Medium Gaussian SVR (RMSE 1.41), Complex tree (RMSE 1.86), Bagged tree (RMSE 1.89), Medium tree & Fine Gaussian SVR (RMSE 1.99), Boosted tree (RMSE 2.00), Cubic SVR (RMSE 2.12) ad linear regression (RMSE 7.02)

Table 6: Regression Analysis results to predict PM based on different features by optimizing parameters

PF	TET	TOFET	BOFP	OOFV	FET	BEFP	EOFV	EFET
Regression Tree								
SDSD	133.48	8.30	3	0.0025541	0.4425	3	0.0025541	0.092335
TP	119.35	6.743	2	28.9728	0.4766	2	28.9728	0.11158
MKDE	150.30	24.79	3	0.0003777	0.3936	3	0.0003777	0.09990
LogEn.	149.027	19.30	2	20.1643	0.31507	2	20.1643	0.09967
Var.	90.95	5.5853	2	0.0002939	0.3080	2	0.0002939	0.09442
Gaussian Process Regression (GPR)								
SDSD	50.227	5.2316	0.010306	0.00028981	0.22078	0.000929	0.00029	0.00029
TP	50.82	5.9016	1.2787	19.7885	0.1838	1627	19.7843	0.19223
MKDE	58.390	9.7709	0.01608	0.0007878	0.15217	0.016247	0.0007878	0.21857
LogEn.	62.810	12.007	0.27425	19.971	0.41503	104.01	19.9704	0.26456
Var.	52.058	5.6543	0.00225	3.7305	0.17283	0.00256	3.73e-5	0.18587
Support Vector Regression (SVR)								
SDSD	162.56	100.6526	1.0013	0.003657	2.7502	1.0013	0.0036569	0.25431
TP	205.57	59.2104	840.56	14.9188	2.5336	845.29	14.9029	1.0541
MKDE	189.8140	73.6534	1.5121	0.0009370	0.10145	0.021298	0.00086	0.075
LogEn.	245.31	129.1313	39.336	8.6735	4.547	39.33	8.6952	4.8549
Var.	90.3693	37.2189	0.004042	0.000319	5.8678	0.0005013	0.0003998	0.09356

Legends:

Predicted Feature (PF), KD tree Entropy (KDE), SDSD (Standard Deviation of Standard Deviation), Total Power (TP), LogEnergy (LE), Variance (Var.), Total elapsed time (TET), Total objective function evaluation time (TOFET), Best observed feasible point (BOFP), Observed objective function value (OOFV), Estimated objective function value (EOFV), Function evaluation time (FET), Best estimated feasible point (BEFP), Estimated objective function value (EOFV), Estimated function evaluation time (EFET)

BOFP/BEFP values denote (Regression tree: Minimum leaf node; GPR: Sigma & SVR: Epsilon)

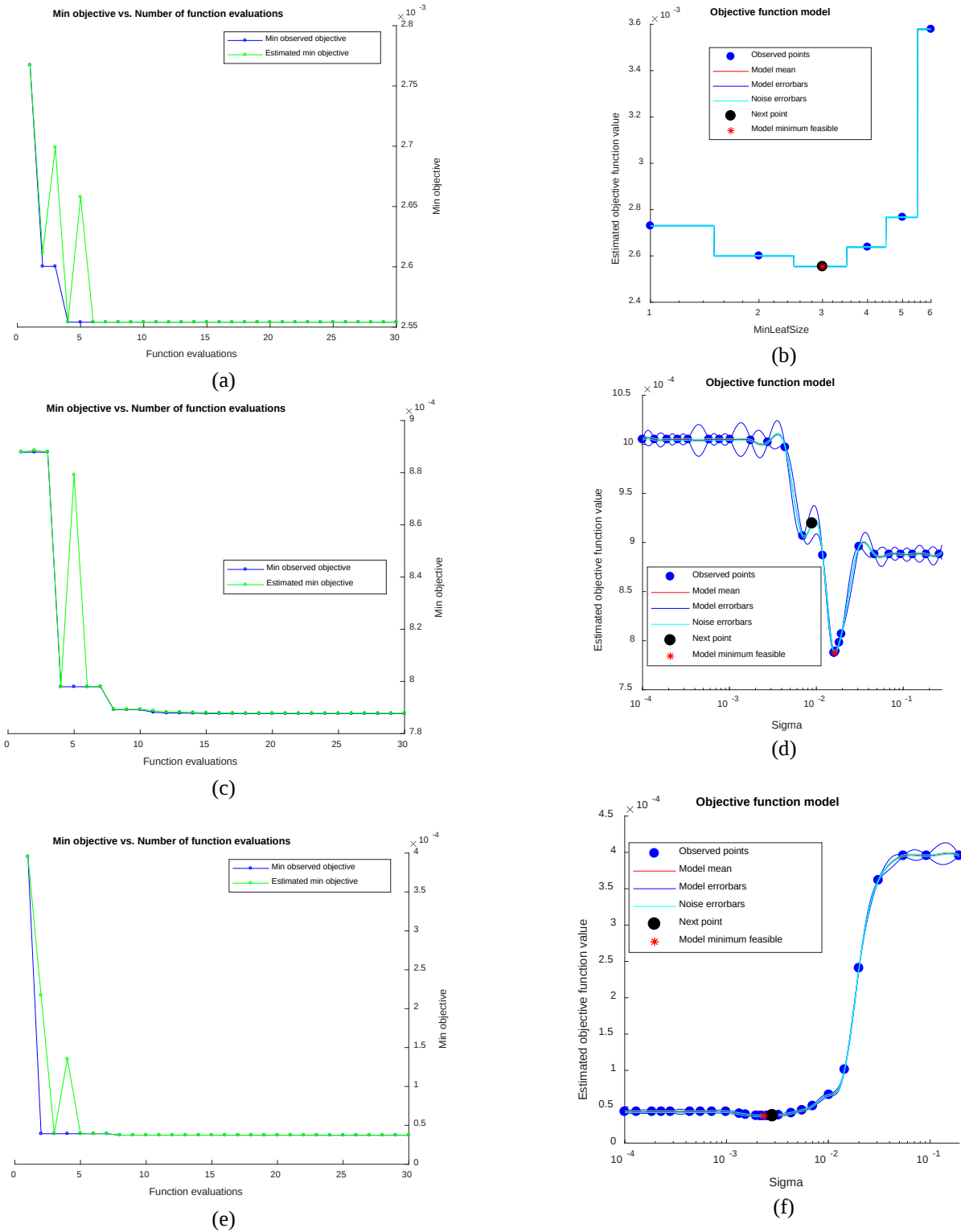


Fig. 4 Regression Analysis performance of minimum observed vs estimated objective functions with (a-b) SDSD features using Regression Tree, (c-d) MKDE features using GRP, & (e-f) variance features using GPR

The machine learning regression, support vector regression and gaussian regression process models were used for prediction based on selected features from 22 features computed from particulate matters PM10.0 and PM2.5. The performance was measured in terms of TET, TOFET, OOFV, EOFV, FET, BEFP, EOFV and EFET etc as reflected in Table 6 and Figure 4 (a-f). Using regression tree model, the best feasible point i.e. minimum leaf nodes for observed and estimated values are found same as reflected in the Table 6 for all selected set of features used in the prediction. Moreover, observed and estimated function evaluation values against each selected feature are also same. To predict the SDSD feature value using regression tree, the BOFP and BEFP is 3 as reflected in Table 6 and Figure 4 (b). The other features such as model minimum feasible, next point, observed points, model mean and error bar are also reflected in the Figure 4 (b, d, f). Using gaussian regression process model, the best observed and estimated feasible points i.e. sigma values to predict the selected set of features are almost similar in most of the cases except TP and LogEn. as depicted in Table 6 and Figure 4. The feasible points to predict SDSD (BOFP: 0.010306, BEFP: 0.00929), TP (BOFP: 1.2787, BEFP: 1627), MKDE (BOFP: 0.01608, BEFP: 0.016247), Log Energy (BOFP: 0.27425, BEFP: 104.01), Variance (BOFP: 0.00525, BEFP: 0.00256). The prediction for MKDE and variance are reflected in the Figure 4 (c-d, e-f). Similarly, using GPR, the observed and estimated functional values have also minimal differences as reflected in Table 6 and Figure 4. The SRV observed and estimated feasible points and functional values to predict the selected set of features have also minimal differences as shown in the Table 6 and Figure 4. In SVR, the BOFP and BEFP denote the value of epsilon.

5. Conclusions

To conclude, the concentrations of indoor PM, specifically PM2.5 and PM10, in indoor environments were monitored at six different sites along roadways in the downtown city of Muzaffarabad. The temporal behavior of PM2.5 and PM10 was analyzed using statistical and nonlinear times series analysis techniques. The statistical measures showed alarmingly high concentration of both PM2.5 and PM10 in the vicinity of Muzaffarabad. The results indicate that PM2.5 and PM10 concentration time series can be modeled using phase space reconstruction by properly selecting the embedding parameters 'm' and 'τ'. The concentration time series of both particulates indicated dominance of some complexity in their concentrations as depicted by higher values of optimal embedding dimension. The positive largest Lyapunov exponent depicts strong chaotic signature in the system dynamics of both particulates. Furthermore, the sensitivity to the

changes in temporal structure of three Poincare plot descriptors Sv, Lv and CCM for PM2.5 and PM10 demonstrated that CCM is more sensitive to changes in temporal variation of the particulate matter time series data compared to descriptors Sv and Lv. The CCM may be affected by short term variation in particulate matter concentration, enabling it to capture short term dynamics of the signal. This feature makes CCM a valuable descriptor to study dynamical fluctuations of particulate matter concentrations that may have implications to develop real time monitoring systems of particulates in the atmosphere. The prediction of indoor particulate matters was made in terms of different errors methods using robust machine learning regression models. The temporal, spectral, complexity, wavelet and statistical features are extracted from indoor PM10.0 and PM2.5 to predict the behavior of these subjects. The performance evaluation results provide the optimal results between observed and predicted values for selected set of features.

References

- [1] Weng, Y. C., Chang, N. B., & Lee, T. Y. (2008). Nonlinear time series analysis of ground-level ozone dynamics in Southern Taiwan. *Journal of environmental management*, 87(3), 405-414
- [2] Chen, J. L., Islam, S., & Biswas, P. (1998). Nonlinear dynamics of hourly ozone concentrations: nonparametric short-term prediction. *Atmospheric environment*, 32(11), 1839-1848.
- [3] Ostro, B. D., Broadwin, R. A. C. H. E. L., & Lipsett, M. J. (2000). Coarse and fine particles and daily mortality in the Coachella Valley, California: a follow-up study. *Journal of Exposure Science and Environmental Epidemiology*, 10(5), 412.
- [4] Schwartz, J., Norris, G., Larson, T., Sheppard, L., Claiborne, C., & Koenig, J. (1999). Episodes of high coarse particle concentrations are not associated with increased mortality. *Environmental Health Perspectives*, 107(5), 339.
- [5] Schwartz, J., & Neas, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology*, 11(1), 6-10.
- [6] Laden F, Neas LM, Dockery DW and Schwartz J. Association of fine particulate matter from different sources with daily mortality in six US cities. *Environmental health perspectives*. 2000; 108: 941.
- [7] Mar, T. F., Ito, K., Koenig, J. Q., Larson, T. V., Eatough, D. J., Henry, R. C., ... & Stölzel, M. (2006). PM source apportionment and health effects. 3. Investigation of inter-method variations in associations between estimated source contributions of PM2.5 and daily mortality in Phoenix, AZ. *Journal of Exposure Science and Environmental Epidemiology*, 16(4), 311-320.
- [8] Janssen, N. A., de Hartog, J. J., Hoek, G., Brunekreef, B., Lanki, T., Timonen, K. L., & Pekkanen, J. (2000). Personal exposure to fine particulate matter in elderly subjects: relation between personal, indoor, and outdoor

- concentrations. *Journal of the Air & Waste Management Association*, 50(7), 1133-1143.
- [9] Albalak, R., Keeler, G. J., Frisanchi, A. R., & Haber, M. (1999). Assessment of PM₁₀ concentrations from domestic biomass fuel combustion in two rural Bolivian highland villages. *Environmental science & technology*, 33(15), 2505-2509.
 - [10] Naeher, L. P., Smith, K. R., Leaderer, B. P., Neufeld, L., & Mage, D. T. (2001). Carbon monoxide as a tracer for assessing exposures to particulate matter in wood and gas cookstove households of highland Guatemala. *Environmental science & technology*, 35(3), 575-581.
 - [11] Cox, C. S. (1995). Physical aspects of bioaerosol particles. *Bioaerosols handbook*, 15-25.
 - [12] Lee, C. K., & Lin, S. C. (2008). Chaos in air pollutant concentration (APC) time series. *Aerosol and Air Quality Research*, 8(4), 381-391.
 - [13] Lee, R. E. (1972). The size of suspended particulate matter in air. *Science*, 178(4061), 567-575.
 - [14] Tan, Z., & Zhang, Y. (2004). A review of effects and control methods of particulate matter in animal indoor environments. *Journal of the Air & Waste Management Association*, 54(7), 845-854.
 - [15] Zhiqiang, Q., Siegmann, K., Keller, A., Matter, U., Scherrer, L., & Siegmann, H. C. (2000). Nanoparticle air pollution in major cities and its origin. *Atmospheric Environment*, 34(3), 443-451.
 - [16] Repace, J. L. (1980). Indoor air pollution, tobacco smoke, and public. *Science*, 208, 464.
 - [17] Kado, N. Y., Colome, S. D., Kleinman, M. T., Hsieh, D. P., & Jaques, P. (1994). Indoor-outdoor concentrations and correlations of PM₁₀-associated mutagenic activity in nonsmokers' and asthmatics' homes. *Environmental science & technology*, 28(6), 1073-1078.
 - [18] Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture notes in mathematics*, 898(1), 366-381.
 - [19] Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3), 285-317.
 - [20] Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical review letters*, 50(5), 346.
 - [21] Abarbanel, H. D., Brown, R., Sidorowich, J. J., & Tsimring, L. S. (1993). The analysis of observed chaotic data in physical systems. *Reviews of modern physics*, 65(4), 1331.
 - [22] Jaeger, L., & Kantz, H. (1997). Homoclinic tangencies and non-normal Jacobians—Effects of noise in nonhyperbolic chaotic systems. *Physica D: Nonlinear Phenomena*, 105(1-3), 79-96.
 - [23] Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2), 1134.
 - [24] Xie, X., Cao, Z., & Weng, X. (2008). Spatiotemporal nonlinearity in resting-state fMRI of the human brain. *Neuroimage*, 40(4), 1672-1685.
 - [25] Kugiumtzis, D. (1996). State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D: Nonlinear Phenomena*, 95(1), 13-28.
 - [26] Kugiumtzis, D. (1996). State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D: Nonlinear Phenomena*, 95(1), 13-28.
 - [27] Chelani, A. B., Gajghate, D. G., ChalapatiRao, C. V., & Devotta, S. (2010). Particle size distribution in ambient air of Delhi and its statistical analysis. *Bulletin of environmental contamination and toxicology*, 85(1), 22-27.
 - [28] Horák, J. (1992). Non-linear models of geophysical hydrodynamics and the problem of forecasting stream patterns II. *Czechoslovak journal of physics*, 42(7), 713-739.
 - [29] Jaeger, L., & Kantz, H. (1997). Homoclinic tangencies and non-normal Jacobians—Effects of noise in nonhyperbolic chaotic systems. *Physica D: Nonlinear Phenomena*, 105(1-3), 79-96.
 - [30] Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2), 1134.
 - [31] Xie, X., Cao, Z., & Weng, X. (2008). Spatiotemporal nonlinearity in resting-state fMRI of the human brain. *Neuroimage*, 40(4), 1672-1685.
 - [32] Kugiumtzis, D. (1996). State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D: Nonlinear Phenomena*, 95(1), 13-28.
 - [33] Chelani, A. B., Gajghate, D. G., ChalapatiRao, C. V., & Devotta, S. (2010). Particle size distribution in ambient air of Delhi and its statistical analysis. *Bulletin of environmental contamination and toxicology*, 85(1), 22-27.
 - [34] Voss, A., Schulz, S., Schroeder, R., Baumert, M., & Caminal, P. (2009). Methods derived from nonlinear dynamics for analysing heart rate variability. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1887), 277-296.
 - [35] Mourot, L., Bouhaddi, M., Perrey, S., Cappelle, S., Henriot, M. T., Wolf, J. P., ... & Regnard, J. (2004). Decrease in heart rate variability with overtraining: assessment by the Poincaré plot analysis. *Clinical physiology and functional imaging*, 24(1), 10-18.
 - [36] Fell, J., Röschke, J., & Schäffner, C. (1996). Surrogate data analysis of sleep electroencephalograms reveals evidence for nonlinearity. *Biological cybernetics*, 75(1), 85-92.
 - [37] Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability?. *IEEE transactions on biomedical engineering*, 48(11), 1342-1347.
 - [38] Horák, J. (1992). Non-linear models of geophysical hydrodynamics and the problem of forecasting stream patterns II. *Czechoslovak journal of physics*, 42(7), 713-739.
 - [39] Tulppo, M. P., Makikallio, T. H., Takala, T. E., Seppanen, T. H. H. V., & Huikuri, H. V. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American journal of physiology-heart and circulatory physiology*, 271(1), H244-H252.
 - [40] Karmakar, C. K., Khandoker, A. H., Gubbi, J., & Palaniswami, M. (2009). Complex Correlation Measure: a novel descriptor for Poincaré plot. *Biomedical engineering online*, 8(1), 17.
 - [41] Piskorski, J., & Guzik, P. (2005). Filtering Poincaré plots. *Computational methods in science and technology*, 11(1), 39-48.

- [42] Rathore, S., Hussain, M., Iftikhar, M. A., & Jalil, A. (2014). Ensemble classification of colon biopsy images based on information rich hybrid features. *Computers in Biology and Medicine*, 47, 76-92.
- [43] Ferland, R. J., Smith, J., Papandrea, D., Gracias, J., Hains, L., Kadiyala, S. B., ... & Herron, B. J. (2017). Multidimensional genetic analysis of repeated seizures in the hybrid mouse diversity panel reveals a novel epileptogenesis susceptibility locus. *G3: Genes, Genomes, Genetics*, g3-117.
- [44] Dheeba, J., Singh, N. A., & Selvi, S. T. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics*, 49, 45-52.
- [45] Hussain, L., Aziz, W., Kazmi, Z. H., & Awan, I. A. (2014). Classification of Human Faces and Non Faces Using Machine Learning Techniques. *International Journal of Electronics and Electrical Engineering*, 2 (2), 116-123
- [46] Hussain, L., Shafi, I., Saeed, S., Abbas, A., Awan, I. A., Nadeem, S. A., ... & Rahman, B. (2017). A radial base neural network approach for emotion recognition in human speech. *IJCSNS*, 17(8), 52.
- [47] Hussain, L., Ahmed, A., Saeed, S., Rathore, S., Awan, I.A., Shah, S.A., Majid, A. Idris, A., Awan, A.A. (2017). Prostate Cancer Detection using Machine Learning Techniques by Employing Combination of Features Extracting Strategies. *Cancer Biomarker*, DOI: 10.3233/CBM-170643.
- [48] Hussain, L., Aziz, W., Nadeem, S. A., & Abbasi, A. Q. (2014). Classification of Normal and Pathological Heart Signal Variability Using Machine Learning Techniques. *International Journal of Darshan Institute on Engineering Research and Emerging Technologies*, 3(2), 13-18.
- [49] Hussain, L., Aziz, W., Khan, A.S., Abbasi, A.Q., Kazmi, ZH., Abbasi, MM. (2015). Classification of Electroencephalography (EEG) Alcoholic and Control Subjects using Machine Learning Ensemble Methods. *Journal of Multidisciplinary Engineering Science and Technology*, 2(1): 126-131
- [50] Van Hoogenhuyze, D., Weinstein, N., Martin, G. J., Weiss, J. S., Schaad, J. W., Sahyouni, X. N., ... & Singer, D. H. (1991). Reproducibility and relation to mean heart rate of heart rate variability in normal subjects and in patients with congestive heart failure secondary to coronary artery disease. *The American journal of cardiology*, 68(17), 1668-1676.
- [51] Tuininga, Y. S., Van Veldhuisen, D. J., Brouwer, J., Haaksma, J., Crijs, H. J., & Lie, K. I. (1994). Heart rate variability in left ventricular dysfunction and heart failure: effects and implications of drug treatment. *Heart*, 72(6), 509-513.
- [52] Bilchick, K. C., Fetcs, B., Djoukeng, R., Fisher, S. G., Fletcher, R. D., Singh, S. N., ... & Berger, R. D. (2002). Prognostic value of heart rate variability in chronic congestive heart failure (Veterans Affairs' Survival Trial of Antiarrhythmic Therapy in Congestive Heart Failure). *The American journal of cardiology*, 90(1), 24-28.
- [53] Ponikowski, P., Anker, S. D., Chua, T. P., Szelemiej, R., Piepoli, M., Adamopoulos, S., ... & Coats, A. J. (1997). Depressed heart rate variability as an independent predictor of death in chronic congestive heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *The American journal of cardiology*, 79(12), 1645-1650.
- [54] Lee, S. H., Lim, J. S., Kim, J. K., Yang, J., & Lee, Y. (2014). Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance. *Computer methods and programs in biomedicine*, 116(1), 10-25.
- [55] Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, 187(2), 1017-1026.
- [56] Faust, O., Acharya, U. R., Adeli, H., & Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure*, 26, 56-64.
- [57] Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE transactions on information technology in biomedicine*, 13(5), 703-710.
- [58] Hussain, L., Seed, S., Awan, I.A., & Idris, A. (2017). Multiscaled Complexity Analysis of EEG Epileptic Seizure using entropy based computational techniques. *Archives of Neuroscience*, doi: 10.5812/archneurosci. 61161.
- [59] Hussain, L., Aziz, W., Saeed, S., Shah, S. A., Nadeem, M. S. A., Awan, I. A., ... & Kazmi, S. Z. H. (2017). Quantifying the dynamics of electroencephalographic (EEG) signals to distinguish alcoholic and non-alcoholic subjects using an MSE based Kd tree algorithm. *Biomedical Engineering/Biomedizinische Technik*. DOI: <https://doi.org/10.1515/bmt-2017-0041>
- [60] Hussain, L., Aziz, W., Alowibdi, J. S., Habib, N., Rafique, M., Saeed, S., & Kazmi, S. Z. H. (2017). Symbolic time series analysis of electroencephalographic (EEG) epileptic seizure and brain dynamics with eye-open and eye-closed subjects during resting states. *Journal of physiological anthropology*, 36(1), 21.
- [61] Zeng, J., & Qiao, W. (2013). Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52, 118-127.
- [62] Santamaría-Bonfil, G., Reyes-Ballesteros, A., & Gershenson, C. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85, 790-809.
- [63] Santamaría-Bonfil, G., Frausto-Solís, J., & Vázquez-Rodarte, I. (2015). Volatility forecasting using support vector regression and a hybrid genetic algorithm. *Computational Economics*, 45(1), 111-133.
- [64] Wang, J., Qin, S., Zhou, Q., & Jiang, H. (2015). Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. *Renewable Energy*, 76, 91-101.
- [65] Argentesi, E., & Filistrucchi, L. (2007). Estimating market power in a two-sided market: The case of newspapers. *Journal of Applied Econometrics*, 22(7), 1247-1266.
- [66] Zeng, J., & Qiao, W. (2011, March). Support vector machine-based short-term wind power forecasting. In *Power Systems Conference and Exposition (PSCE)*, 2011 IEEE/PES (pp. 1-8). IEEE.
- [67] Cesa-Bianchi, N., Mansour, Y., & Shamir, O. (2015, June). On the complexity of learning with kernels. In *Conference on Learning Theory* (pp. 297-325).

- [68] Mohammadi, K., Shamshirband, S., Anisi, M. H., Alam, K. A., & Petković, D. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, 91, 433-441.
- [69] Zhu, X., & Genton, M. G. (2012). Short-Term Wind Speed Forecasting for Power System Operations. *International Statistical Review*, 80(1), 2-23.
- [70] Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., & Hussain, O. K. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied soft computing*, 13(2), 947-958.
- [71] Chen, K., & Yu, J. (2014). Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Applied Energy*, 113, 690-705.
- [72] Liu, D., Niu, D., Wang, H., & Fan, L. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, 62, 592-597.
- [73] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [74] Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G. A., & Torricelli, P. (2011). Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling*, 222(8), 1471-1478.
- [75] Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.
- [76] Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73(364), 699-705.
- [77] Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 56-66.
- [78] Butler, A., Haynes, R. D., Humphries, T. D., & Ranjan, P. (2014). Efficient optimization of the likelihood function in Gaussian process modelling. *Computational Statistics & Data Analysis*, 73, 40-52.
- [79] Kapoor, A., Grauman, K., Urtasun, R., & Darrell, T. (2007, October). Active learning with gaussian processes for object categorization. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (pp. 1-8). IEEE.
- [80] Bernardo, J., Berger, J., Dawid, A., & Smith, A. (1998). Regression and classification using Gaussian process priors. *Bayesian statistics*, 6, 475.

Biography of Correspondence Author



Dr. Lal Hussain is a Programmer at Quality Enhancement Cell, University of Azad Jammu and Kashmir, Muzaffarabad Pakistan. He obtained his MS in Communication and Networks from Iqra University, Islamabad, Pakistan in 2012 with Gold medal. He received Ph.D. from Department of Computer Science & Information Technology, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan in February 2016. He worked as visiting PhD researcher at Lancaster University UK for six months under HEC International Research Initiative Program and worked under the supervision of Dr. Aneta Stefanovska, Professor of Biomedical Physics, Physics Department, Lancaster University, UK. He is author of more than 10 publications of highly reputed peer reviewed and Impact Factor Journals. He presented various talks at Pakistan, UK and USA. His research interest includes Biomedical Signal Processing with concentration in complexity measures, Time-Frequency representation methods, Cross Frequency Coupling to analyze the dynamics of neurophysiological and physiological signals. His research interest also includes Neural Networks and Machine Learning classification, regression and prediction techniques for detection of cancer, epilepsy, Brain Dynamics and Diseases (i.e. autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD), Alzheimer's Disease, Brain Tumor), image processing and segmentation etc.