Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 86 (2016) 277 – 280

Procedia Computer Science

2016 International Electrical Engineering Congress, iEECON2016, 2-4 March 2016, Chiang Mai, Thailand

Improvement on PM-10 forecast by using hybrid ARIMAX and Neural Networks Model for the summer season in Chiang Mai

Rati Wongsathan[a]*, Supawat Chankham[b]

[a] *Electrical Engineering Department, North-Chiangmai University HangDong Chiang Mai Thailand 50230*
[b] *Electrical Engineering Department, North-Chiangmai University HangDong Chiang Mai Thailand 50230*

Abstract

Since the air monitoring stations do not provide the relation between other toxic gas and meteorological parameters with the particulate matter up to 10 micrometer or PM-10.The influence of meteorological as well as correlation with other toxic gas is investigated and used them to forecast PM-10 in the case of Chiang Mai province of Thailand. In this paper an attempt to develop hybrid models of an Autoregressive Integrated Moving Average (ARIMA) model with other exogenous variables (ARIMAX) and Neural Networks (NNs), the two hybrid models, i.e. hybrid ARIMAX-NNs model and hybrid NNs-ARIMAX model were implemented to forecast PM-10 for highly season during January-April of Chiang Mai Province. Simulation results of hybrid model are compared with the results of ARIMA, ARIMAX and NNs model. The experimental results demonstrated that the hybrid NNs-ARIMAX model outperformed best over the hybrid ARIMAX-NNs model, ARIMAX model, NNs model, and ARIMA model respectively. In this case study and maybe other cases, it has proved that the NNs model should be priori captured and filtered the non-stationary non-linear component while the fully linearly stationary residuals were accurately predicted by ARIMAX model later.

Peer-review under responsibility of the Organizing Committee of iEECON2016

Keywords: PM-10; Hybrid ARIMAX and Neural Networks.

1. Introduction

Chiang Mai, the largest city in the north of Thailand, severely experienced with the pollution related to PM-10 for a decade. PM-10 start climbing to the dangerous level especially between January to April, dry-season aridity and rising temperatures coincide with forest fire, wood and agricultural burning. By comparison, PM-10 in Bankok

* Corresponding author. Tel.: +66(81)2893400; fax: +66(53)819998.
*E-mail address:*rati1003@gmail.com

Peer-review under responsibility of the Organizing Committee of iEECON2016
doi:10.1016/j.procs.2016.05.062

has only 40-50 during this same period. PM-10's sources go beyond farming, forest and grass burning, e.g only Mae Chaem one of the district of Chiang Mai alone produces and burns over 37,000 tons of corncob waste every year [1]. For our assumption that PM-10 is nonlinear and complex model, the capture of advantage both ARIMA and NNs model is alternate selected and implemented which gives the forecast result better than any single model. However, the residual usually severe occurs in the high season period. To solve this problem, other exogenous variables which relate to PM-10 are considered to include in an ARIMA model and is referred to an ARIMAX model.

2. Methodology and Methods

In this work, special emphasis is focused for high season period with highly disturbance by various factors and implemented the hybrid model on PM-10 forecast. The data are collected from year 2011-2013 for both PM-10 and exogenous data which are 4 toxic gas variables i.e. *CO*, O_3, NO_2 and SO_2 and 4 meteorological variables i.e. gust wind (*GW*), temperature (*T*), pressure (*P*), and relative humidity (*H*) [2]-[3]. The data in year 2011-2012 was used for model training and the data in year 2012-2013 was used to test the performance of the model.

2.1. ARIMAX model

It assumes that input sequences are expressed by $\{X_{1t}\},\ldots,\{X_{Kt}\}$, dependence sequence is represented by $\{Y_t\}$, and ARIMA(p,d,q)X(K) model can be described as following,

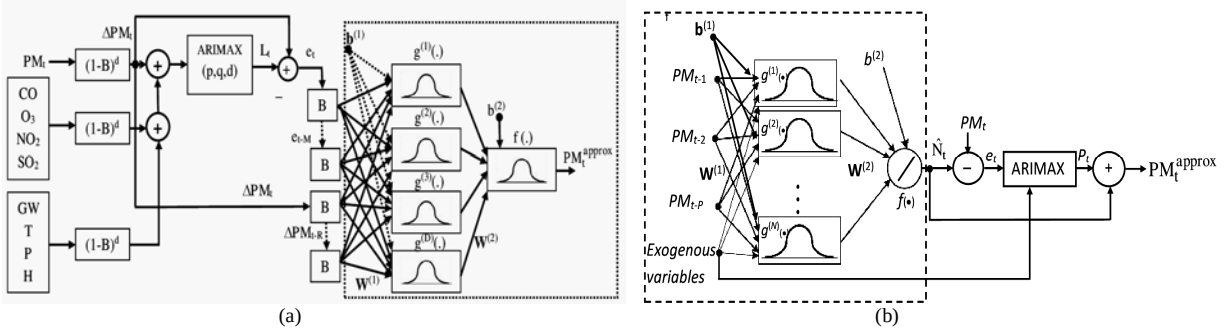
$$\Delta^d Y_t = \delta + \sum_{i=1}^{K} \mu_i X_{t-i} + \sum_{i=1}^{P} \varphi_i Y_{t-i} + \sum_{i=0}^{Q} \theta_i \varepsilon_{t-i} \quad . \tag{1}$$

Where μ_i, φ_i, and θ_i denote the coefficient parameters, *K*, *P*, and *Q* denote the maximum time lag related to the input sequences, dependence sequence and residuals respectively, and δ is a constant.

According to the [4]-[5], the basically procedure step uses to set up ARIMAX model is similar to ARIMA model. Stationary test on both $\{X_{it}\}$ and $\{Y_t\}$ is preliminary examined by ACF and PACF plot or unit root test by augmented Dickey-Fuller (ADF) test. The first differencing is applied to the time series data for non-stationary case. After series are identified as the stationary, the best fit parameters of an ARIMA model were estimated according to its order *p* and *q* by PACF and ACF plot considering. The MLR is fitted the model and the insignificance variable is eliminated by notation of P-value statistics. Diagnostic checking is used to examine at the last step by the several statistics assumption of the residuals such as Chi-Square test or the correlation of the residual plot.

2.2 Hybrid ARIMAX-NNs model

It may be assume to consider PM-10 times series to be composed of a linear autocorrelation structure (L_t) and a nonlinear component (N_t) as, $Y_t = L_t + N_t + e_t$, where e_t is the residuals at time *t*. In the proposed model, there are mainly two stages which has illustrated in the diagram of Fig. 1(a). In the first stage ARIMAX operates to forecast by using the *K* historical data, *Q* past error values, and the exogenous variables. The residual is then generated and provided to NNs which used this error altogether with historical of ΔPM data for the final PM-10 forecast. In the design experiment of generalized NNs, several factors including number of input node (in this case is the time lag length) and number hidden node are properly selected for accuracy and rapidly convergence of solution.



(a) (b)

Fig. 1. The struture of (a) hybrid ARIMAX-NNs model, and (b) hybrid of NNs-ARIMAX modle

2.3 Hybrid NNs-ARIMAX model

Model of the hybrid NNs-ARIMAX model is shown in Fig. 1(b), the designed structure NNs with *P* input node, *N* hidden node and one output node of NNs model was adopted to forecast solution N_t in the first stage. The