Evaluation of effective factors on air pollution using optimized cellular automata: A case study of Tehran

Alireza Hamoudzadeh[1], Saeed Behzadi[2*]

[1]MSc Student in Geographic Information Systems, Department of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran

[2]Assistant Professor in Surveying Engineering, Department of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran, behzadi.saeed@gmail.com

Abstract:

Vehicles and traffic congestion have been known as the main reasons for air pollution in urban areas, and Cellular Automata (CA) holds a great promise for predicting this hazard. Urban air pollution is a complex phenomenon and many factors involve in its distribution and diffusion. In this paper, the traffic map was used as the source of the air pollutant. Also, the prediction of urban pollution has been done using different data sources such as green space, buildings, wind direction and speed. The coefficient of these factors got estimated with Genetic Algorithm, and a comparison between different modes of the model got done. With considering the effect of these factors an accuracy of 58.4% was obtained.

Keywords: Cellular Automata, Air Pollution, Geographic Information Science (GIS), Spatial Big Open Data

1. Introduction

Urban pollution has always been considered as a great health hazard especially in metropolitan areas and big cities; it contributes significantly to air pollution problems on local, regional, and national scales. One of the main causes of this hazard is the vehicles' exhaust emissions (Westerholm and Egebäck 1994, Kean et al. 2000, Hoekman 1992, Gao et al. 2020) such as lead, particulate matter (PM$_{2.5}$ (PM with an aerodynamic diameter less than 2.5 μm) and PM$_{10}$ with concentrations of particles less than 10 μm in

1

diameter), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NO_x), hydrocarbons, ozone and toxic substances (Chin 1996). Exposure to Particle Matters (PM) increase the chance of cardiovascular diseases (CVDs) (Makri and Stilianakis 2008, Burroughs Peña and Rollins 2017) and infant mortality and reduce the life expectancy (Fotourehchi 2016). It also have a major impact on asthma, chronic obstructive pulmonary disease, and lung cancer (Guo et al. 2016, Kurt, Zhang and Pinkerton 2016) and affects lung development in children (Gauderman et al. 2004).

The diffusion phenomenon is a complex process with dynamic changes in spatiotemporal aspects (Ai et al. 2016). Many approaches can be used to predict air pollution distribution. For instance, Zeinalnezhad et al. (Zeinalnezhad et al. 2020) employed Autoregressive Integrated Moving Average, Logistic Regression techniques, and Adaptive Neuro-Fuzzy Inference System, or Leong et al. (Leong, Kelani and Ahmad 2020) used a Support Vector Machine (SVM) for predicting the Air Pollution Index (API), while a spatiotemporal convolutional long short-term neural network got used to predict the pollution by Wen et al. (Wen et al. 2019).

Besides these methods, Cellular Automata (CA) can be applied to various problem types due to its dynamic behavior and spatial variability (Guariso and Maniezzo 1992, Hamoudzadeh and Behzadi 2021), as it has been used in architecture (Herr and Ford 2016), forest fire growth (Vahidnia et al. 2013), urban planning and spatial growth (Maithani 2010, Iltanen 2012), fluid simulation (Heintz et al. 2017), traffic simulation (Behzadi and Alesheikh 2014), and air pollution prediction. CA (Neumann and Burks 1966) is a discrete model that gets used for modeling and simulating physical systems, phenomena, and spatiotemporal patterns (Deutsch 2013, Cosentino, Vescio and Amato 2013).

CA has 4 characteristics that make it distinct from Agent-Based Models (ABM): firstly, it is consisted of a grid of cells, while in the ABMs, the number of agents varies throughout the program. Secondly, it has a unique grid-based neighborhood that is constant during the process, while the agent's neighbors can change in the ABMs. Thirdly, some rules change the cell status; and finally, in the CA, the rules are based on the neighborhoods, while in ABMs, the rules are based on the interaction of the agents even if they are away.

This method has been used to forecast and model the distribution of air pollution and monitor the Air Quality Index (AQI). Benjavanich et al. used an approach to forecast the pollutants around an urban area,

given recent historic sensor streams using cellular automata which parameters are learned and adapted online by an evolutionary algorithm. With a similar approach, Huynh et al. (Huynh et al. 2020) simulated and analyzed the atmospheric quality using cellular automata with rules of gravity, diffusion, and wind. Xiao-ping (Xiao-ping et al. 2013) et al. used a conventional diffusion algorithm based on Cellular Automata (CA), that considers water flow. Yan-feng et al. (Yan-feng et al. 2020) studied Particulate Matter (PM) emission of some typical cellular automata Velocity-Dependent-Randomization (VDR) model, and TT model with slow-to-start rules under periodic condition and open boundary conditions. In a similar research, Salcido and Carreon-Sierra (Salcido and Carreón-Sierra 2017) used a Fukui-Ishibashi and Nagel-Schreckenberg Traffic Cellular Automata to estimate the space-time distribution of the air pollutant emission rates.

Meteorological elements such as wind are another important factor that impacts air pollution, which affects the direction of the pollution distribution. The wind has two different components: speed and direction, which affect transportation (Kim et al. 2015) and concentration (Padmanabhamurty 1975) of the air pollutants.

Greenspace, parks, and urban forests affect the pollution and noises absorption emitted from the vehicles (Oh 2012, Jo 2002, Yang et al. 2005, Liu and Shen 2014). To estimate the impact of these elements, the Genetic Algorithm (GA), a meta-heuristic optimization method, got used.

Genetic Algorithm (GA) (Goldberg 1989) is a randomized search algorithm that does the optimization with primary chromosomes using three different stages of reproduction, crossover, and mutation, where reproduction is a process based on the objective function (fitness function) of each string. This objective function identifies how good a string is (Roetzel, Luo and Chen 2020). Crossover is defined as the mating of the last population, and mutation is random changes with a small probability in the value of the string position.

Vehicular exhaust emissions are known as one of the main reasons for air pollution in urban regions. This paper aims to predict urban air pollution dynamically using Cellular Automata with the effect of wind, green spaces, and buildings. In Section 2 the data source, case study, methods and processes, pollution absorption, and distribution will be elaborated. In the third section, the result will be explained, and a comparison between the real pollution values and the predicted ones will be done.

2. **Data**

Sample field: In this research, Tehran's metropolitan area (figure 1) got used. Tehran is located at latitude 35°43′00.12″ North longitude 51°24′00.00″ East and consisted of 22 regions with 60 main expressways always facing heavy traffic congestions.

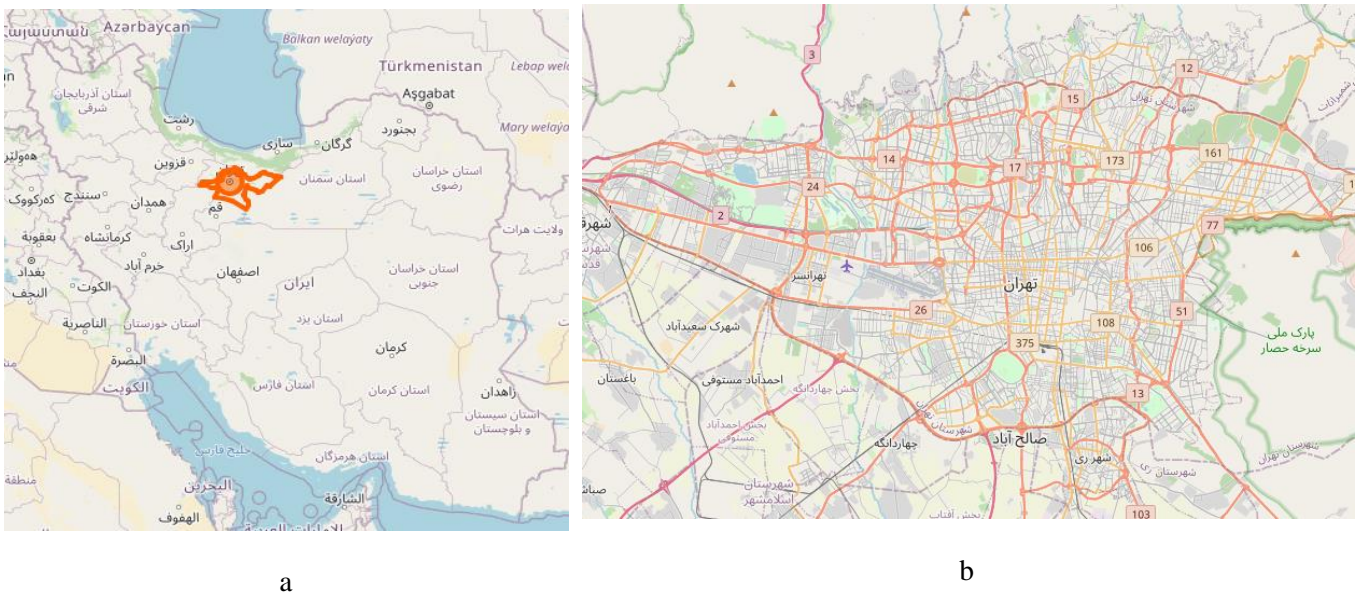a   b

Figure 1- a) Position of Tehran in Iran- b) Tehran (from Openstreetmap.org)

Traffic monitoring: Real-time traffic volume (more than 3,000,000 observations) of 24hours from 10 AM of August 10th to 11th got extracted from Neshan, a Persian navigation app and traffic Application Programming Interface (API), and transformed into raster form and used as the main source of traffic distribution (figure 2).
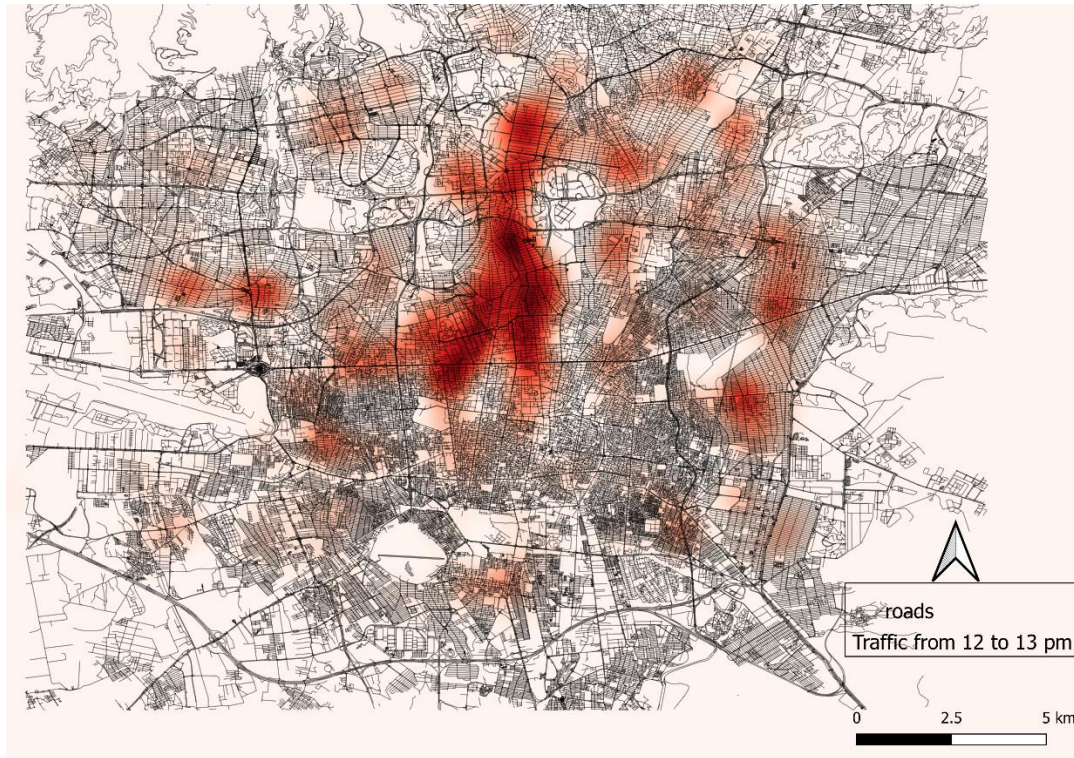
Figure 2- Traffic heat-map of 12 to 1 pm

Wind and pollution measurements: Wind and pollution data (PM_{10} and $PM_{2.5}$) got captured from the Tehran Air Quality Control Company (AQCC) with 21 active stations every 3 hours.

Green area and building map: an OpenStreetMap (OSM) shapefile got used for this factor. On one hand, the green area is classified into 4 groups (Table 1), where each group has a density level and height. Density is classified into 3 groups: High, Medium, and Low. The height of the trees also got classified into two groups: high and low.

Table 1- Green area classification

Green area types	Density	Height
Jungle park	High	High
City park	Medium	High
Local park	Low	High
Greenspace	Low	Low

5

On the other hand, buildings (Table 2) have two important factors for absorbing the pollutants: Structures heights and floor area ratio. Due to the lack of data for structures height in Tehran, this factor got classified based on observations, but the floor area ratio is obtained from the Comprehensive Plan of Tehran City. The multiplier of density and height of the classes are calculated using the GA.

Table 2- Tehran regions height and building density

Region	*Floor Area Ratio*	*Height*
1,2,3	High	High
8,9,10,11,17	High	Medium
5,15,16,18,19,20,21	Low	Low
4,6,7,12,13,14	Medium	Medium

3. Methodology

3.1 Modeling of Traffic and air pollution

Air pollution is a phenomenon, which is affected by different factors. Motor vehicle traffic is known as the source of air and noise pollution (Ozer et al. 2009), blight, vibration (Harrison 2004), and congestion. PM is one of the emissions of petrol vehicles that remains in the atmosphere for a long time, which appears mainly in the form of smoke and suspended particles.

In this study, five 2D matrices are considered as input layers (green space, buildings, wind speed and direction, and traffic). Each layer's cell size is 300 meters, which means that each cell shows an area of 90000 square meters. First, we normalize the traffic volume, then the traffic volume is used to calculate the pollution. Meaning, wherever is the traffic volume higher, there are more pollutants.

The traffic map is created as the base map for every hour, then the other factors that affected the pollution are added in the next steps. Each cell has a value that represents the pollution in that area; the wind moves the pollution based on the direction and the speed and leaves a trace on the primary cells. Each cell has

6

wind characteristics, green space's height, and density type, and buildings' height and density. With the existence of green space or buildings, the amount of air pollution is reduced (figure 3).
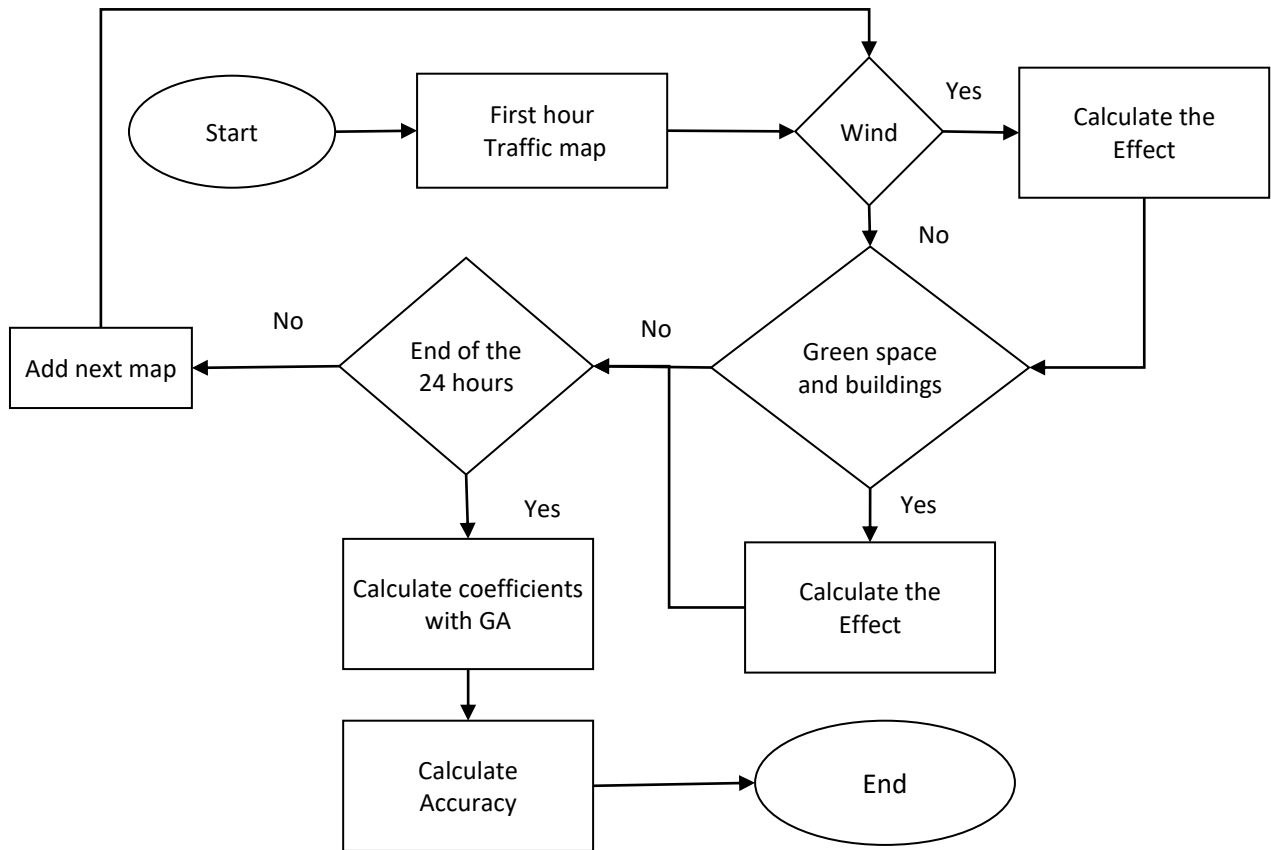

Figure 3- Schema of the algorithm

The traffic of each hour is added to the previous pollution map, and the effect of the wind, green spaces, and the buildings will be calculated again. At the last iteration, the real pollution, and the determined pollution based on the proposed method are compared with a Pearson's linear correlation coefficient (Equation 1), where U is the actual (PM$_{2.5}$ and PM$_{10}$) pollution obtained from the AQCC stations, and C is the calculated pollution from the Cellular Automata.

Equation 1- Pearson's linear correlation coefficient

$$\text{corr}(U, C) = \frac{\sum_{i=1}^{n}(X_{U,i} - \overline{X_U})(Y_{C,i} - \overline{Y_C})}{\sqrt{\sum_{i=1}^{n}(X_{U,i} - \overline{X_U})^2 \sum_{j=1}^{n}(Y_{C,j} - \overline{Y_C})^2}}$$

7

Equation (1) is used as the fitness function for evaluating the coefficient of each factor and its effect.

3.2. Calculating the coefficient of the factors

A Genetic Algorithm is used with 50 feasible nonlinear populations, which calculates the effect of wind, green space (density and height), and buildings (based on height and floor area ratio) absorption. Each coefficient is between 0 and 1, where 1 means that the element completely absorbs the pollution, and 0 means it has no effect. After each iteration, the CA uses the newly calculated coefficient to predict the pollution. Then the obtained result get compared to the actual pollution (Equation 1) to calculate the cost function. In each iteration, the GA tries to optimize the coefficient and improve the cost function. If the algorithm cannot obtain a better result, it will stop.

These coefficients increase the reliability and accuracy of the model. As a simple comparison is drawn between the model with and without the effect of the meteorological and building factors, 73% improvement is noticed in the results.

4. Results

Diffusion of the pollutant is a complex process that is influenced by various factors, and this simulation can only represent a simple distribution of the pollution
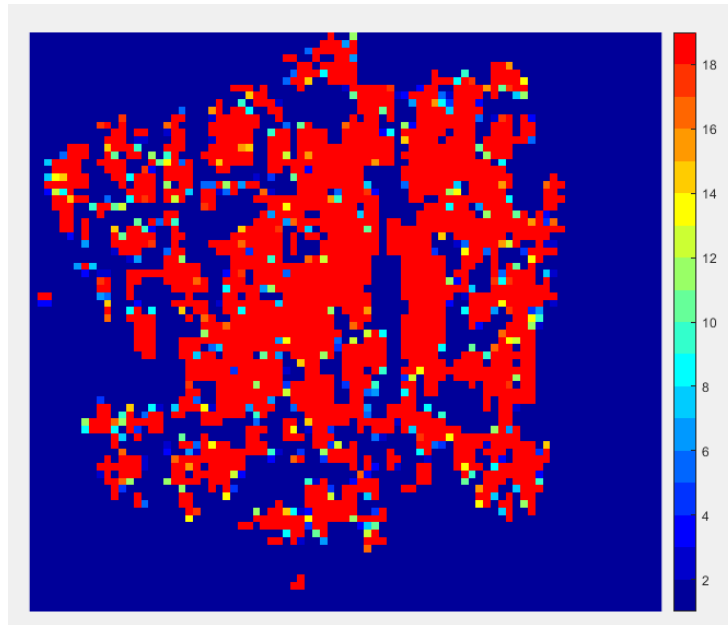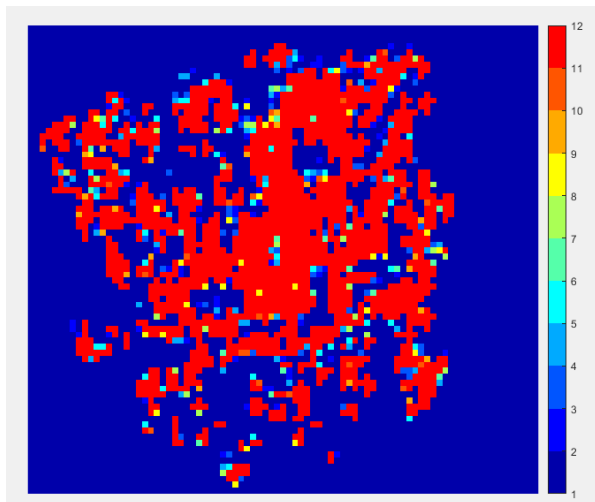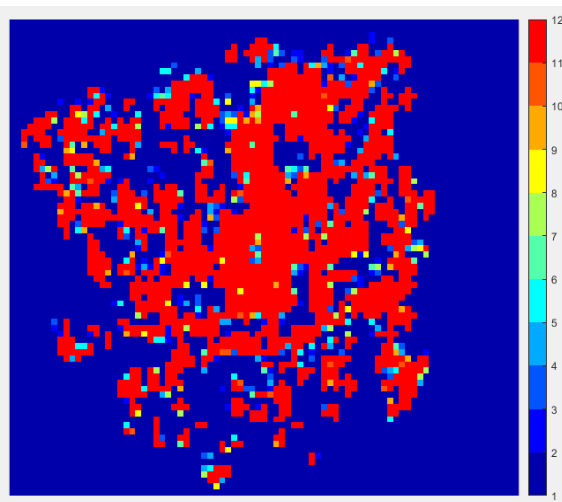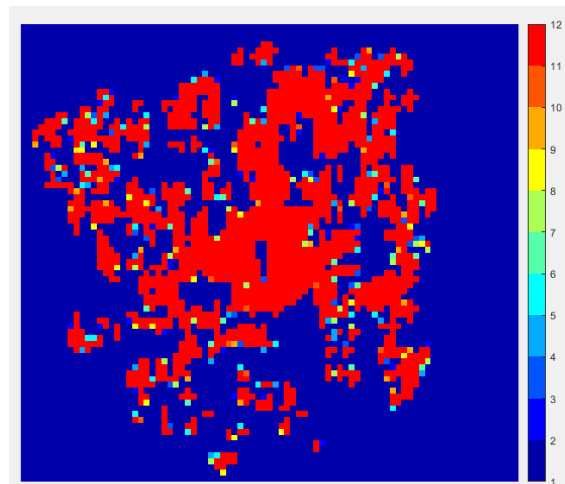
Figure 4- Pollution map of August 10th, 10 am

a

b

c

Figure 5- a) Pollution prediction with the effect of elements, b) Without the effect

c) Actual air pollution (at 10 am 11th August)

Figure 4 shows the pollution map of 10 am of August 10th, wherein the next 24 hours the wind was mostly blowing from north-west to south-east from 110 to 165 degrees from 1 to 3 m/s. With a comparison between Figure 4 and 5, it is shown that the pollution has diffused to the southeast and the pollution (PMs) has been decreased in the north.

Model	Green area density (G_d)	Green area Height (G_h)	Buildings Density (B_d)	Buildings Height (B_h)	Wind effect	Correlation with the real pollution	Mean Square Error (MSE)
Full effect of elements	*1*	*1*	*1*	*1*	*1*	*0.0819436151604*	*221.007674262*
Without elements	*0*	*0*	*0*	*0*	*0*	*-0.155171819816*	*176.5431890403*
GA	*0.12125656854*	*0.17471192390*	*0.0165684564*	*0.04302724770*	*0.33363248525*	*0.584111449796*	*44.78947368421*

Table 3- Models and correlation

Three scenarios have been evaluated as a comparison to understand the effect of coefficients (Table 3). In the first scenario, the factors are set to have a full effect (the wind carries the pollution away, and green area and building existence also absorbs the pollutions), and the accuracy of 8% got obtained in comparison with the AQCC stations.

In the Second scenario, the effect of elements was completely ignored, meaning that each layer only got added up to the previous ones. In this scenario, a correlation of 0.15 got obtained when it was compared to the real pollution index.

In the third scenario, the effect of the elements was calculated with GA. The Genetic Algorithm Calculated the effect of the green-space absorption about 30% overall, meaning that the parks and other urban green spaces play an important role in reducing pollution. Results show that the green area height has about 5% more effect on reducing pollution in comparison with density. Meaning that the height of the trees is more important than their density.

On the other hand, buildings only absorbed about 6% of pollution, and the height of the buildings also played a more important role than the density. The wind was an important factor as well, which had 33% effectivity in pollutants diffusion. Generally, the GA obtained 58% accuracy for predicting pollution.

Table 4- Multipliers of density and height of Buildings and Green area

Multipliers	High	Medium	Low
Green area density (M_{gd})	1	0.6845131661	0.4163275712
Green area height (M_{gh})	1	-	0.1731389546
Buildings density (M_{bd})	1	0.7853446217	0.3516458453
Green area height (M_{bh})	1	0.7143545216	0.1384548326

The effect of buildings and green areas' height and density got calculated with GA. The result is a multiplier for each factor (Table 4).

Equation 2- Green area and building effect

$$Green\ area\ effect = \ M_{gd} \times G_d + M_{gh} \times G_h$$

$$Buildings\ effect = \ M_{bd} \times B_d + M_{bh} \times B_h$$

These multipliers change the effect of each factor based on the density and height of green areas and buildings. The numbers also show that the density has more effect than height. The effects of height and density are calculated with equation 2.

5. Conclusion

Due to the hazard that air pollution creates for health, it is important to predict the diffusion of the vehicular pollutants as the main source. In this paper, the emitted pollutants from vehicles exhaust got used as the main source of urban air pollution. The traffic map got processed and transformed into a pollution map, and the effect of buildings and green areas was determined using the Genetic Algorithm. Because of the different heights and densities in the area, a multiplier got used to increase the effectivity of the model, which was also determined with GA. In the end, by comparing the data from AQCC stations and the model, it was able to predict air pollution with 58.4% accuracy.

The wind pattern is complex and changes every minute, so the pollution stations do not report this, which holds us from obtaining better accuracy. The effect of buildings and green areas on the wind was also another complexity that this paper faced. On the other hand, many other meteorological phenomena affect the distribution of pollutants, temperature for instance.

References:

Ai, J., S. Wu, Z. Wei, H. Xu & G. Liu. 2016. *Diffusion Model Based on Cellular Automata and Its Application*.

Behzadi, S. & A. Alesheikh (2014) Cellular Automata vs. Object-Automata in Traffic Simulation. *International Journal of Remote Sensing Applications,* 4**,** 61.

Burroughs Peña, M. S. & A. Rollins (2017) Environmental Exposures and Cardiovascular Disease: A Challenge for Health and Development in Low- and Middle-Income Countries. *Cardiol Clin,* 35**,** 71-86.

Chin, A. T. H. (1996) Containing air pollution and traffic congestion: Transport policy and the environment in Singapore. *Atmospheric Environment,* 30**,** 787-801.

Cosentino, C., B. Vescio & F. Amato. 2013. Cellular Automata. In *Encyclopedia of Systems Biology,* eds. W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota, 381-385. New York, NY: Springer New York.

Deutsch, A. 2013. Spatiotemporal Pattern Formation. In *Encyclopedia of Systems Biology,* eds. W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota, 1963-1963. New York, NY: Springer New York.

Fotourehchi, Z. (2016) Health effects of air pollution: An empirical analysis for developing countries. *Atmospheric Pollution Research,* 7**,** 201-206.

Gao, J., H. Chen, K. Dave, J. Chen & D. Jia (2020) Fuel economy and exhaust emissions of a diesel vehicle under real traffic conditions. *Energy Science & Engineering,* 8**,** 1781-1792.

Gauderman, W. J., E. Avol, F. Gilliland, H. Vora, D. Thomas, K. Berhane, R. McConnell, N. Kuenzli, F. Lurmann, E. Rappaport, H. Margolis, D. Bates & J. Peters (2004) The effect of air pollution on lung development from 10 to 18 years of age. *N Engl J Med,* 351**,** 1057-67.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc.

Guariso, G. & V. Maniezzo (1992) Air quality simulation through cellular automata. *Environmental Software,* 7**,** 131-141.

Guo, Y., H. Zeng, R. Zheng, S. Li, A. G. Barnett, S. Zhang, X. Zou, R. Huxley, W. Chen & G. Williams (2016) The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis. *Environmental Research,* 144**,** 60-65.

Hamoudzadeh, A. & S. Behzadi (2021) Predicting user's next location using machine learning algorithms. *Spatial Information Research,* 29**,** 379-387.

Harrison, M. 2004. *Vehicle refinement: controlling noise and vibration in road vehicles*. Elsevier.

Heintz, C., M. Grunwald, S. Edenhofer, J. Hähner & S. von Mammen. 2017. *The game of flow - cellular automaton-based fluid simulation for realtime interaction*.

Herr, C. M. & R. C. Ford (2016) Cellular automata in architectural design: From generic systems to specific design tools. *Automation in Construction,* 72**,** 39-45.

Hoekman, S. K. (1992) Speciated Measurements and Calculated Reactivities of Vehicle Exhaust Emissions from Conventional and Reformulated Gasolines. *Environmental Science & Technology,* 26**,** 1206-1216.

Huynh, H. X., P. T. Luu, H. H. Luong, N. Duong-Trung, M. T. Truong, N. Q. T. Tang & T. C. T. Tran. 2020. Analysis of Atmospheric Quality based on Cellular Automata Simulation. In *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, 27–33. Hanoi, Viet Nam: Association for Computing Machinery.

Iltanen, S. 2012. Cellular Automata in Urban Spatial Modelling. In *Agent-Based Models of Geographical Systems,* eds. A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty, 69-84. Dordrecht: Springer Netherlands.

Jo, H.-k. (2002) Impacts of urban greenspace on offsetting carbon emissions for middle Korea. *Journal of Environmental Management,* 64**,** 115-126.

Kean, A. J., R. A. Harley, D. Littlejohn & G. R. Kendall (2000) On-Road Measurement of Ammonia and Other Motor Vehicle Exhaust Emissions. *Environmental Science & Technology,* 34**,** 3535-3539.

Kim, K. H., S.-B. Lee, D. Woo & G.-N. Bae (2015) Influence of wind direction and speed on the transport of particle-bound PAHs in a roadway environment. *Atmospheric Pollution Research,* 6**,** 1024-1034.

Kurt, O. K., J. Zhang & K. E. Pinkerton (2016) Pulmonary health effects of air pollution. *Current opinion in pulmonary medicine,* 22**,** 138-143.

Leong, W. C., R. O. Kelani & Z. Ahmad (2020) Prediction of air pollution index (API) using support vector machine (SVM). *Journal of Environmental Chemical Engineering,* 8**,** 103208.

Liu, H.-L. & Y. S. Shen (2014) The Impact of Green Space Changes on Air Pollution and Microclimates: A Case Study of the Taipei Metropolitan Area. *Sustainability (Switzerland),* 6**,** 8827-8855.

Maithani, S. (2010) Cellular Automata Based Model of Urban Spatial Growth. *Journal of the Indian Society of Remote Sensing,* 38**,** 604-610.

Makri, A. & N. I. Stilianakis (2008) Vulnerability to air pollution health effects. *International Journal of Hygiene and Environmental Health,* 211**,** 326-336.

Neumann, J. & A. W. Burks. 1966. *Theory of self-reproducing automata*. University of Illinois press Urbana.

Oh, S.-C. (2012) Using an Adaptive Search Tree to Predict User Location. (Using an Adaptive Search Tree to Predict User Location). *Journal of Information Processing Systems,* 8**,** 437–444.

Ozer, S., H. Yilmaz, M. Yeşil & P. Yeşil (2009) Evaluation of noise pollution caused by vehicles in the city of Tokat, Turkey. *Scientific Research and Essays,* 4**,** 1205-1212.

Padmanabhamurty, B. (1975) The Role of Wind in Pollution Dispersion. *Journal of the Air Pollution Control Association,* 25**,** 956-957.

Roetzel, W., X. Luo & D. Chen. 2020. Chapter 6 - Optimal design of heat exchanger networks. In *Design and Operation of Heat Exchangers and their Networks,* eds. W. Roetzel, X. Luo & D. Chen, 231-317. Academic Press.

Salcido, A. & S. Carreón-Sierra (2017) Air Pollutant Emissions in the Fukui-Ishibashi and Nagel-Schreckenberg Traffic Cellular Automata. *Journal of Applied Mathematics and Physics,* 5**,** 2140-2161.

Vahidnia, M. H., A. A. Alesheikh, S. Behzadi & S. Salehi (2013) Modeling the spread of spatio-temporal phenomena through the incorporation of ANFIS and genetically controlled cellular automata: a case study on forest fire. *International Journal of Digital Earth,* 6**,** 51-75.

Wen, C., S. Liu, X. Yao, L. Peng, X. Li, Y. Hu & T. Chi (2019) A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of The Total Environment,* 654**,** 1091-1099.

Westerholm, R. & K.-E. Egebäck (1994) Exhaust emissions from light-and heavy-duty vehicles: chemical composition, impact of exhaust after treatment, and fuel parameters. *Environmental health perspectives,* 102**,** 13-23.

Xiao-ping, R., D. Zhen, L. Jin & S. Xian-feng (2013) Simulation of Point Source Pollution Diffusion Using a Velocity Field-cellular Automata Coupled Method. *Information Technology Journal,* 12**,** 5424-5431.

Yan-feng, Q., X. Yu, W. Xue, C. Bing-ling & W. Yi (2020) Study on particulate matter emissions from traffic by cellular automaton model with slow-to-start effect. *arXiv preprint arXiv:2011.03674*.

Yang, J., J. McBride, J. Zhou & Z. Sun (2005) The urban forest in Beijing and its role in air pollution reduction. *Urban Forestry & Urban Greening,* 3**,** 65-78.

Zeinalnezhad, M., A. G. Chofreh, F. A. Goni & J. J. Klemeš (2020) Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. *Journal of Cleaner Production***,** 121218.