J. Software Engineering & Applications, **2010, 3: 374-383**
doi:10.4236/jsea.2010.34042 Published Online April 2010 (http://www.SciRP.org/journal/jsea)

Time Series Forecasting of Hourly PM10 Using Localized Linear Models

Athanasios Sfetsos, Diamando Vlachogiannis

Environmental Research Laboratory, INTR-P, National Centre for Scientific Research "Demokritos", Attikis, Greece.
Email: ts@ipta.demokritos.gr

Received October 15th, 20009; revised November 4th, 2009; accepted November 8th, 2009.

ABSTRACT

*The present paper discusses the application of localized linear models for the prediction of hourly PM10 concentration values. The advantages of the proposed approach lies in the clustering of the data based on a common property and the utilization of the target variable during this process, which enables the development of more coherent models. Two alternative localized linear modelling approaches are developed and compared against benchmark models, one in which data are clustered based on their spatial proximity on the embedding space and one novel approach in which grouped data are described by the same linear model. Since the target variable is unknown during the prediction stage, a complimentary pattern recognition approach is developed to account for this lack of information. The application of the developed approach on several PM*10 *data sets from the Greater Athens Area, Helsinki and London monitoring networks returned a significant reduction of the prediction error under all examined metrics against conventional forecasting schemes such as the linear regression and the neural networks.*

Keywords: *Localized Linear Models, PM*10 *Forecasting, Clustering Algorithms*

1. Introduction

Environmental health research has demonstrated that Particulate Matter (PM) is a top priority pollutant when considering public health. Studies of long-term exposure to air pollution, mainly to PM, suggest adverse long- and short-term health effects, increased mortality (e.g. [1,2]), increased risk of respiratory and cardiovascular related diseases (e.g. [3]), as well as increased risk of developing various types of cancer [4]. Hence, the development and use of accurate and fast models for forecasting PM values reliably is of immense interest in the process of decision making and modern air quality management systems.

In order to evaluate the ambient air concentrations of particulate matter, a deterministic urban air quality model should include modelling of turbulent diffusion, deposition, re-suspension, chemical reactions and aerosol processes. In recent years, an emerging trend is the application of Machine Learning Algorithms (MLA), and particularly, that of the Artificial Neural Networks (ANN) as a means to generate predictions from observations in a location of interest. The strength of these methodologies lies in their ability to capture the underlying characteristics of the governing process in a non-linear manner, without making any predefined assumptions about its properties and distributions. Once the final models have been determined, it is then a straight-forward and exceedingly fast process to generate predictions. However, ANN have also inherent limitations. The main one is the extension of models in terms of time period and location; this always requires training with locally measured data. Moreover, these models are not capable of predicting spatial concentration distributions.

Owing to the importance and significant concentrations of PM in major European cities, there is an increasing amount of literature concerned with the application of statistical models for the prediction of point PM values. For the purposes of the EU-funded project APPETISE, an inter-comparison of different air pollution forecasting methods was carried out in Helsinki [5]. Neural networks demonstrated a better forecasting accuracy than other approaches such as linear regression and deterministic models.

In [6], Perez *et al.* compared predictions produced by three different methods: a multilayer neural network, linear regression and persistence methods. The three methods were applied to hourly averaged PM2.5 data for the years of 1994 and 1995, measured at one location in the downtown area of Santiago, Chile. The prediction errors for the hourly PM2.5 data were found to range from 30% to 60% for the neural network, from 30% to 70% for the persistence approach, and from 30% to 60% for the linear regression, concluding however that the neural network gave overall the best results in the prediction of the hourly concentrations of PM2.5.

In [7], Gardner undertook a model inter-comparison using Linear Regression, feed forward ANN and Classification and Regression Tree (CART) approaches, in application to hourly PM10 modelling in Christchurch, New Zealand (data period: 1989-1992). The ANN method outperformed CART and Linear Regression across the range of performance measures employed. The most important predictor variables in the ANN approach appeared to be the time of day, temperature, vertical temperature gradient and wind speed.

In [8], Hooyberghs *et al.* presented an ANN for forecasting the daily average PM10 concentrations in Belgium one day ahead. The particular research was based upon measurements from ten monitoring sites during the period 1997-2001 and upon the ECMWF (European Centre for Medium-Range Weather Forecasts) simulations of meteorological parameters. The most important input variable identified was the boundary layer height. The extension of this model with further parameters showed only a minor improvement of the model performance. Day-to-day fluctuations of PM10 concentrations in Belgian urban areas were to a larger extent driven by meteorological conditions and to a lesser extent by changes in anthropogenic sources.

In [9], Ordieres *et al.* analyzed several neural-network methods for the prediction of daily averages of PM2.5 concentrations. Results from three different neural networks (feed forward, Radial Basis Function (RBF) and Square Multilayer Perceptron) were compared to two classical models. The results clearly demonstrated that the neural approach not only outperformed the classical models but also showed fairly similar values among different topologies. The RBF shows up to be the network with the shortest training times, combined with a greater stability during the prediction stage, thus characterizing this topology as an ideal solution for its use in environmental applications instead of the widely used and less effective ANN.

The problem of the prediction of PM10 was addressed in [10], using several statistical approaches such as feed-forward neural networks, pruned neural networks (PNNs) and Lazy Learning (LL).The models were designed to return at 9 a.m. the concentration estimated for the current day. The forecast accuracy of the different models was comparable. Nevertheless, LL exhibited the best performances on indicators related to average goodness of the prediction, while PNNs were superior to the other approaches in detecting the exceedances of alarm and attention thresholds.

In view of the recent developments in PM forecasting, the present paper introduces an innovative approach based on localized linear modelling. Specifically, two alternative localized liner modelling approaches are developed and compared against benchmark models such as the linear regression and the artificial neural networks. The advantage of the proposed approach is the identification of the finer characteristics and underlying properties of the examined data set through the use of suitable clustering algorithms and the subsequent application of a customized linear model on each one. Furthermore, the use of the target variable in the clustering stage enhances the coherence of the localized models. The developed approach is applied on several data sets from the monitoring networks of the Greater Athens Area and Helsinki, during different seasons.

2. Modelling Approaches

Time series analysis is used for the examination of a data set organised in sequential order so that its predominant characteristics are uncovered. Very often, time series analysis results in the description of the process through a number of equations (Equation (1)) that in principle combine the current value of the series, y_t, to lagged values, y_{t-k}, modelling errors, e_{t-m}, exogenous variables, $\mathbf{x}_{t-j}$, and special indicators such as time of the day. Thus, the generalized form of this process could be written as follows:

$$y_t = f\,(y_{t-k},\ \mathbf{x}_{t-j},\ e_{t-m} \mid \textit{various } k,j,m \textit{ and special indicators}) \qquad (1)$$

2.1 Linear Regression

This approach uses linear regression models to determine whether a variable of interest, y_t, is linearly related to one or more exogenous variable, $\mathbf{x}_t$, and lagged variables of the series, y_t. The expression that governs this model is the following:

$$y_t = c + \sum^{k} \beta_k y_{t-k} + \sum^{j} \gamma_j \mathbf{x}_{t-j} + \varepsilon \qquad (2)$$

The coefficients c, β, γ are usually estimated from a least squares algorithm. The inputs should be a set of statistically significant variables, defined under Student's t-test, estimated from the examination of the correlation coefficients or using a backward elimination selection procedure from a larger initial set.

2.2 Artificial Neural Networks (ANN)

The multi-layer perceptron or feed-forward ANN [11] has a large number of processing elements called neurons, which are interconnected through weights (w_{iq}, v_{qj}). The neurons expand in three different layer types: the input, the output, and one or more hidden layers. The signal flow is from the input layer towards the output. Each neuron in the hidden and output layer is activated by a nonlinear function that relies on a weighted sum of its inputs and a neuron-specific parameter, called bias, b. The response of a neuron in the output layer as a function of its inputs is given by Equation (3), where f_1 and f_2 can be sigmoid, linear or threshold activation functions.

$$y_i = f_1 \left(\sum_{q=1}^{l} w_{iq} f_2 (\sum_{j=1}^{m} (v_{qj} x_j + b_j)) + b_q \right) \quad (3)$$

The strength of neural networks lies in their ability to simulate any given problem from the presented example, which is achieved from the modification of the network parameters through learning algorithms. In this study, the Levenberg-Marquardt [12] algorithm is applied because of its speed and robustness against the conventional back-propagation.

The most important issue concerning the introduction of ANN in time series forecasting is "generalization", which refers to their ability to produce reasonable forecasts on data sets other than those used for the estimation of the model parameters. This problem has two important parameters that should be accounted for. The first is data preparation, which involves pre-processing and the selection of the most significant variables. The second embraces the determination of the optimum model structure that is closely related with the estimation of the model parameters. Although, there is no systematic approach, which can be followed [13], some useful insight can be found using statistical methods such as the correlation coefficients.

The second aspect can be jointly tackled under the cross-validation training scheme. The data set is split into three smaller sets the training (TS), the evaluation or validation (ES) and the prediction or testing (PS) sets. The model is initialized with a few parameters. The next step is to train the model using data from the training set and when the error of the evaluation set is minimized, the model parameters and configuration are stored. The number of parameters is then increased and a new network is trained from the beginning. If ES error is lower compared to the previously found minimum, then the parameters of this new model are stored. This iterative process is terminated when a predefined number of iterations are reached (**Figure 1**).

In this study, ES was formed using a Euclidean metric withholding a percent value (here 25% is used) of the TS data that are located nearest to other data. The strength of this approach lies in the fact that TS covers more distinct characteristics of the process, thus, allowing for the development of a model with better generalization capabilities.

2.3 Nearest Neighbours

This class of hybrid models includes a local modelling and a function approximation to capture recent dynamics of the process. The underlying aim of these predictors is that segments of the series neighbouring under some distance measure may correspond to similar future values. This claim was endorsed by the work of Farmer and Sidorowich [14] that showed that the chaotic time-series prediction is several orders of magnitude better using local approximation techniques rather than universal approximators. The tricky part in these models is the selection of the embedding dimension, which effectively determines segments of the series, and the number of neighbours. Initially, it is required to estimate the embedding dimension d and time delay τ of the attractor as follows:

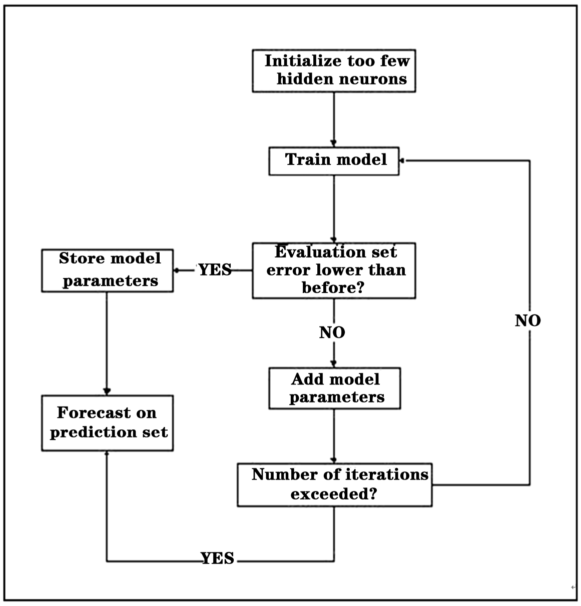
Initialize too few hidden neurons
Train model
Evaluation set error lower than before?
YES
Store model parameters
NO
NO
Add model parameters
Forecast on prediction set
Number of iterations exceeded?
YES

Figure 1. Iterative cross-validation training

$$Y(t) = \left[y(t-\tau), ..., y(t-(d-1)\tau), \mathbf{x}(\mathrm{t}-\mathrm{j}) \right] \quad (4)$$

In this study, a value of $\tau = 1$ was used and $Y(t)$ had the same parameters as the linear regression model. The number of neighbours was not pre-determined but was set to vary between predefined limits. A small number of neighbours increase the variance of the results whereas a large number can compromise the local validity of a model and increase the bias of results. Once the nearest neighbours to $Y(t)$ have been identified, an averaging procedure is followed in the present study to generate predictions.

2.4 Local Models with Clustering Algorithms (LMCA)

The idea behind the application of clustering algorithms in time series analysis is to identify groups of data that share some common characteristics. On each of these groups, the relationships amongst the members are modelled through a single equation model. Consequently, each of the developed models has a different set of parameters. The process is described in the following steps:

1) Selection of the input data for the clustering algorithm. This can contain lagged and/or future characteristics of the series, as well as other relevant information.

C(t) = [y_t, $y_{t-k,}$ $\mathbf{x}_{t-j}$]. Empirical evidence suggests that the use of the target variable y_t is very useful to discover unique relationships between input-output features. Additionally, higher quality modelling is ensured with the function approximation since the targets have similar properties and characteristics. However, this occurs to the expense of an additional process needed to account for this lack of information in the prediction stage.

2) Application of a clustering algorithm combined with a validity index or with user defined parameters, so that n_{cl} clusters will be estimated.

3) Assign all patterns from the training set to the $\mathrm{n_{cl}}$ clusters. For each of the clusters, apply a function approximation model, $y_t = f_i(y_{t-k}, \mathbf{x}_{t-j}),\ i = 1...n_{cl}$, so that n_{cl} forecasts are generated.

Successful application of this method has been reported on the prediction of locational electricity marginal prices [15], Mckay Glass and daily electric load peak series [16], the A and D series of the Santa Fe forecasting competition [17] and hourly electric load [18].

In this study, the k means clustering algorithm was selected [19]. It is a partitioning algorithm that attempts to directly decompose the data set into a set of groups through the iterative optimization of a certain criterion. More specifically, it re-estimates the cluster centres through the minimization of a distance-related function between the data and the cluster centres. The algorithm terminates when the cluster centres stop changing.

The optimal number of clusters is determined using a modified cluster validity index, CVI, [20], which is directly related to the determination of the user-defined (here the number of clusters) parameters of the clustering algorithm. Two indices are used for showing an estimate of under-partitioning (U_u) and over-partitioning (U_o) of the data set:

$$U_u = \frac{1}{c}\sum_{i=1}^{c} MD_i \qquad U_o = \frac{c}{d_{\min}} \tag{5}$$

MD_i is the mean intra-cluster distance of the i-th cluster. Here, $d_{\min}$ is the minimum distance between cluster centres, which is a measure of intra-cluster separation. The optimum number is found from the minimization of a normalized combinatory expression of these two indices.

2.5 Hybrid Clustering Algorithm (HCA)

The hybrid clustering algorithm is an iterative procedure that groups data, based on their distance from the hyper-plane that best describes their relationship. It is implemented through a series of steps, which are presented below:

1) Determine the most important variables.

2) Form the set of patterns H(t) = [y_t, y_{t-k}, $\mathbf{x}_{t-k}$].

3) Select the number of clusters n_h.

4) Initialize the clustering algorithm so that n_h clusters are generated and assign patterns.

5) For each new cluster, apply a linear regression model to y_t using as explanatory variables the remaining of the set H_t.

6) Assign each pattern to a cluster based on their distance.

7) Go to 5) unless any of the termination procedures is reached.

The following termination procedures are considered: a) the maximum pre-defined number of iterations is reached and b) the process is terminated when all patterns are assigned to the same cluster as in the previous iteration in 6). The selection of the most important lagged variables, 1), is based on the examination of the correlation coefficients of the data.

The proposed clustering algorithm is a complete time series analysis scheme with a dual output. The algorithm generates clusters of data, the identical characteristic of which is that they “belong” to the same hyper-plane, and synchronously, estimates a linear model that describes the relationship amongst the members of a cluster. Therefore, a set of n_h linear equations is derived (Equation (6)).

$$\hat{y}_{t,i} = a_{o,i} + \sum a_{i,k} y_{t-k} + \sum b_{i,j} X_{t-j} \quad , \ i = 1 \ldots n_h \tag{6}$$

Like any other hybrid model that uses the target variables in the development stage, the model requires a secondary scheme to account for this lack of information in the forecasting phase. For HCA and LMCA, the only requirement is the determination of the cluster number, n_h and n_{cl} respectively, which is equivalent to the estimation of the final forecast.

The optimum number of HCA clusters is found from a modified cluster validity criterion. An estimate of under-partition (U_u) of the data was formed using the inverse of the average value of the coefficient of determination (R_i^2) on all regression models. U_o indicates the over-partitioning of the data set, and $d_{\min}$ is the minimum distance between linear models (Equation (7)). The optimum number is found from the minimization of a normalized combinatory expression of these two indices.

$$U_u = \frac{1}{\frac{1}{h}\sum_{i=1}^{h} R_i^{\,2}} \qquad U_o = \frac{c}{d_{\min}} \tag{7}$$

2.6 Pattern Recognition

A pattern recognition scheme with three alternative approaches was then applied to convert the LMCA and HCA output to the final predictions. Initially, a conventional clustering (k-means) algorithm was employed to identify similar historical patterns in the time series. The second was to determine n_{cl}/ n_h at each time step, using information contained in the data of the respective cluster.

(p1) Select a second data vector using *only* historical observations $P_t = [y_{t-k}, \mathbf{x}_{t-k}]$

(p2) Initialize a number of clusters n_k

(p3) Apply a k-means clustering algorithm on P_t.

(p4) Assign data vectors to each cluster, so that each of the n_k clusters should contain k_m, m = 1,…, n_k data.

To obtain the final forecasts the following three alternatives were examined:

(M1) From the members of the *k*-th cluster find the most frequent LMCA / HCA cluster, *i.e.* n_{cl}/ n_h number.

(M2) From the members of the *k*-th cluster estimate the final forecast as a weighted average of the LMCA/ HCA clusters. Here p_i is the percentage of appearances of the LMCA / HCA cluster in the *k*-th cluster data.

$$y_t = \sum_i p_i y_{t,n} \qquad i = 1,...,k \ \ and \ \ n = 1,...,n_h \text{ or } n_{cl} \tag{8}$$

(M3) From the members of the *k*-th cluster estimate the final forecast as a distance weighted average of the HCA clusters.

$$y_t = \sum_i t_i y_{t,n} \qquad i = 1,...,k \ \ and \ \ n = 1,...,n_h \text{ or } n_{cl}$$

$$d_i = \left\| P_t - P_i \right\| \tag{9}$$

$$t_i = \frac{d_i^{-a}}{\sum_i d_i^{-a}} \text{ and } a = 2$$

The optimal number of clusters for the pattern recognition stage was determined using the modified compactness and separation criterion for the k-means algorithm discussed previously in section "Local Models with Clustering Algorithms".

3. Data Description and Results

The previously described forecasting methodologies were applied to eight different data sets both univariate and multivariate. The data sets were hourly PM10 concentration values from the monitoring network in the Greater Athens Area and in the cities of Helsinki and London, spanning over different seasons. It should be clarified that meteorological data were available only from the Helsinki station. The results returned by the applied algorithms for each station are discussed separately in the following sections.

In addition to the combined LMCA / HCA – PR methodology, the ideal case of a perfect knowledge of the n_{cl} / n_h parameter is also presented. This indicates the predictive potential, or the least error that the respective methodology could achieve. Also, the base-case persistent approach ($y_t = y_{t-1}$) is presented as a relative criterion for model inter-comparison amongst different data sets. The ability of the models to produce accurate forecasts was judged against the following statistical performance metrics:

Root Mean Square Error

$$RMS = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(O_i - P_i)^2} \tag{10}$$

Normalized RMS

$$NRMS = \frac{\sum_{i=1}^{k}(O_i - P_i)^2}{\sum_{i=1}^{k}(\overline{O} - O_i)^2} \tag{11}$$

Mean Absolute Percentage Error

$$MAPE = 100\frac{1}{k}\sum_{i=1}^{k}\frac{|O_i - P_i|}{O_i} \tag{12}$$

Index of Agreement

$$IA = 1 - \frac{\sum_{i=1}^{k}(O_i - P_i)^2}{\sum_{i=1}^{k}\left(|\overline{O} - P_i| + |\overline{O} - O_i|\right)^2} \tag{13}$$

Fractional Bias

$$FB = \frac{(\overline{O} - \overline{P})}{0.5 * (\overline{O} + \overline{P})} \tag{14}$$

3.1 Greater Athens Area – Aristotelous Str

The selected station from the Greater Athens Area monitoring network was Aristotelous Str. It is located at 23°43′39′′ North and 37°59′16′′ West, at an elevation height of 95 m above ground level. It is characterized as an urban station, positioned in the city centre with traffic dominated emissions. The training and the prediction sets covered the periods from 1/7/2001 to 14/8/2001 and 15/8/2001 to 31/8/2001, respectively.

The analysis revealed that the most influential variables were PM_{t-1}, PM_{t-2}, PM_{t-24}, PM_{t-25} and an indicator for the time of the day. This data set was used for the development of all methodologies and the input set for the pattern recognition scheme. The results on **Table 1** indicate that with the exception of NN, all other conventional approaches demonstrate a reduction of the prediction error by approximately 6% on the basis of the RMS error compared to the base case persistent method. The difference between LR and ANN was not found to be statistically significant, although the later was marginally better under all criteria.

The application of the local linear models was able to reduce the predictive error by an order of magnitude depending on the pattern recognition scheme that was applied. Both LMCA and HCA are capable of reaching exceedingly lower prediction error, with IA above 0.98, if all n_{cl}/n_h clusters are predicted correctly at each time step. **Figure 2** presents a graphical description of the prediction error of the HCA-perfect cluster forecast. The HCA coupled with the M3 scheme returned the overall best prediction error that was approximately 8% lower than that of the persistent approach.

3.2 Greater Helsinki Area – Kallio

The data from the Helsinki monitoring network were from the suburban station of Kallio, with co-ordinates 25°52′92′′ W and 66°75′47′′ N and elevation height of 21 m above sea level. The training set was from 3/9/2003 to 9/11/2003, whereas the unknown prediction set spanned from 10/11/2003 to 30/11/2003.

The developed models for the prediction of PM10 val-

Table 1. Prediction results from Aristotelous

	RMS	NRMS	MAPE	d	FB	Nu of clusters
Persistent	9.5596	0.3112	13.006	0.9223	–0.0002	
LR	9.0193	0.277	12.6536	0.9007	0.0052	
ANN	8.9311	0.2716	12.3984	0.9152	0.0037	
NN	10.117	0.3485	14.3699	0.892	–0.0094	24
LCMA	n_{cl} = 4					n_k = 32
Perfect	4.6355	0.0732	7.2108	0.9813	–0.0043	
M1	9.6748	0.3187	13.3351	0.8999	–0.0107	
M2	9.0637	0.2797	12.434	0.9121	–0.0052	
M3	9.0559	0.2793	12.3804	0.9108	–0.009	
HCA	n_{cl} = 8					n_k = 13
Perfect	2.1522	0.0158	2.857	0.9961	–0.0002	
M1	9.6085	0.3144	12.5104	0.9105	–0.0134	
M2	8.8787	0.2684	12.3668	0.915	0.0048	
M3	8.8153	0.2646	12.3368	0.9178	0.0046	

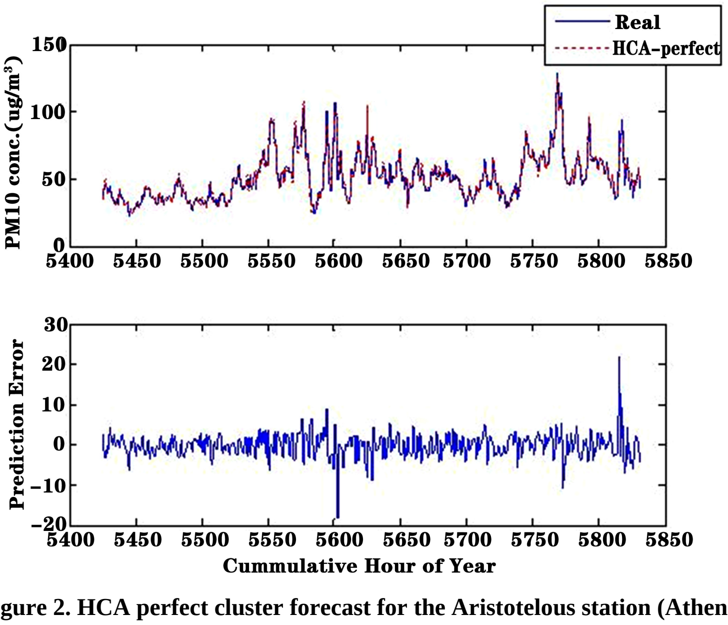

Figure 2. HCA perfect cluster forecast for the Aristotelous station (Athens)

Table 2. Linear regression model details for Helsinki 1

Variable	Coef.	St. Error	t-stat.	Variable	Coef.	St. Error	t-stat.
c	4.7626	1.1361	4.1921	T_{t-1}	1.0627	0.2753	3.8601
PM_{t-1}	0.7611	0.0247	30.8584	T_{t-2}	–1.0446	0.274	–3.8128
PM_{t-2}	0.0622	0.0246	2.5319	u_{t-1}	–0.749	0.2213	–3.3847
PM_{t-24}	0.0232	0.0136	1.7008	u_{t-2}	0.6094	0.2216	2.7493
RH_{t-1}	0.2055	0.0547	3.7582	v_{t-1}	0.6673	0.2242	2.9767
RH_{t-2}	-0.2361	0.0547	-4.317	v_{t-2}	–0.4508	0.2257	–1.9968

Table 3. Prediction Results from Helsinki 1

		RMS	NRMS	MAPE	d	FB	Nu of clusters
Persistent		5.1208	0.2793	33.3564	0.9301	0.0001	
LR		4.9654	0.2626	36.4317	0.9073	–0.0139	
ANN		5.1722	0.2849	39.5785	0.9085	–0.0591	
NN		5.6876	0.3446	43.8667	0.857	–0.0489	13
LCMA	n_{cl} = 3						n_k = 61
Perfect		3.033	0.098	18.1484	0.9724	0.0038	
M1		5.1044	0.2775	37.1176	0.9295	–0.0119	
M2		4.892	0.2549	37.5676	0.9193	0.0008	
M3		4.8416	0.2497	36.905	0.9229	0.0049	
HCA	n_{cl} = 7						n_k = 19
Perfect		1.5653	0.0261	8.9912	0.9932	–0.0051	
M1		5.2179	0.29	42.6351	0.9072	0.021	
M2		4.8139	0.2468	37.4036	0.9203	–0.0018	
M3		4.7612	0.2415	36.7128	0.9239	–0.0006	

ues from Helsinki contained meteorological parameters that were identified using a combination of statistical correlation properties and stepwise linear regression, discarding all those that were judged statistically as not significant under Student's t-test. The finally selected parameters and their estimation from the least squares fit are shown on **Table 2**.

The prediction results (**Table 3**) demonstrate that the forecasting ability of the conventional models is somewhat similar to that of the base-case persistent approach. The large prediction error of the ANN can be partly explained by the linear nature governing process that relates PM10 values to lagged values and from the over-fitting of the applied training scheme. The introduction of the LMCA and HCA localized models coupled with the M3 pattern recognition scheme returned the least overall prediction error that was approximately 5.5% and 7% respectively lower on the RMS criterion and double under NRMS. **Figure 3** shows the values of the prediction error of the LMCA-M3 modelling approach.

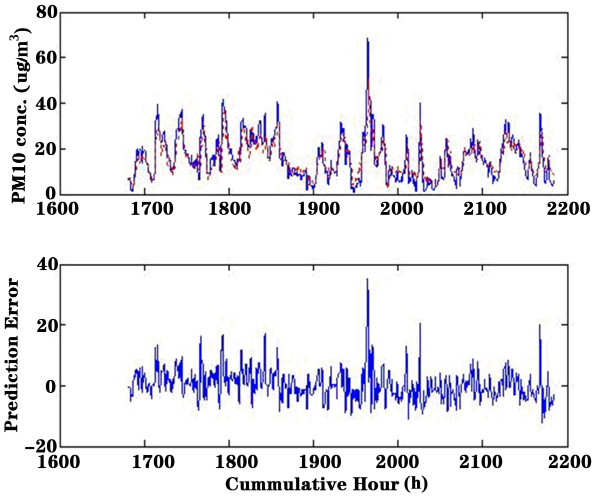

Figure 3. Prediction and error with LCMA – M3 approach

3.3 Greater London Area – Bloomsbury

The data from the Greater London Area were from the

Bloomsbury station located in the city centre of London (51°31'24" N, 0°7'54" W), characterised as an urban background station. The training set was selected to cover the period from 1/9/2005 to 22/10/2205, whereas the unknown prediction set comprised data ranging from 23/10/2005 to 6/11/2005.

The stepwise regression with a threshold value for the t-statistic of ± 1.96, corresponding to the 95% confidence interval, revealed as the most significant values PM_{t-1}, PM_{t-2}, PM_{t-24}. Additionally, an indicator for the time of the day was utilized. That data set was used for the development of all methodologies while the input set for the pattern recognition scheme. The analysis of the results (**Table 4**) indicated that none of the conventional forecasting approaches managed to return consistently lower prediction errors than the base case persistent approach. The least prediction error was returned from the ANN that was 3.6% lower than the persistent approach on the basis of the RMS error.

The developed localized linear model (HCA) has significant forecasting potential, as it can be observed in **Figure 4**, under the assumption of a perfect knowledge of the future cluster in the pattern recognition stage. The percentage improvement over the bench-mark persistent approach ranged from 40-70%. Similar results were found for the other two data sets

4. Discussion

The development and application of accurate models for forecasting PM concentration values in a rather fast and efficient manner is of primary concern in modern air quality management systems. The applied LR and ANN are nowadays mature approaches that have been integrated in many operational systems and could be used for the benchmarking of novel methodologies. The results of this work yielded that for the majority of the examined data sets, the linear approach marginally outperforms ANN. This indicates that the underlying process could possess predominantly linear characteristics.

The main focus of this work was the development and application of novel localized linear models. These were based on clustering algorithms as a means to identifying

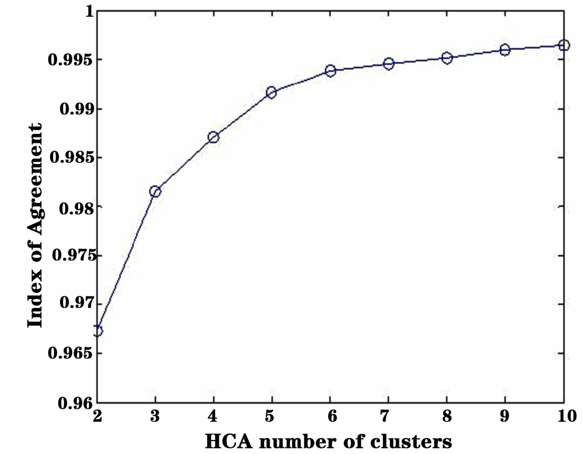

Figure 4. Index of agreement for HCA – perfect cluster forecast

Table 4. Prediction results from London Bloomsbury

	RMS	NRMS	MAPE	d	FB	Remarks
Persistent	4.4165	0.272	16.4202	0.9282	–0.0007	
LR	4.3119	0.2593	17.281	0.9257	0.0206	
ANN	4.256	0.2526	16.9101	0.9266	0.0221	
NN	5.1193	0.3655	22.466	0.8933	0.0075	14
LCMA	**n_{cl} = 4**					**n_k = 16**
Perfect	4.1665	0.2421	17.0711	0.9324	0.0193	
M1	4.4947	0.2817	17.9071	0.9137	–0.0136	
M2	4.2704	0.2543	17.051	0.9228	0.0069	
M3	4.3005	0.2579	17.9051	0.9229	0.021	
HCA	**n_{cl} = 7**					**n_k = 29**
Perfect	1.2401	0.0214	5.1061	0.9945	0.0064	
M1	4.3246	0.2608	16.8142	0.8877	–0.0188	
M2	4.2795	0.2554	16.8375	0.9163	0.0097	
M3	4.2513	0.2521	16.9179	0.8812	0.0046	

similar properties of the time series. The LMCA identified clusters based on their proximity on the embedding space, whereas HCA identified grouped points that were described by the same linear model. As both approaches included the target variable in the model development stage, a pattern recognition scheme was needed to account for this lack of information in the prediction stage.

The final prediction model was reached with the use of the modified CVI coupled with a pattern recognition scheme. The results suggested M3 as the most effective choice, because it produced consistently the least prediction error, under all metrics. For the RMS and MAPE errors, the improvement over the persistent approach ranged from 3.5% (London) to 7.7% (Athens and Helsinki). This value was almost doubled for NRMS and IA for the respective data sets. The HCA produced the least prediction error on every single examined data set, compared both to conventional approaches and the LCMA.

5. Conclusions

This paper introduced the application of localized linear models for forecasting hourly PM10 concentration values using data from the monitoring networks of the cities of Athens, Helsinki and London. The strength of this innovative approach is the use of a clustering algorithm that identifies the finer characteristics and the underlying relationships between the most influential parameters of the examined data set and subsequently, the development of a customized linear model. The calculated clusters incorporated the target variable in the model development phase, which was beneficial for the development of more coherent localized models. However, in order to overcome this lack of information in the prediction stage a complementary scheme was required. For the purposes of this study, a pattern recognition scheme based on the concept of weighted average distance (M3) was developed that consistently returned the least error under all examined metrics. The calculated results show that the proposed approach is capable of generating significantly lower prediction error against conventional approaches such as linear regression and neural networks, by at least one order of magnitude.

6. Acknowledgements

The assistance of members of OSCAR project (funded by EU under the contract EVK4-CT-2002-00083), for providing the Helsinki data for this research, is gratefully acknowledged. The Department of Atmospheric Pollution and Noise Control of the Hellenic Ministry of Environment, Physical Planning and Public Works is also acknowledged for the provision of the Athens data. The data from London are from the UK air quality archive (http://www.airquality.co.uk).

REFERENCES

[1] K. Katsouyanni, “Ambient Air Pollution and Health,” *British Medical Bulletin*, Vol. 68, 2003, pp. 143-156.

[2] E. Samoli, A. Analitis, G. Touloumi, J. Schwartz, H. R. Anderson, J. Sunyer, L. Bisanti, D. Zmirou,˙ J. M. Vonk, J. Pekkanen,˙ P. Goodman,˙ A. Paldy,˙ C. Schindler and K. Katsouyanni, “Estimating the Exposure-Response Relationships between Particulate Matter and Mortality within the APHEA Multicity Project,” *Environmental Health Perspectives*, Vol. 113, 2005, pp. 88-95.

[3] R. D. Morris, “Airborne Particulates and Hospital Admissions for Cardiovascular Disease: A Quantitative Review of the Evidence,” *Environmental Health Perspectives*, Vol. 109, Supplement 4, 2001, pp. 495-500.

[4] E. G. Knox and E. A. Gilman, “Hazard Proximities of Childhood Cancer in Great Britain from 1953-1980,” *Journal of Epidemiology and Health*, Vol. 51, 1997, pp. 151-159.

[5] J. Kukkonen, L. Partanen, A. Karppinen, J. Ruuskanen, H. Junninen, M. Kolehmainen, H. Niska, S. Dorling, T. Chatterton, R. Foxall and G. Cawley, “Extensive Evaluation of Neural 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*, Vol. 37, 2003, pp. 4539-4550.

[6] P. Perez, A. Trier and J. Reyes, “Prediction of PM2.5 Concentrations Several Hours in Advance Using Neural Networks in Santiago, Chile,” *Atmospheric Environment*, Vol. 34, 2000, pp. 1189-1196.

[7] M. W. Gardner, “The Advantages of Artificial Neural Network and Regression Tree Based Air Quality Models,” Ph.D. Dissertation, School of Environmental Sciences, University of East Anglia, Norwich, 1999.

[8] J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens and O. Brasseur, “A Neural Network Forecast for Daily Average PM10 Concentrations in Belgium,” *Atmospheric Environment*, Vol. 39, No. 18, 2005, pp. 3279-3289.

[9] J. B Ordieres, E. P. Vergara, R. S. Capuz and R. E. Salazar, “Neural Network Prediction Model for Fine Particulate Matter (PM2.5) on the US-Mexico Border in El Paso (Texas) and Ciudad Juαrez (Chihuahua),” *Environmental Modelling & Software*, Vol. 20, No. 5, 2005, pp. 547-559.

[10] G. Corani, “Air Quality Prediction in Milan: Feed-Forward Neural Networks, Pruned Neural Networks and Lazy Learning,” *Ecological Modelling*, Vol. 185, No. 2-4, 2005, pp. 513-529.

[11] C. Lin and C. Lee, “Neural Fuzzy Systems,” Prentice Hall, Upper Saddle River, 1996.

[12] M. Hagan and M. Menhaj, “Training Feed-Forward Networks with the Marquardt Algorithm”, *IEEE Transactions on Neural Networks*, Vol. 5, 1996, pp. 989-993.

[13] T. Chernichow, A. Piras, K. Imhof, P. Caire, Y. Jaccard, B. Dorizzi and A. Germond, “Short Term Electric Load

Forecasting with Artificial Neural Networks," *Engine Intelligent Systems*, Vol. 2**,** 1996, pp. 85-99.

[14] J. D Farmer and J. J. Sidorowich, "Predicting Chaotic Dynamics, Dynamic Patterns in Complex Systems," In: J. A. S. Kelso, A. J. Mandell and M. F. Shlesinger, Ed., *World Scientific*, 1988, pp. 265-292.

[15] Y. Y. Hong and C. Y. Hsiao, "Locational Marginal Price Forecasting in Deregulated Electricity Markets Using Artificial Intelligence," *IEE Proceedings of Generation Transmission Distribution*, Vol. 149, No. 5, 2002, pp. 621-626.

[16] J. Mitchell and S. Abe, "Fuzzy Clustering Networks: Design Criteria for Approximation and Prediction," *IEICE Transactions on Information and Systems*, Vol. E79D, No. 1, 1996, pp. 63-71.

[17] A. B. Geva, "Hierarchical-Fuzzy Clustering of Temporal-Patterns and its Application for Time-Series Prediction," *Pattern Recognition Letters*, Vol. 20, No. 14, 1999, pp. 1519-1532.

[18] M. Djukanovic, B. Babic, O. J. Sobajic and Y. H. Pao, "24-hour Load Forecasting," *IEE Proceedings – C*, Vol. 140, 1993, pp. 311-318.

[19] J. B. McQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkley Symposium on Mathematical Statistics and Probability*, Berkeley, 27 December 1965-7 January 1966, pp. 281-297.

[20] D. J. Kim, Y. W. Park and D. J. Park, "A Novel Validity Index for Determination of the Optimal Number of Clusters," *IEICE Transactions on Information and Systems*, Vol. E84-D, No. 2, 2001, pp. 281-285.