The Visualization of Taiwan's Air Quality Map Grid Using IDW for Open Data Diffusion

Ting-Yu Kuo
Department of Computer Science,
Tunghai University, Taichung City, Taiwan, (R.O.C.)
Chao-Tung Yang*
Department of Computer Science,
Tunghai University, Taichung City, Taiwan, (R.O.C.)
http://ithu.tw/orcid
Shun-Hung Teng
Department of Computer Science,
Tunghai University, Taichung City, Taiwan, (R.O.C.)
Yuan-An Chen
Department of Computer Science,
Tunghai University, Taichung City, Taiwan, (R.O.C.)
Hsing-Chung Chen†
Department of Computer Science and Information Engineering,
Asia University, Taiwan, (R.O.C.)
Department of Medical Research,
China Medical University Hospital, Taiwan, (R.O.C.)

Abstract

With the current growth of climate and air pollution, there are many mobile apps or websites of all types that can monitor air quality in real-time. However, most of them are integrated official Open Data and numerical scale on the map throughout the websites or programs. Only where the official station is to be set up, there is numerical air quality, and most areas do not have a clear reference to air quality value. For these reasons, we decided to create an air quality map to expand the start point. Python is used to import Taiwan's 1KM*1KM grid map Shapefile and then use the web crawler to sort the PM 2.5 values from the Civil IOT Taiwan and Academia Airbox open data into a format to make Taiwan's 1KM*1KM air quality map. Then we apply Inverse Distance Weighting to infer those blank grids that do not have PM 2.5 values to complete the entire map and allow users to use on the website.

Keywords: PM 2.5, air quality map, open data, inverse distance weighting method, IDW, Open Data Diffusion

1 Introduction

In recent years, issues of air pollution has been the focus of everyone's attention, from motor vehicles and industrial, floated abroad The pollutants are polluting the air in Taiwan, and people's awareness of health and the environment is gradually rising. In view of this, a wide range of products that monitor air quality, as well as many applications or websites that can instantly query air quality, are springing up. But are these products or websites really available to users in different parts of Taiwan.

G. Sutcliffe, A. Voronkov (eds.): easychair 1.0, 2008, volume 1, issue: 1, pp. 1-11

*Corresponding Author: ctyang@thu.edu.tw
†Corresponding Author: shin8409@ms6.hinet.net, cdma2000@asia.edu.tw

This study first considers the distance as the main consideration factor to speculate that there is no monitoring station. The method adopted is the inverse distance weighting method. The concept of the inverse distance weighting method is to estimate each unknown value and use it to approximate the known point. The value is used to perform the weighting operation. The weight given is calculated according to the distance and is Taiwan according to the size of the value 1KM*1KM colored in the map grid of. By outputting the map on the web page, the user can access the browser. The operation uses the map, the mouse clicks on the grid to see the current air quality data of the grid, and the interactive map makes the user more convenient to operate the map, and the usage is easier to use, hoping to integrate effective information to the user. The platform for the query.

2 Methods

2.1 Python

Python is an object-oriented, high-level programming language and a literal translation language.Python emphasizes the readable, easy-to-learn, easy-to-learn (simplified and clear syntax features) of the programming language and speeds up the development of the program. It is easy to use, can be used for all kinds of difficulty, and can be run on most systems. Developed with the concept of reducing development and maintenance costs. This study uses Python as the main development language.

2.2 The inverse distance weight

The method is called Inverse Distance Weighted (IDW). It is estimated for each unknown value. The weighted operation is performed by using the value of its neighbouring known points. The weight given is calculated according to the distance. The concept is a point of unknown value, which is inversely proportional to the distance of the value of the known point around it. The farther the distance is, the smaller the degree of influence is. The degree of influence is the power of the distance between the unknown point and the known point. Said. The formula is as follows:

$$f(x,y) = \left[\sum_{N}^{i=1} w(d_i) z_i\right] \div \left[\sum_{N}^{i=1} w(d_i)\right] \tag{1}$$

Where $w(d_i)$ is the weight equation, which z_i is i of the the value first known point, which d_i is i the distance from the point to the unknown point. The size is determined by the inverse ratio of the power. Although there are a variety of spatial analysis interpolation methods that can be used to estimate areas where there are no values, according to [1] the paper shows that the concentration of PM 2.5 meshing has better simulation results for using the inverse distance weight method. We have used this method to fill in areas that have no value.

2.3 QGIS

QGIS (formerly known as Quantum GIS) is a free software desktop GIS software. It provides the ability to display, edit, and analyze data. QGIS can be used to edit geographic data, such as: Shapefile. This study uses this software to format the grid layer and rearrange the grid number.

3 System Design and Implementation

3.1 System Architecture

The System Architecture is shown in Fig. 1. Implemented in the Linux Ubuntu operating system, the air quality data sources using open networking people biological information and the Academia Sinica's Open Data. The data through Python performs the inverse distance weighting analysis operation. Then integrate the calculated data into the layer that is colored in the Shapefile and overlap with the map. Finally, use the Apache server to set up a website, so that users can query the results map information through the browser.

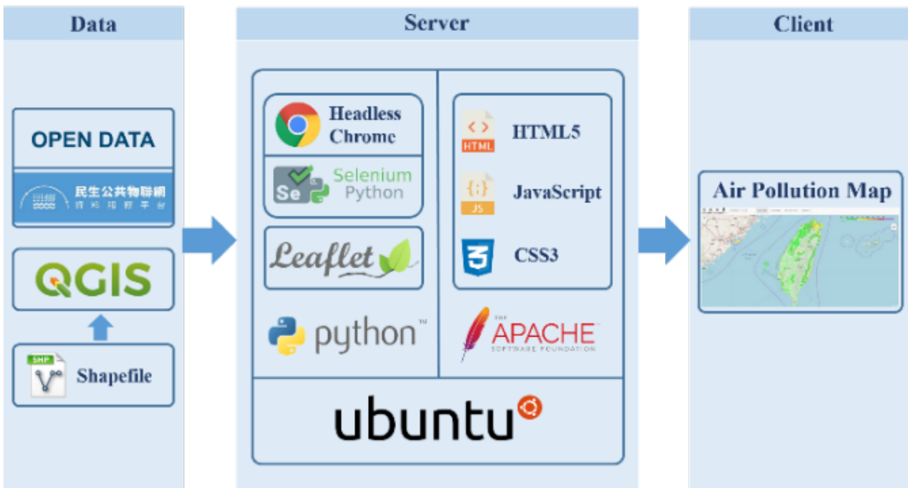

Figure 1: System Architecture

3.2 Experimental Procedure

3.2.1 Data Acquiring

First, in Python, grab the [2] micro-sensor data of the Min Bio-networking data service platform and [3] the data of the air box of the Academia Sinica, and obtain the data type. For JSON, save the data into Data Frame after finishing. If there is no data on the station failure or stop display on the same day, we will use 0 to indicate that the total amount of data is about 3,000. The data will be saved as csv file for subsequent operation. Figure 2 shows the Min Bio-Based Data Service Platform page.

Figure 2: Min Bio-Data Service Platform

3.2.2 Grid Layer

Based on the Taiwan grid layer used in this study, the data format is in Shapefile, read into Python by Geopandas, presented in Data Frame, and found that its grid number is not the exact order as shown in Fig. 3.

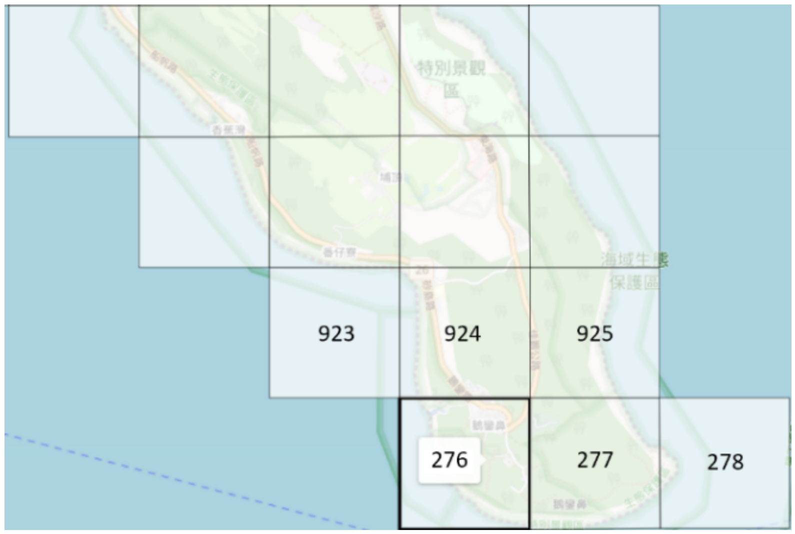

Figure 3: unordered mesh number

In order to make the next process smoothly, the present study used QGIS numbered rearranged. After input the expression, the id attribute of the layer was integrated with the sorting location, as shown in Fig. 4.

Finally obtained number of grid layers is completed, the order of left to right, from the bottom up, as shown in Fig. 5.

3.2.3 Inverse Distance Weighting (IDW) Method

IDW uses the ability to derive the characteristics of unknown points, and uses the inverse distance weighting method to fill the grid without the station data as much as possible. But, if the unknown value is too far from the station, it will make the value after IDW has no reference value. Therefore, we set the priority reference to the station within 3 km from the unknown point as the IDW data. If there is no station in the range, increase the distance to 5 km. If there is still no station, set the PM 2.5 value to 0 to ensure it would not make the value meaningless because the distance is too far. Nor will it result in the weight being diluted and not spreading due to too many known points. Figure 6 describes a schematic diagram showing the weights being diluted by reference to many known points. It can be seen that there is still a good air color around the area where the air quality is poor.

The amount of data is about 3,000, and the grid has a total data about 37,000 grids, also will involve multiple layers of loops and square operations. In order to speed up, Pandas used the Data Frame column operation instead of the For Loop and the List to write the IDW algorithm, which greatly shortened the time required for the operation. Table 1 shows the data obtained after actually testing the two methods.

3.2.4 The values are merged into the layer

Each grid has its corresponding PM 2.5 value, and to present different colors, we need to complete the IDW value, merge it into the Data Frame of the grid according to the corresponding grid number, and

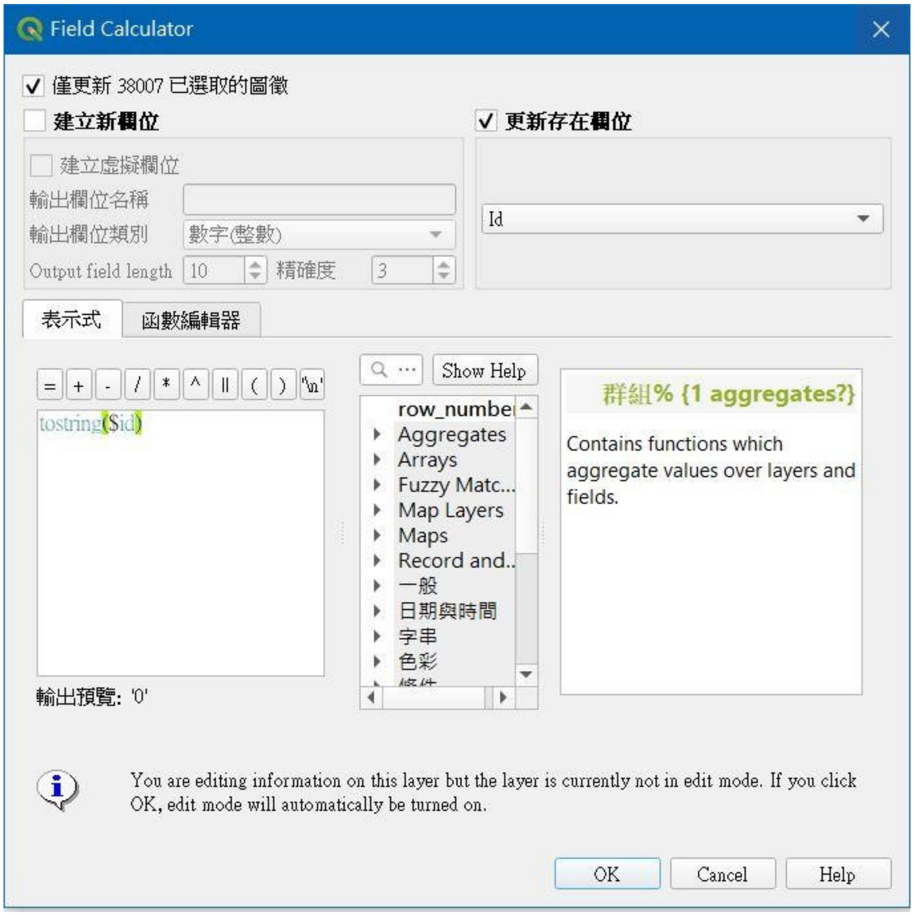

Figure 4: QGIS re-arrangement

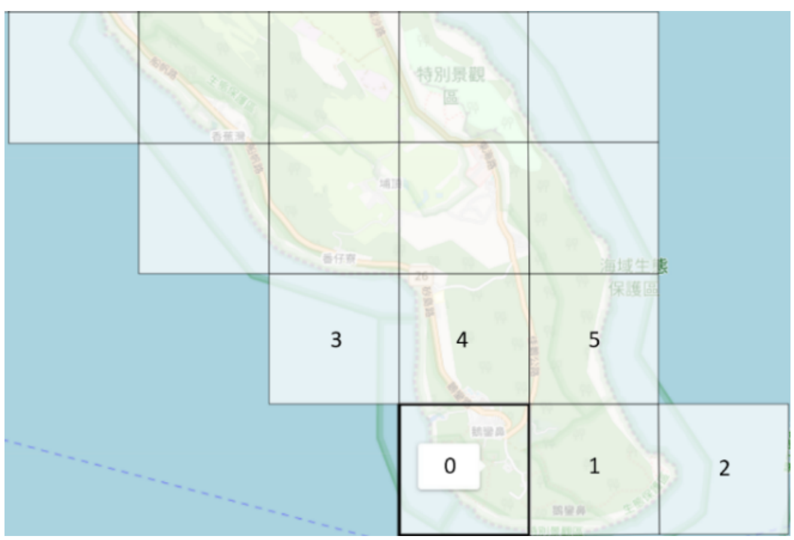

Figure 5: Completed sorting grid number

then write the color settings. The color of the presentation refers to the EPA's classification of the severity of PM 2.5. It is divided into 9 colors as shown in Figure 7, taken from the EPA data. After the IDW

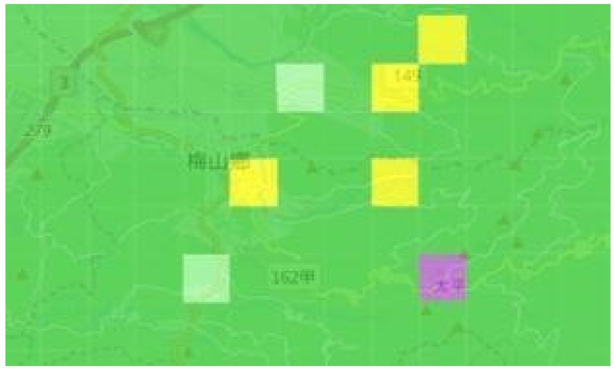

Figure 6: Weights are diluted Schematic

Table 1: Execution time comparison

Usage method	**For loop + List**	**Pandas**
execution time is	about 0:11:37	about 0:00:08

operation, the value of the grid is 0. In addition to more convenient viewing, it also speeds up the loading of the page. Finally, the grid layer is superimposed with the Leaflet map to complete the creation of the grid map.

分類	低	低	低	中	中
指標等級	1	2	3	4	5
PM2.5 濃度 ($\mu g/m^3$)	0 - 11	12 - 23	24 - 35	36 - 41	42 - 47

分類	中	高	高	高	非常高
指標等級	6	7	8	9	10
PM2.5 濃度 ($\mu g/m^3$)	48 - 53	54 - 58	59 - 64	65 - 70	> 71

Figure 7: PM 2.5 indicator comparison

3.2.5 Historical record playback

In order to better understand the concentration change of PM 2.5 in each region, this study uses python suite Selenium to control the web driver, the browser used is Headless Chrome browser. Its advantages there is no interface, so it is not easy to make mistakes in the execution of the schedule. Also, it is set to automatically capture the current Taiwan grid map every 30 minutes and store it. Finally combine the images in 12 hours into an animation and store them.

3.2.6 Webpage construction

Apache is applied as the server of the web page. Bootstrap.css is introduced into the web page, with JavaScript as an aid to make the layout more neat and convenient. The grid map and the playback animation of the historical playback are embedded on the web page, and the SSL certificate as shown in Fig. 8 is applied to make the function of the webpage operate smoothly.

Figure 8: SSL Certificate

4 Experimental Results

This study mainly combines the Air Pollution Index of the Environmental Protection Agency, the Minsheng Public Internet of Things and the Air Box Open Data to create a map of 1KM*1KM squares with inverse distance weighting method to spread PM 2.5. The system's website address is: https://pm25.thu.edu.tw

Because the EPA micro-sensor and the Academia Sinica air box have different sensing principles, and the setting environment of the two is not the same, so the value is slightly different. The system is divided into two categories: AS-Airbox (Academia Sinica Air Box) and EPA-Sensor (EPD Micro Sensor) for user reference. At present the results of an interactive web page has the following features: (1) Overall system uses color visual way to present the user interactive map, as shown in Figure 9. (2) If the user clicks on the map grid (Fig. 10), a pop-up label will display the PM 2.5 value after the grid IDW and the latitude and longitude of the grid center, in conjunction with the function of obtaining the user's position.

(3) The map has special markings for each air quality monitoring point, which can directly confirm the value. Top right corner of the map layer for the switch to enable or disable the tag as described in Fig. 11.

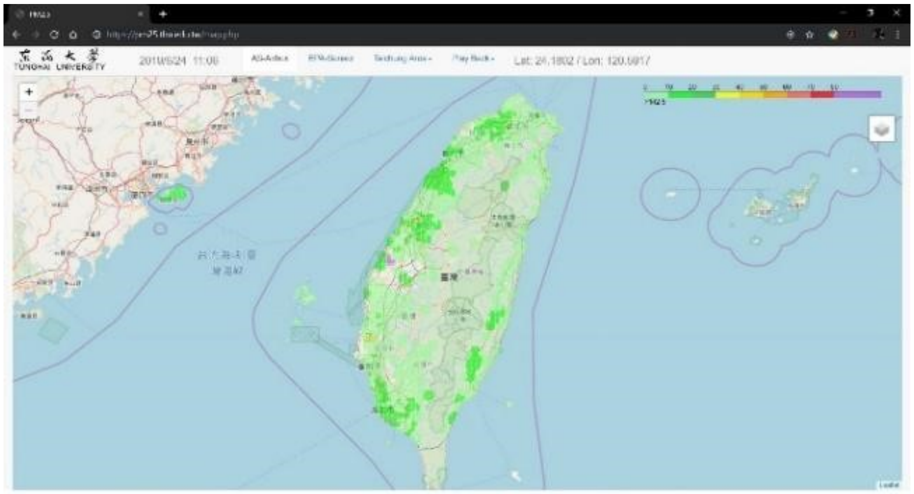

Figure 9: Web Page screen

Figure 10: Click on the grid icon

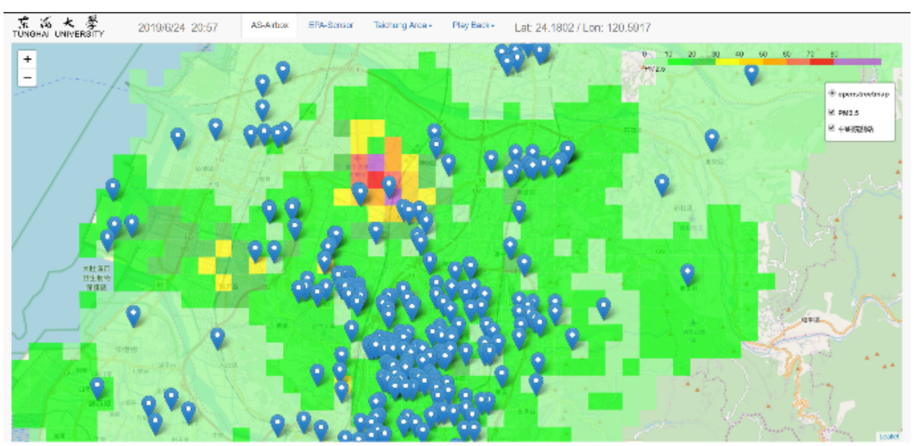

Figure 11: FIG position display station

(4) History playback (Play Back) can be seen more clearly the change in the air quality of each region within 12 hours, and take appropriate precautions, as shown in Fig. 12.

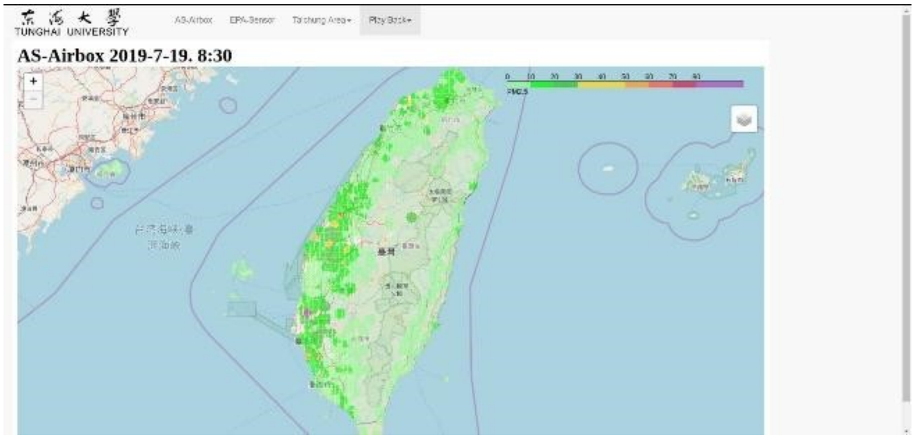

Figure 12: History Record Playback

(5) Considering that most users will use mobile devices to browse, it is set to replace the original button column with a function button when the screen is compressed. Click to display the drop-down menu to browse and select the grid map, make the picture more concise and does not occupy layout, easier to use, as shown in Fig. 13.

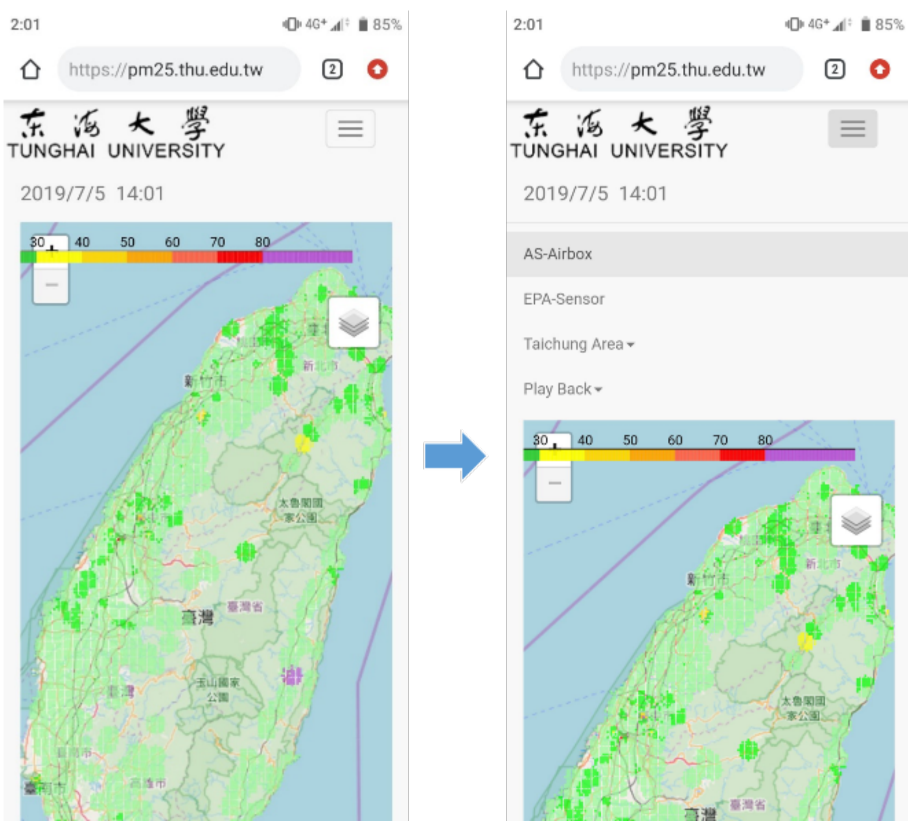

Figure 13: Mobile Device Screen

The system can use by browsing the web page using a browser, which can quickly and easily see the instantaneous air quality in most parts of Taiwan. Therefore, this system can be as an early warning system, like it is a precautionary measure to wear masks and suspend outside for strenuous exercise to avoid physical discomfort. The government can also use this platform to understand the locations and time periods where air pollution is likely to occur, and to strengthen the ban on factories or units that emit excessive emissions. To maintain the air quality of Taiwan.

5 Conclusions and Future Works

5.1 Conclusion

At present, the production of this system has been largely completed and can be used smoothly on the web page. Compared with the purely labelled station location and data of most air quality maps, if we refer to the map after the system has been calculated, it can be seen the current PM 2.5 concentration in the surrounding area more accurately. It is also observed from the map that sometimes the data of some of the grids is particularly high, which may be caused by pollution from the factory operation. Therefore, in the future, the icons and locations of the factories will be added as much as possible to observe whether the specific plants have an impact on the air quality.

5.2 Future Works

In the future, reference will be made to more information in the field of spreading the air pollution concentration to provide more precise predictions. In terms of web page presentation, it is expected to gradually increase the flexibility of web pages, such as: automatically acquiring the location of the user, automatically locking the area, implementing the prediction of future regional values, improving the interface of the web page, etc. It will enable more people to have a platform for easy access to air quality information.

Acknowledgments

This work was sponsored by the Ministry of Science and Technology (MOST), Taiwan, under Grant No. 106-3114-M-029-001-A and 108 -2119-M029-001-A, and MOST 107-2221-E-468 -015. This work was also supported in part by Asia University, Taiwan, and China Medical University Hospital, China Medical University, Taiwan, under Grant ASIA-107-CMUH-05. This work was also supported in part by Asia University, Taiwan, UMY, Indonesian, under Grant 107-ASIA-UMY-02.

References

[1] Zhe-Hao Lai, "Establishing Air Quality Background Concentration of Meshing by Using Objective Analysis in Taiwan", National Taiwan University Environmental Engineering Department, 2013.

[2] Civil IOT Taiwan Data Service Platform, https://ci.taiwan.gov.tw/dsp/environmental.aspx

[3] Academia Sinica Airbox Open Data, https://pm25.lass-net.org/data/last-all-airbox.json

[4] Yung-Hong Wu, "Characteristics if Airborne Particulates in Taichung Urban Area", National Chung Hsing University, Environmental Engineering Department.

[5] Kuei-Chin Wu, "Analysis on the source of PM 2.5 in Central Taiwan Air Quality District", National Chung Hsing University, Environmental Engineering Department.

[6] Chang, Teng-Wei, "The Impact of Stationary Pollution Sources on the Concertation of PM 2.5 in Taichung Area under Different Weather Types", Tunghai University, Environmental Engineering Department.

[7] Hung-Yi Lu, Yee-Lin Wu1, Justus Kavita Mutuku1, Ken-Hui Chang, "Various Sources of PM 2.5 and their Impact on the Air Quality in Tainan City, Taiwan"

[8] Yong-Sheng Ye, "Estimating PM 2.5 Concentrations for Ungauged Sites in Taiwan Using Machine Learning Approaches", National Central University Computer Science Department.

[9] Yen-Ching Chang, Yi-Shiang Shiu, Feng-Cheng Lin, "Applying Spatial Regression Model to Explore the Relationship among Urban Green Areas, PM 2.5 and Land Surface Temperature", Journal of Photogrammetry and Remote Sensing.

[10] Selenium with headless Chrome, https://jiayi.space/post/zai-ubuntufu-wu-qi-shang-shi-yong-chrome-headless

[11] QGIS reset index, http://gis.rchss.sinica.edu.tw/qgis/?p=2625

[12] Folium 0.9.1 documentation, https://python-visualization.github.io/folium/modules.html