Project 2: PM 2.5 Prediction Based on Simulation of Pollutant Dispersion Checkpoint Report

Team member: Maozhi He, Xiaoyu Pan, Ke Zhang

1. Abstract

Unacceptable particulate matter(PM) levels are increasing the impacts on our health. As a result, it will be helpful in improving our air condition if we can predict or forecast the concentrations of PM. In our study, the concentration of PM 2.5 is studied and predicted by using CA, Neural Network MLP and Markov Chain models. The background of our system is basically about Metro Atlanta. The data we gained includes not only the concentration of PM2.5 in each place but also temperature, humidity, wind, pressure, date and radiation which are all correlated with the concentration element. Those correlated elements we include in our data are all parts of our system and have essential impacts on the state of concentration of PM2.5. For example, according to some physical theories, high temperature has the ability to enhance the activity of particles in our system. And if we have higher humidity in the system, those gas pollution in the air will have more chances to converse into particulate matter. Pressure is the element to show the stability of the environment that particulate matters are staying in. In our study, we will analyze the concentration of PM2.5 in our system for different timelines and predict the behaviours of them in future hours according to the relationship we build in our system.

The goal of the project is to find out whether there is an efficient way to predict PM2.5 on a large scale. The potential application of an accurate prediction is wide. For instance, if we can predict high PM 2.5 concentration in a certain area for the next few days, we can alert the public of the situation and urge them not to participate in outdoor activities and take precautionary measures. It will definitely help those who have existing respiratory problems and have great public health benefits.

2. Description of the system being studied

The forecasting of PM2.5 relies on the simulation of the spread of the pollutants. Thus the system relies highly on the fundamental rules of aerodynamics and particle transmissions. In this study, the system will contain several pollution sources to begin with, each representing a major output point PM2.5 over a period of time. As the simulation progresses, the particle pollutants will be carried to other regions by either atmospheric activities or thermodynamic diffusion. Alongside the transmission of the particles, dilution and decomposition will also occur due to physical and chemical activities. And the following parameters are critical for the system: generation rate, diffusion rate, dilution rate and decomposition rate. These parameters are greatly affected by the local geographical and weather conditions, such as the temperature, humidity, wind speed altitude, season and etc.

In order for such a complex system to function correctly, it is important for us to determine the correct relations between data and the parameters. As well as adjusting the system conditions accordingly.

3. Expected approaches and platform of development

We will use these three approaches: CA, MLP Neural Network and Markov Chain. For the platform of development, we will use Jupyter Notebook in this project.

4. Literature review

In the world of contamination modeling, the main focus has always been choosing the right approach. The earliest mathematical models include Gaussian models, Lagrange models and complex systems like Euler models[1]. These models suffer from being either too simple or too complex. With the development of data analysis and machine learning, researchers try to combine practical models with new simulation techniques. These methods include Support Vector Machine[2], Neural Networks[3] and K-Nearest Neighbors. The new strategy of adapting big data and regression methods shows significant progress in terms of prediction accuracies, but shows limitations at the same time. The overlook of physical-chemical principles and details puts a question mark on the credibility of the result. Also the massive amount of data monitoring and management required by these methods will cause dramatic increment in research costs. Other studies also explore new possibilities by adapting methods like Nonlinear Regression[4] and Exponential smoothing.

In terms of PM2.5 prediction, researchers have developed multiple models to realize it. For example, Ordieres utilized a neural network prediction model to predict the PM2.5 concentration on the US-Mexico border. Three different topologies of neural networks (MLP, RBF and SMLP) have been used and compared. Their results showed that all neural approaches performed better than the classical models. The results also demonstrated that there are very little differences between each topology. In Dong's work, the researchers demonstrated a way to use a hidden semi-Markov model for PM2.5 prediction. They pointed out that the simple hidden Markov model has limited use because of its short-term memory of past history. Their results of prediction in Chicago O'Hare Airport showed high accuracy for the next 24 hours. Furthermore, Deng proposed a method using cellular automata to predict PM2.5 generation and diffusion in Beijing. It incorporated the factor of source and weather into calculation. Their results showed that it is a reliable prediction.

5. Current state of the project

At this point, we have finished the first approach: Cellular Automata.

PART I: CA Model

1. Description of the Model

Our CA model is composed of several basic components which are cellular space, cellular components, cellular neighbours, cellular states and evolution rules. In our study, the cellular space is the diffusion region of PM2.5. Each cell in our model is surrounded by other eight cells. The cells in our system are made by the contaminated places we choose in our study. The evolution rules can be affected by several matters but we mainly choose the generation, dilution, diffusion and settlement of PM2.5. From our literature researches, we found that the generation of PM2.5 can be summarized by one formula:

(1) $$C(t+1) = C(t) + k_i * (f(T,H))$$

C represents the state of the cell which is the concentration of PM2.5 in that cell. ki is the emission rate for the pollution resource.f(T, H) is a function gained from cell's temperature and humidity data. In our first state study, we choose one zip code from Atlanta to be the pollution resource.

The diffusion of particles in air can be divided into free movement based on concentration difference and wind spread influenced by wind speed and wind direction.

For our system, the formula that describes free movement is:

(2) $$Dispersion = (C_c - C_n) * f(Alt,T)$$

Where Cc is the concentration of the center cell and Cn is the concentration of the neighbour cell. The coefficient that times the concentration difference is gained related to the cell's altitude and temperature data.

In order to describe wind spread, we come up with a formula:

(3) $$C_N(t+1) = C_N(t) + C_c(t) * f(W) * \sin(\deg)$$

Where f(W) is a function gained from a cell's wind speed and deg represents the wind direction in that cell.

In our System, concentration of PM2.5 will be reduced in two processes: dilution and deposition. For dilution, we have:

(4) $$C(t+1) = C(t) * (1 - f(W,T))$$

Where f(W, T) is a function gained from a cell's wind speed and temperature.

For deposition, we have:

(5) $$C(t+1) = C(t) - d_i$$

Where di is the rate of precipitation.

6. Initial Results

At this point, a couple of tests were performed on the CA model. In the first case , wind speed in all cells was set to 0 km/h, meaning the transmission and dilution effect by the wind was completely ignored. As shown in Figure 1 after simulating for 100 unit time, the contaminants were spread evenly around the source cell due to the diffusion effect. Here we can see how the particles travel slowly and aggregate in a small area without wind. The second test was based on real time data (data of the Atlanta area collected on April 4th) but changing the wind direction to 315° North West of all cells to see how wind would affect the system. Figure 2 shows a clear path of the transmission of PM2.5 along the North West direction as well as the dilution effect done by the wind. When this is compared to the result demonstrated in the prior test, we can notice two major differences, one being the contaminants traveling with the wind to further regions at a much faster speed. And secondly, the wind movement contributes greatly in diluting the pollutants, dropping the concentration of the source cell from 10ug/m3 to 3ug/m3. This agrees with the common knowledge of a much clearer sky on a windy day, which is promising evidence that this model is simulating correctly. The final study performed was to simulate the model according to the actual data collected. This time the model supposes a pollutant source representing the heavy traffic during rush hours, this is definitely interesting to see how the PM2.5 moves around the city.

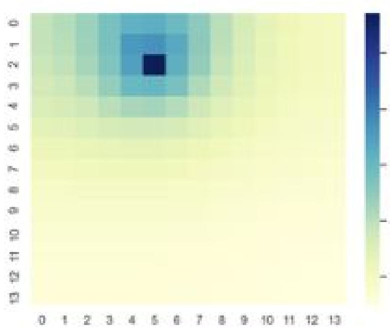

Fig1. No Wind Considered

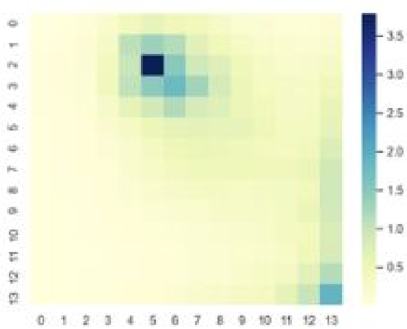

Fig2. With Wind Degree of 315 (NW)

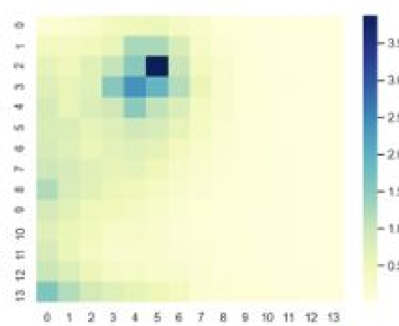

Fig3. Real Time Data From Atlanta

Based on the tests we have completed and the above figures, we can conclude that our model is working properly on different conditions. Our next step will be investigating how particles move within a period of time and how accurately this model will predict the concentration of PM2.5. Additionally, we will finish the other two approaches and compare the three of them.

Division of Work:

Maozhi He: Theoretical modeling and coding for ML approaches for PM2.5 prediction.
Xiaoyu Pan: Research and data collection. Coding and analyzing for three approaches.
Ke Zhang: Writing tutorials. Coding and analyzing for three approaches.
Github repository: https://github.com/kevinzhang99/CSE6730-project2

References:

1. The MR-CA Models for Analysis of Pollution, Sources and Prediction of PM2.5, Fang Deng, Member, IEEE, Liqiu Ma, Xin Gao, and Jie Chen, Senior Member, IEEE. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS
2. A novel hybrid-Garch model based on ARIMA and SVM for PM2.5 concentrations forecasting, Ping Wang, Hong Zhang, *, Zuodong Qin, Guisheng Zhang. Atmospheric Pollution Research
3. J. B. Ordieres, E. P. Vergara, and R. S. Capuz, "Neural network prediction model for fine particulate matter (PM 2.5) on the US–Mexico border in EL Paso (Texas) and Ciudad Juárez (Chihuahua)," Environ. Model. Softw., vol. 20, no. 5, pp. 547–559, 2005.
4. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations, W. Geoffrey Cobourn. Atmospheric Environment 44 (2010) 3015-3023
5. Ordieres, J. B., et al. "Neural network prediction model for fine particulate matter (PM2. 5) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua)." Environmental Modelling & Software 20.5 (2005): 547-559.
6. Dong, Ming, et al. "PM2. 5 concentration prediction using hidden semi-Markov model-based times series data mining." *Expert Systems with Applications* 36.5 (2009): 9046-9055.
7. Deng, Fang, et al. "The MR-CA Models for Analysis of Pollution Sources and Prediction of PM 2.5." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*49.4 (2017): 814-820.