RESEARCH ARTICLE OPEN ACCESS

PREDICTION OF ENVIRONMENTAL POLLUTION USING NEURAL NETWORK

Ms. N. Zahira Jahan, M.C.A., M.Phil.,[1], R.Dhanalakshmi[2],

[1]Associate Professor, Department of Computer Applications,

Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

[2]Final MCA, Department of Computer Applications,

Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

Email : [1]zahirajahan1977@gmail.com, [2]pragalyaanbu@gmail.com

Abstract. In general environmental pollution is one of the most considerable one as challenges in a smart megacity terrain. In denitrifying 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.

This project performed the pollution prediction using sarima regression technique and also classification using neural network approach with a comparative study to analyze the best model for accurately predicting the air quality with reference to data size and processing time. Current research has used different machine learning tools for pollution prediction; however, comparative analysis of these techniques is often required to have a better understanding of their processing time for multiple datasets.

In analyzing the standard of atmospheric terrain is an important condition for the much longer survival of living people on environment. A pure suitable atmospheric terrain is needed for the healthy development of mortal beings. Now the problem of haze, photochemical problems in the air, and global warming is formerly a crucial issue of global concern. Current implementations 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. The project is designed using R Studio. The coding language used is R 3.4.4.

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

I. INTRODUCTION

Air quality evaluation is an important way to monitor and control air pollution. The characteristics ofair supply affects its suitability for a specific use.A few air pollutants, called criteria air pollutants, are common throughout the United States. These pollutants can injure health, harm the environment and cause property damage. The current criteria pollutants are:

- Carbon Monoxide (CO)
- Lead (Pb)
- Nitrogen Dioxide (NO_2)
- Ozone (O_3)
- Particulate matter (PM)
- Sulfur Dioxide (SO_2).

The Air Quality System (AQS) contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from over thousands of monitors. AQS also contains meteorological data, descriptive information about each monitoring station (including its geographic location and its operator), and data quality assurance/quality control information. AQS data is used to assess air quality, assist in Attainment/Non-Attainment designations, evaluate State Implementation Plans for Non-Attainment Areas, perform modeling for permit review analysis, and other air quality management functions. AQS information is also used to prepare reports for Congress as mandated by the Clean Air Act

- **Air Quality Standards**

Office of air quality planning and standards (OAQPS) manages EPA programs to improve air quality in areas where the current quality is unacceptable and to prevent deterioration in areas where the air is relatively free of contamination. To accomplish this task, OAQPS establishes the National Ambient Air Quality Standard (NAAQS) for each of the criteria pollutants. There are two types of standards - primary and secondary.

- **Primary standards**: They protect against adverse health effects;
- **Secondary standards:** They protect against welfare effects, such as damage to farm crops and vegetation and damage to buildings.

Because different pollutants have different effects, the NAAQS standards are also different as shown in Table 2. Some pollutants have standards for both long-term and short-term averaging times. The short-term standards are designed to protect against acute, or short-term, health effects, while the long-term standards were established to protect against chronic health effects.

According to the researchers Elias and Nicholas, modelling of atmospheric pollution phenomena till now has been based mainly on dispersion models that provide approximation of the complex physicochemical processes involved. While the sophistication and complexity of these models has increased over the years, use of these techniques in the frame of real-time atmospheric pollution monitoring seems not suitable in terms of performance, input data requirements and compliance with the time constraints of the problem. Instead, human expert's knowledge has been primarily applied in Air Quality Operational Centers for the real-time decisions required, while mathematical models have been used mostly for off-line studies of the phenomena involved. As per Air pollution phenomena have been measured by using physical reality as the start point. And then, for example, these data traditionally have been coded into differential equations. However, these kinds of techniques have limited accuracy due to their inability to predict extreme events.

Table 11. AQI Classification

AQI	Air Pollution Level
0–50	Excellent
51–100	Good
101–150	Lightly Polluted
151–200	Moderately Polluted
201–300	Heavily Polluted
300+	Severely Polluted

Air pollution important parameter called air quality index (AQI) which quantifies air quality in a region as shown in Table 1.1 It is a number used by government agencies to communicate to the public how polluted the air is currently or how polluted it is forecasted to become. As the AQI increases, an increasingly large percentage of the population is likely to be exposed, and people might experience increasingly severe health effects. Different countries have their own air quality indices, corresponding to different national air quality standards.

II. LITERATURE REVIEW

In this paper 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.

In this paper 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.

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. 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 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.

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 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, comprising enlarged air and water pollution and reduced water force, lacking casing and sanitation aenities and business overcrowding. The studies of Bloom etal., Glaeser 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.

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).

III. PROPOSED METHODOLOGY

Utmost real life problems need an optimal and respectable result rather than calculating them precisely at the cost of degraded performance, time and space complications. Thus, it's necessary to carry out the analysis using sarima model. Sarima model includes a complete set of styles for time series analysis, ratiocination, and control. The data set is taken for 2 times and coming five times vaticinator is carried out. The main objective of attribute selection process is to find the minimal number of applicable attributes from a given input dataset as comma separated values described by their attributes.

In The algorithm used in existing system is carried out, in addition, K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) are applied for classification and so it is found to be faster especially if the data set is having more number of records. Moreover, naïve bayes classification is also carried out. Probability of date wise CO level is also calculated. The input dataset contains air pollution data with humidity information also. 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.

- Even if the data set becomes too larger this can easily predict the required output in short time.
- The proposed ANN model yields better accuracy even if the test record is no exactly matched with any of the training data records.
- CO is identified percent wise using conditional probability.
- If the outlier data is more the proposed method can easily proceed without any restrictions.

IV. FINDINGS

Figure 4.1 AIR QUALITY DATA SET SOME RECORDS

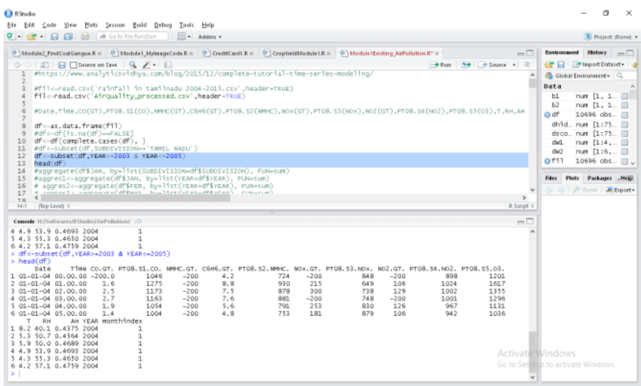


FIGURE 4.2 AIR POLLUTION PREDICTION VALUE FOR GIVEN TIME SERIES DATA

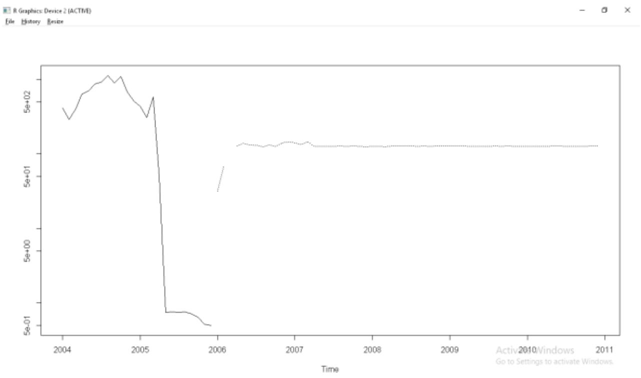


FIGURE 4.3 AIR POLLUTION PREDICTED VALUES

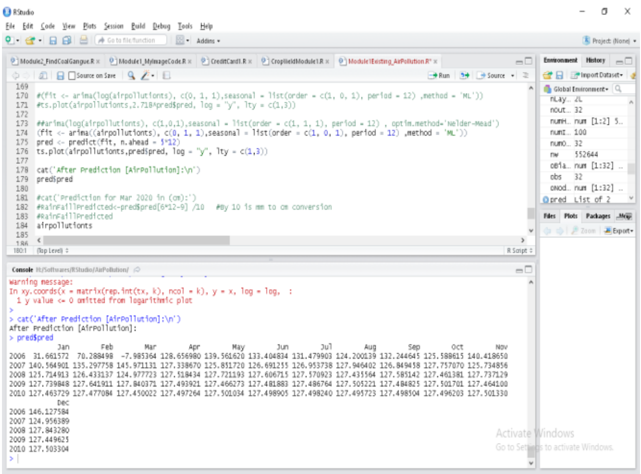


FIGURE 4.4 [POSITVE VALUE RECORDS ARE DIFFERENT FROM NEGATIVE VALUE RECORDS]

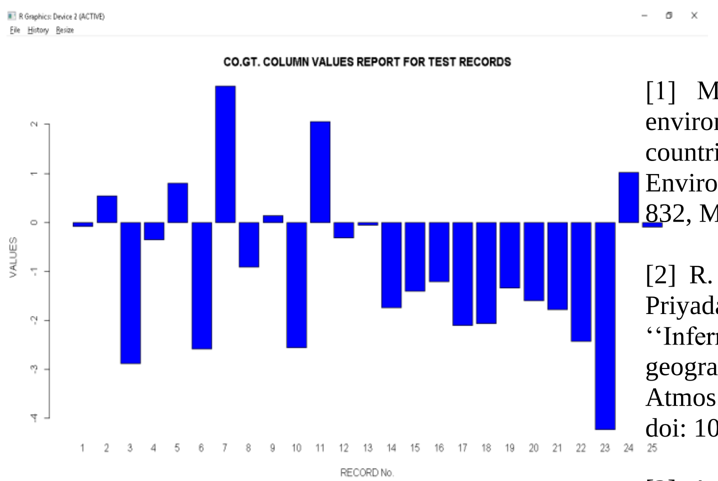
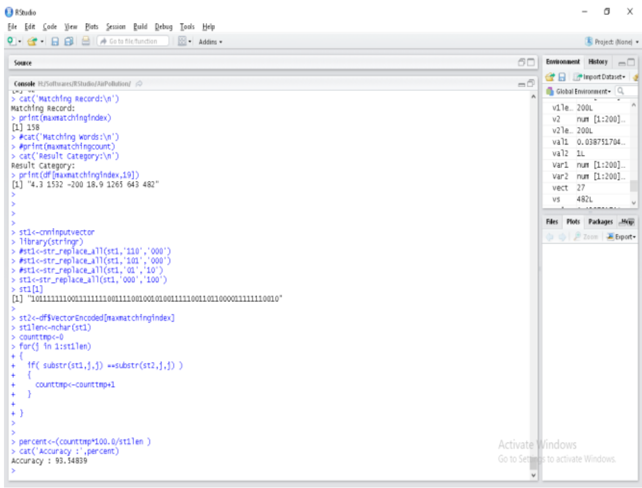


FIGURE 4.5 FIND NEURAL NETWORK BASED TEST DATA'S CATEGORY AND CLASSIFICATION ACCURACY



V. CONCLUSION

Thus the system has tested well. The output accuracy level has increased more than existing system. This new technique is producing an enhanced concept over the environmental pollution prediction within novel data mining techniques; Sarisma model and ANN algorithm. The artificial neural network finds out the air pollution level during prediction process. The proposed technique ANN is producing more than 90% accuracy in classification, while existing model gives below 90% accuracy during prediction. In future work, we will look for the more enhancements to get better result over the Air pollution prediction by increasing the metrics and pollution suggestion within different types of attributes and also provide the suggestion in unavailability 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.