International Journal of Computer Techniques -– Volume 9 Issue 2, Mar 2022

RESEARCH ARTICLE OPEN ACCESS

TIME SERIES ANALYSIS AND FORECASTING OF AIR POLLUTION PARTICULATE MATTER USING SARIMA AND SVM APPROACH

Mr.C.Mani M.C.A.,M.Phil.,M.E., [1] ,S.Arunkumar[2]

[1]Associate Professor, Department of Computer Science and Engineering, Nandha Engineering College (Autonomus),Erode,Tamilnadu,India.

[2]Final MCA,Department of Computer Application,Nandha Engineering College (Autonomus),Erode,Tamilnadu,India.

Email: [1]cmanimca@gmail.com,[2]arunkumarmca2000@gmail.com

Abstract: Air pollution is one of the major environmental challenges in a smart megacity terrain. Real- time monitoring of pollution data enables the metropolitans to dissect the current business situation of the megacity and take their opinions consequently. Deployment of Big data Analytical Tool grounded detectors has vastly changed the dynamics of prognosticating air quality. The quality of the atmospheric terrain is an important condition for the long-term survival of humans on earth. A clean suitable atmospheric terrain is needed for the healthy development of mortal beings. Current development of country's frugality, transportation and assiduity with the enhancement of urbanization, environmental pollution problems have gradationally come prominent, but this is contrary to people's vision of pursuing a high- quality life. Now the problem of haze, photochemical problems in the air, and global warming is formerly a crucial issue of global concern. Being exploration has used different machine literacy tools for pollution vaticination; still, relative analysis of these ways is frequently needed to have a better understanding of their processing time for multiple datasets. This design performed the pollution vaticination using sarima retrogression fashion and also SVM bracket approach with a relative study to dissect the stylish model for directly prognosticating the air quality. SVM bracket is performed for pollution estimation using multiple available data sets. The design is designed using R Studio. The rendering language used is R3.4.4.

Keywords: Data Mining, Air Pollution, Time Series Analysis, Sarima Model.

I. INTRODUCTION

The quality of the atmospheric terrain is an important condition for the long- term survival of humans on earth. A clean suitable atmospheric terrain is needed for the healthy development of mortal beings (1). Still, with the fast development of country's frugality and assiduity and nonstop increase in position of urbanization, air is seriously defiled. Increased air pollution affects the people physical health, and increases threat of respiratory infections, heart complaint, and lung cancer.

Because of frequent environmental pollution accidents and severe bank pollution incidents across the country, government and public are veritably concerned about this air pollution. 2016 WHO report said that seven million people worldwide die each time as the exposure result to medium (out-of-door inner) air pollution.

Utmost susceptible to air pollution are people who are senior, veritably youthful, with pre-existing respiratory conditions or low socioeconomic status. Particulate matter (PM), ozone (O3), nitrogen dioxide (NO2), and sulfur dioxide (SO2) are the adulterants with the strongest substantiation of health goods.

Particulate matter is one of the atmospheric adulterants caused by burning coal as an energy source or from bus-mobile exhaust. Due to the appearance of PM, the conformation of different chemical composition, flyspeck size distribution, and other physical and chemical parcels are veritably different in the atmosphere, and their behavioral impact causing different natural goods on mortal health.

Thus, it's important important for every country and metropolitan metropolises to keep control and cover the PM2.5 in atmosphere to maintain good health and safe terrain. This study proposes an approach for vaticination of unborn soothsaying of air quality of metropolitan megacity. According to the Air Quality Life Index (AQLI), Pakistan was ranked the 5th most weakened country in the world in 2016. The National Environmental Quality Norms for Ambient Air cover sulfur dioxide (SO2), nitrogen dioxide

(NO2), ozone (O3), suspended particulate matter, particulate matter (PM10 and PM2.5), lead (Pb), and carbon monoxide (CO). In 2017 the Smog Commission was established to probe the reasons for the poor air quality across Lahore and to formulate plans to ameliorate it.

Lahore, Faisalabad, and Gujranwala are the metropolises most affected by gauze in Pakistan with high situations of air pollution. In November 2019 the Punjab education department closed public and private seminaries due to poor air quality. It also banned children from any out-of-door conditioning until the end of December and needed children to wear antihaze masks during class hours. In November 2019, three teenagers sought legal action against the government of Punjab, for the "violation of their abecedarian right to a clean and healthy terrain" demanding critical action be taken.

Another study shows that children living and attending academy in a veritably high PM2.5 region had a significantly advanced blood pressure compared to children with lower exposure. Study shows that business- related civic pollution can contribute to their after threat of hypertension and cardiovascular complications, indeed in children, if that increased blood pressure persists. It finds that the advanced PM2.5 with advanced gauze leads towards the adverse health impact. The study reviews the current situation of Lahore and concluded that the current situation is likely to deteriorate due to the lack of an applicable action plan on the part of the government and the incapability of the authorities concerned to take note of the inflexibility of the situation.

In this paper we characterize the air adulterants in the ambient air of Lahore, Pakistan; examine periodic variations; use backward air mass circles to identify pollution sources; perform a correlation and retrogression analysis of the measured adulterants ((CO), (NO), (SO2), (O3)) with particulate matter; and compare these results with the source of product of .those patches. This study gives the complete analysis of relationship of adulterants as well as there source of generations by line styles. Eventually, we use a time series model for the vaticination of particulate matter attention.

II. LITERATURE REVIEW

In this paper (1) the authors stated that The broad end of this study is to estimate empirically the impacts of urbanization along with some other explicatory variables on environmental declination measured by carbon dioxide emigrations for four countries from the South Asian Association for Regional Cooperation (SAARC) region videlicet India, Bangladesh, Sri Lanka, and Pakistan.

Periodic time series data over the period of 1982 – 2013 are used. After, employing colorful applicable and required statistical tests, the system of least places has been employed as an logical fashion for parameters estimation. The least places estimate reveals that the impact of urbanization on the terrain plant is blended. In cases of Bangladesh and India, the relationship between urbanization growth and terrain plant is significantly negative, while, the impact of urbanization on terrain is significantly positive in case of Sri Lanka and insignificantly positive for Pakistan during the period under

the study. The findings of the study suggest that policy makers need to formulate applicable policy for long term civic planning which can clearly help to alleviate largely CO2 emigrations/ environmental pollution.

The term urbanization refers to a process where large-scale labor force is going from an agribusiness-grounded frugality to civic- grounded artificial frugality. This kind of metamorphosis is a trend of profitable and social development. It seems that, at the original stage of profitable development process, the environmental declination is likely to be low. Most probably, industrialization, modernization, and rapid-fire increase of urbanization contribute to environmental pollution which is a thoughtful environmental concern across the world, but more dangerous for developing nations. While with effective development, a high- contaminating artificial frugality turns into technology- grounded frugality and as such the degree of ecological pollution is condensed (11).

Global warming is a hot issue for experimenters as it's a crucial global concern. It has been noted that enhanced position of carbon dioxide (CO2) emigrations is one of the major causes of planetary heating. The artificial sector is seen also responsible for pollution as compared to the services sector (12). In the same way, energy is one of the factors of product and plays pivotal part in the profitable development and growth process. Indeed, energy is needed for massive occupants abiding in the civic areas as well as the artificial sector, where some limited consumption of energy causes environmental declination.

The study of Peng and Bao (13) shows that vast consumption of natural coffers creates further emigrations of artificial adulterants. Eventually, the artificial growth leads to environmental quality neglect. The resource-ferocious growth together with rapid-fire urbanization growth, the growing consumption of a rising middle class and growing product patterns have robust negative consequences in saddening profitable development prospects for the countries (14).

Mortal- convinced environmental neglect is the utmost concern and a multifaceted global issue facing the macrocosm as a whole. Environmental declination is projected to have substantial goods on natural resource systems, and in this manner changes in the natural terrain can impact mortal aliment as well as profitable conditioning. The present study focuses substantially on probing the impact of urbanization on environmental declination in the environment of India, Bangladesh, Sri Lanka, and Pakistan.

The harmonious growth of urbanization, profitable development and the terrain is a pivotal field of exploration that seems to syndicate the social and natural lores. In a

study, Li and Ma (15) note that in proposition, urbanization growth, profitable development and the terrain are connected by a sequence of positive and negative impacts. Generally, in numerous countries along with urbanization, rapid-fire profitable growth occurs where movement of populations from pastoral areas to metropolises and municipalities have been observed.

Also, the growing urbanization affects the terrain by shifting the situations of polluting emigrations as a result of the change in product and shifts in the population's patterns after shifting from pastoral areas to civic areas.

On the other hand, urbanization has also created abundant environmental problems ranging from the domestic to the global scale (16, 17), comprising enlarged air and water pollution and reduced water force (18, 19), lacking casing and sanitation amenities and business overcrowding (20). The studies of Bloom etal. (21) and Glaeser (22) expound that urbanization accelerates the range of complaint and worsens ills similar as crime, poverty and dangerous environmental quality.

The central ideal of the present study is to examine the impacts of urbanization along with some other named explicatory variables, videlicet pastoralist land, technology, influx, and total population on environmental declination measured by CO2 emigrations in four countries from the SAARC region videlicet India, Bangladesh, Sri Lanka, and Pakistan. For empirical inquiry, periodic time series data ranging from 1982 to 2013 are used. This study employed all possible individual test settings including descriptive analysis, periodical correlation analysis, white general hetero scedasticity test and D.W. test, which reveal that there's no serious problem of diversity and periodical correlation in the model of environmental declination used for the named four SAARC countries.

Data have been checked for stationarity using the ADF test. The Johanson cointegration results indicates that there exists four cointegrating relationship among the variables. The OLS estimate uncovers that the impact of technology measured by energy use is positive on environmental declination and pastoralist land has statistically significant negative impact on terrain of the countries under the study. The results also reveal that the population growth has a positive impact on terrain in cases of Bangladesh and India. The empirical results on urbanization and terrain relationship are blend.

The relationship between urbanization and terrain plant is significantly negative in cases of Bangladesh and India; indicating that in these two countries sustained urbanization seems to be easing in dwindling CO2 emigration and latterly cover terrain. While the impact of urbanization on environmental declination plant is significantly positive in case of Sri Lanka, still, in case of Pakistan the relationship between these two variables are positive but not significant statistically. Therefore, in cases of Sri Lanka and Pakistan, the results inferring that urbanization contributes to environmental declination because due to growing civic population, the operation of structure, energy, and transport upsurges and as a result of tenant shift from agrarian to artificial sector might enlarge pollution of the terrain.

In this paper (2) the authors stated that India is one of the most weakened countries in the world, where several major metropolises are facing serious environmental consequences as a result of rapid-fire pollution growth. The ideal of this exploration is to dissect air pollution trends with respect to colorful geographical locales, in order to have a global view of the damage caused, so that applicable conduct can be developed in the future to help air pollution.

In this regard, the weakened database was established grounded on the data handed by the Central Pollution Control Board; Ministry of Environment, Forest, and Climate Change (India). These data demonstrate the periodic growth of SO2, NOx, and particulate matter (PM)2.5 from 2015 to 2018 and were recorded at colorful monitoring stations in three metropolises, videlicet, Delhi, Bengaluru, and Chennai.

The results show that SO2, NOx, and PM2.5 were from different transport modes, both small and large-scale power generations (from diesel, coal and gas factory), diligence, constructions, and domestic cuisine. Overall, there was an adding trend, day by day, in India. The result distributed the considered areas into the following four classes critically weakened (CP), largely weakened (HP), relatively weakened (MP), and low weakened (LP). The results will help in the assessment of pollution for the metropolises delved in this exploration

Air pollution is a serious problem that affects the lives of billions of people every time (Louati etal. 2018; Son and Louati 2016). According to the World Health Organization (WHO), further than 25 of deaths around the world may be directly linked to pollution (Amal etal. 2018).

The global complaint assessment stated that in 1990 and 2016,0.99 million and0.78 million unseasonable deaths were due to particulate matter (PM)2.5 pollution, independently (GBD MAPS Working Group 2018). In Asia, the loftiest pollution situations recorded were in 2015, when 35 of deaths were due to air pollution. Pollution is especially serious in big metropolises, which are facing numerous difficulties in balancing air quality and the terrain.

From South Korea to Thailand and India, defiled air is harming the largest metropolises. Air pollution is also big issues of concern for China, as it's a major source of dust patches, and it's full of manufactories spread across China that regularly emit thick bank. Although numerous programs and laws have been proposed to reduce pollution situations, air pollution is still a big problem that needs further sweats in order to have safe air quality situations.

In India, air pollution is a more severe problem. Moment, India incorporates 640 sections, and out of this,

27 exceeded the periodic standard value of 40 µg/ m3 in 1998, and 45 exceeded it in 2010 (Guttikunda etal. 2019). Therefore, of the top 10 most weakened metropolises in the world list released by the World Health Organization, nine metropolises are from India, and among them, Delhi remains at sixth position (Yuda 2019). The WHO database further mentioned that Delhi is the most weakened megacity for PM 10 (Donkelaar etal. 2016; WHO 2018). Therefore, this situation is a complicated issue, which explosively influences mortal health, and a demanding result is required in order to maintain the air quality position.

Monitoring air quality plays a vital part in controlling the situations of pollution (Camastra etal. 2019). Therefore, it's veritably important to quantify the air quality situations by different locales in order to see the effect of air pollution on mortal health. Schwela (2012) stated that an air pollution chart could be veritably useful for managing air quality and its goods. Still, air quality operation stations (AQM) aren't always available in developing metropolises because of their set-up costs. Nonetheless, in order to determine air pollution goods directly, the result is to increase number of the AQM stations, to allow for wider and further comprehensive content.

The air quality indicator (AQI) is used for the effective assessment of air quality. This process transforms the data of colorful adulterants into a single number or value. AQI is distributed into the following six different orders good, satisfactory, relatively defiled, poor, veritably poor, and severe. The AQsub-index was developed for tracking eight adulterants (PM 2.5, PM 10, SO2, NO2, CO, NH3, O3, and Pb). The computation of the AQsub-index is grounded on the ambient attention of air adulterants, which is a direct function of the attention (e.g., thesub-index for PM2.5 is 75 at a attention of 45 µg/ m3, 51 at a attention of 31 µg/ m3, and 100 at a attention of 60 µg/ m3). In Table 2, the AQI orders for SO2, NOx, and PM2.5 are given (CPCB 2016).

The government of India has started a monitoring program named the National Air Quality Monitoring Program (NAMP) to regularly cover air quality. As per the report of September 2018, four measure adulterants, SO2, NO, PM2.5, and PM 10, can be covered regularly using NAMP. There are 703 air quality stations across 307 metropolises in India. This program is fully controlled by the Central Pollution Control Board in collaboration with state pollution control board (Pant etal. 2018). Substantially, the air quality stations are stationed in civic areas (Donkelaar etal. 2016; Gordon etal. 2018; Guttikunda etal. 2019; WHO 2018).

Table 2.1. Examples of studies using HYSPLIT for transport and dispersion calculations

Application	Location	Brief description	Reference (s)
Radionuclides	Marshall Islands	Deposition of fallout	Moroz et al. (2010)
	(central Pacific), Nevada Test Site (United States), Semipala-tinsk Nuclear Test Site (Kazakhstan)	from atmospheric nuclear tests	
	AREVA NC La Hague nuclear processing plant (northwestern France)	Krypton-85 air concentra-tions	Connan et al. (2013)
	80-km range around Fu-kushima reactor (Japan)	Temporal behavior of plume trajectory, con-centration, deposition, and radiation dosage of cesium-137	Challa et al. (2012)
	Fukushima and adjacent prefectures (Japan)	Air parcel transport and dispersion to interpret io-dine, tellurium, and cesium measurements	Kinoshita et al. (2011)
Wildfire smoke	CONUS	U.S. National Weather Service Smoke Forecasting System	Rolph et al. (2009)

III. PROPOSED METHODOLOGY

The main purpose of feature selection process is to find the minimum number of relevant attributes from a given Air pollution dataset described by their attributes

(features). Most real life problems require an optimal as well as acceptable solution than calculate them precisely at degraded performance cost, time and space complexities. Therefore, it is must to carry out this analysis using the sarima model. Sarima model includes a full set of methods for the time series analysis, prediction, as well as control. The data set is taken for past 2 years and next 5 years prediction is carried out. The drawbacks are:

- Classification is not being carried out in past year data set.
- Only future year pollution is found out.
- Heavy or low pollution classification is not being carried out.
- Not suitable when dataset grows.

In proposed system, all the existing methodology is carried out. In proposed system also, data set is taken from meteorological sites of India and then saved in excel files as 'comma separated values' files. The data contains air pollution data in addition with humidity information. These data are taken for text pre-processing first and then converted into time series data set format and future year prediction is carried out.

In addition, KNN and SVM are applied for classification and so it is found to be suitable especially if the data set is having more number of records contains outlier data. A wide variety of ground water level records can be taken for classification purpose and predicting a new model at the same time increasing the efficiency. Moreover, naïve bayes classification is also carried out. Probability of datewise CO level is also calculated. The advantages are:

- The SVM model yields better accuracy only if the test record is exactly matched with any of the training data records.
- SVM is applied so could be preferred when the data set grows larger.
- SVM is applied so could be preferred when the outlier data is more.
- Conditional Probability helps in percent wise CO for all given dates.

IV. FINDINGS

Figure 4.1 AIR QUALITY DATA SET SOME RECORDS

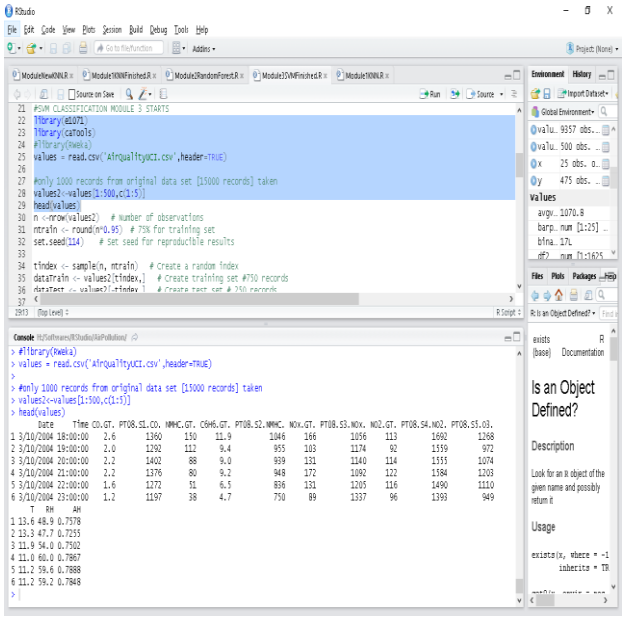


Figure 4.2 SUMMARY OF SUPPORT VECTORS

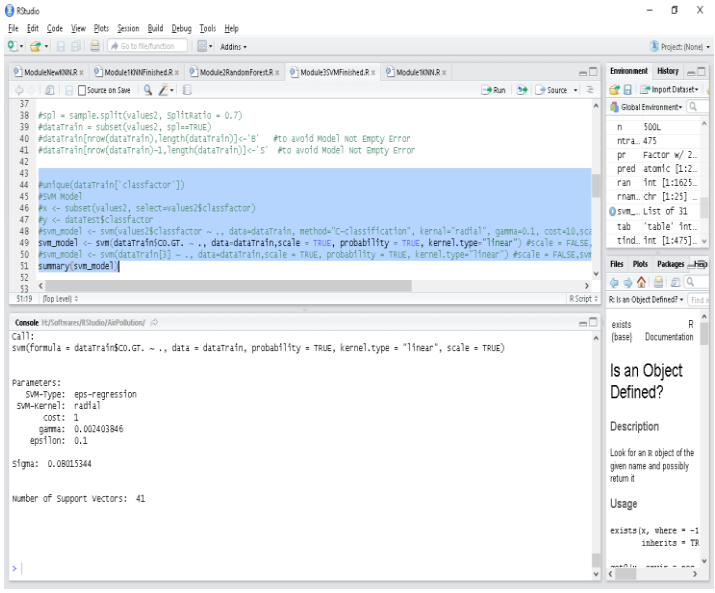
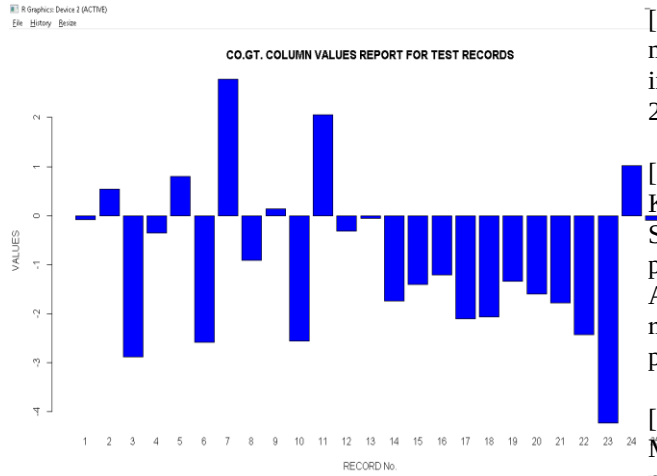


Figure 4.3 SVM PROBABILITY FOR RECORDS [POSITVE VALUE RECORDS ARE DIFFERENT FROM NEGATIVE VALUE RECORDS]



V. CONCLUSION

The main exploration work of this study is to first dissect the current status of Lahore air quality, and conclude that the main adulterants in the atmosphere of various cities, PM10, PM2.5, SO2, CO, and NO2 from 2014 to 2019 in the air are changing and there's a strong correlation between them. On this base, the impact of several applicable factors, similar as geographic position, meteorological conditions, and mortal conditioning (coal burning, transportation, dust, population, and programs) on the air quality of various cities are described. The proposed fashion is producing an enhanced conception over the Air pollution within new data mining ways; Sarima model and SVM bracket. The SVM bracket finds out the air pollution position during the process. The proposed fashion SVM is producing further than 90 delicacy in bracket, while sarima model gives below 90 delicacy during the process. In unborn work, the study can be enhanced further to get better result over the air pollution by adding the criteria and pollution suggestion within different types of attributes and also give the suggestion in attainability of Air pollution specialist or experts.

REFERENCES

[1] M. Azam and A. Q. Khan, ‘‘Urbanization and environmental degradation: Evidence from four SAARC countries-Bangladesh, India, Pakistan, and Sri Lanka,’’ Environ. Prog. Sustain. Energy, vol. 35, no. 3, pp. 823–832, May 2016, doi: 10.1002/ep.12282.

[2] R. Sharma, R. Kumar, D. K. Sharma, L. H. Son, I. Priyadarshini, B. T. Pham, D. Tien Bui, and S. Rai, ‘‘Inferring air pollution from air quality index by different geographical areas: Case study in India,’’ Air Qual., Atmos. Health, vol. 12, no. 11, pp. 1347–1357, Nov. 2019, doi: 10.1007/s11869-019-00749-x.

[3] A. F. Stein, R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, ‘‘NOAA’s HYSPLIT atmospheric transport and dispersion modeling system,’’ Bull. Amer. Meteorol. Soc., vol. 96, no. 12, pp. 2059–2077, 2015, doi: 10.1175/bams-d-14-00110.1

[4] K. Kalpakis, D. Gada, and V. Puttagunta, ‘‘Distance measures for effective clustering of ARIMA time-series,’’ in Proc. IEEE Int. Conf. Data Mining, Nov./Dec. 2001, pp. 273–280, doi: 10.1109/icdm.2001.989529.

[5] A. Malik, A. Kumar, S. Kim, M. H. Kashani, V. Karimi, A. Sharafati, M. A. Ghorbani, N. Al-Ansari, S. Q. Salih, Z. M. Yaseen, and K.-W. Chau, ‘‘Modeling monthly pan evaporation process over the Indian central Himalayas: Application of multiple learning artificial intelligence model,’’ Eng. Appl. Comput. Fluid Mech., vol. 14, no. 1, pp. 323–338, Jan. 2020.

[6] Ghorbani, M. A., Deo, R. C., Yaseen, Z. M., Kashani, M. H., & Mohammadi, B. (2017). Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. Theoretical and Applied Climatology, 133, 1119–1131. doi:10.1007/s00704-017-2244-0

[7] Doorenbos, J., & Pruitt, W. O. (1977). Guidelines for predicting crop water requirements (Irrigation and Drainage Paper No. 24, FAO). doi:10.2514/6.2014-2117.

[8] Griffiths, J. F. (1966). Another evaporation formula. Agricultural Meteorology, 3(3–4), 257–261.

[9] Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of the surface heat flux and evaporation using large-scale parameters. Monthly Weather Review, 100, 81–92

[10] Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology, 528, 312–320. doi:10.1016/j.jhydrol.2015.06.052

[11] Grossman, G., & Krueger, A. (1995). Economic growth and the environment, Quarterly Journal of Economic, 110, 353–372.

[12] Neumayer, E. (2003). Are left-wing party strength and corporatism good for the environment? Evidence from panel analysis of air pollution in OECD countries, Ecological Economics, 45, 203–220.

[13] Peng, S., & Bao, Q. (2006). Economic growth and environmental pollution: An empirical test for the environmental Kuznets Curve hypothesis in China, Research on Financial and Economic Issues, 3–17.

[14] United Nations Economic and Social Commission for the Asia Pacific (UNESCAP). (2008). Greening Growth in Asia and the Pacific, Bangkok: UNESCAP.

[15] Li, S., & Ma, Y. (2014). Urbanization, economic development and environmental change, Sustainability, 6, 5143–5161.