Calibration model for air quality data

Luling Duan 1 and Ninggui Duan 2,*

1 Department of Mathematics and Information Science, Guangxi College of Education, Nanning, China
2 Information and Educational Technology Center, YouJing Medical University for Nationalities, Baise, China

*Corresponding author e-mail: duanningsgui@ymcn.edu.cn

Abstract. This article focused on the National Monitoring and control stations and self-built two dust four gas, and the data were analysed, taking into account the humidity, temperature, wind speed, pressure, precipitation of the five meteorological factors, the difference between the two data analysis, and made the establishment of a mathematical model, the self-built Point data was calibrated.

1. Introduction
The effect of air pollution on the ecological environment and human health hazards, through PM2.5, PM10, CO, NO₂, SO₂, O₃ (“two dust four gas”) concentration of real-time monitoring, can better grasp the air quality, timely pollution sources to take appropriate measures[1]. Annex 1 gives the national control point for PM2.5, PM10, CO, NO₂, SO₂, O₃ in a period of time of detection data, taking into account the humidity, temperature, wind speed, pressure, precipitation five meteorological factors on the impact of air quality, Annex 2 gives the National Control Point neighbour self-built Point data (corresponding to the National Control Point:

1. Analysis of National Control Point data and self-built Point Data.
2. Analysis of the factors that caused the difference between the data of the National point of control and the data of the point of origin.
3. With reference to the National Control Point data, establish a suitable mathematical model of the self-built Point data for calibration.

2. Analysis of the problem
Question one:
First of all, we need to collate the data of the National Control Point and the self-built point. According to excel software data filtering, data from the data table can be found in the number of days of data missing. So we need to filter the data again, sort out the number of days complete, data complete month, we found that December 2018, as well as March 2019 these two months of data is complete. And then find the complete month every day “two dust four gas” average concentration value, by MATLAB software [2], for December each build Point and the National Control Point “two dust four gas” data were compared to obtain a graph, after comparing the graph, data analysis.

Second, with excel software on March the National Control Point and the self-point “two dust four gas” average concentration, drawn monthly line chart, and then press the monthly numerical points were compared by the ups and downs of the line segment, it can be seen that there are differences in the

1. 2. 3.
concentrations of “two dust four gas” measured at overseas control points and self built points. It can also be judged from a large range of line segments that the meteorological parameters affecting “two dust four gas” are temperature, wind, air pressure, precipitation and so on. In contrast, the undulating degree of polyline little “two dust four gas”, they can also be inferred by temperature, wind, pressure, precipitation and other meteorological parameters is small, the value is relatively stable.

Question two:
Using excel software, the data sheets of Annex 1 and Annex 2 were collated on a monthly basis, averaging the data of PM2.5, PM10, CO, NO₂, SO₂, O₃ concentrations for the 8-month period from November 2018 to June 2019. concentration variation pattern. The factors causing the change were analyzed by the change of the monthly average concentration bar graph of the self-built Point and the National Control Point “Two dust four gas”, and finally the correlation analysis was carried out.

Question three:
According to the National Control Point Data, Self-Built point data for screening, using excel to find the data close to the whole point of time, while calculating the average data of the National Control Point and the self-built point per day, sorting out the “two dust four gas” and the average daily concentration of meteorological parameters data. Then use the SPSS software [3] were “two dust four gas” and five meteorological parameters regression automatic linear model analysis, excluding the impact of meteorological factors, make the difference between the corresponding national control point and self-point maximum impact meteorological factors related regression scatter plot, then the self-built point Meteorological parameter data into the function, the data obtained after calibration.

3. Symbol description

| Symbols | Meaning                      | Unit |
|---------|------------------------------|------|
| Y       | Concentration                |      |
| X₁      | Average wind speed           | m/s  |
| X₂      | Average pressure             | Pa   |
| X₃      | Average precipitation        | mm/m²|
| X₄      | Average temperature          | °C   |
| X₅      | Average humidity             | rh%  |
| r       | Measure of concentration trend|      |

4. Model assumptions
1. Assuming that some of the missing data in the Annex has little effect on the calculation results.
2. Assuming only consider the environmental factors given in the title (temperature, humidity, wind, pressure, precipitation).
3. Assuming that the monitoring data from the National Control Point is not affected by the weather always maintain accurate.
4. Assuming that each state control point and the distance between adjacent self-built point is equal.
5. Under the assumption of zero drift, range drift and unconventional gaseous pollutants exist, there is a difference from the point of data, but can run normally.
5. Model establishment and solution

5.1. Establish the model and solving the problem one

5.1.1. Compare the concentration of “two dust and four gas” between the National Control Point and the self-built point in December. According to the literature [4], due to the electrochemical gas sensor miniature air quality detector used after prolonged use will produce a certain zero drift and range drift. There are some differences between the data values collected by the self-built Point and the National Control Point. According to the data table we can find some days of data missing, so the use of excel software data integration, the data of the National Control Point and the self-built point utilization average of a data point of the day, the concentration of “two dust and four gas” per day and the meteorological parameters of the self-built point per day are calculated by averaging method. From the collated data tables, we found that some days of data were missing, while December was the month in which data could be collected in full. Therefore, we analyse the data using the months in which we can collect the complete data for the entire month [5]. The data collected by the National Control Point collated with the self-built point of the collected data were compared using MATLAB software to calculate a more intuitive chart. Then following PM2.5, for example, to the point of departure and self-contrast image points, as shown in Figure 1:

Figure 1. PM2.5 comparison image of national control point and self built point.

Figure 1 shows that in December, since the point because of temperature, humidity, wind speed, pressure, precipitation and other meteorological factors, its PM2.5 graphics are above the national control point, in the first 1 to the first two days, self-point PM2.5 is much higher than the national control point, in the first 3 to 5 days, the two curves fell, reaching the lowest point, in the first 6 to 13 days, the curve continued to rise, the self-built point of PM2.5 even reached 117, then, has continued volatility, volatility is relatively large.

5.1.2. Comparison of the concentration of the National Control Point and the self-built Point “two dust four gas” in March. With Excel software, were obtained by the state control point and self-built two dust four gas at the temperature, humidity, wind, pressure, precipitation contrast graphics, through the ups and downs of the line segment, you can see the state control point and self-built point measured “two dust four gas” the concentration of the difference, from the large segment to determine the the effect of precipitation and other meteorological parameters is relatively small, the value is relatively stable. We take March 2019 as an example, we get the average concentration of “two dust and four gas” in the National Control Point and self-built point, as shown in Figure 2 and Figure 3.
Figure 2. National Control Point March 2019 "two dust four gas " average concentration line chart.

Figure 3. Self-built Point March 2019"two dust four gas " average concentration line chart

From Figure 2, Figure 3, regardless of the state control point or self-point, the measured CO concentration did not change much, relatively stable. Therefore, the temperature, humidity, wind, pressure, precipitation and other meteorological parameters, the concentration of CO did not have much effect. The concentration of CO in other months also has little effect (see support material for details).

5.2. Establishing the model and solving the problem two

5.2.1. Analysis of monthly mean concentration data of “two dust four gas ” between the National Control Point and the self-built point. Since the table data in Annex 1 and Annex 2 is more than enough, we will deal with the concentration of “two dust and four gas” on a monthly basis. we have calculated the concentration of “two dust and four gas” for a total of 8 months from November 2018 to June 2019, and made a monthly average concentration bar of the National Control Point, as shown in Figure 4:

Figure 4. National Control Point November 2018 to June 2019 “two dust four gas” monthly average concentration.
As can be seen from Figure 4, the monthly average concentration of “two dust four gas” has changed in these eight months, and the monthly average concentrations of PM2.5 and PM10 have increased significantly from November to February, while the monthly average concentration of O3 has increased significantly from November 2018 to June 2019. The concentration of “two dust four gas” has a seasonal change.

5.2.2. The change of concentration of “two dust four gas” between the National Control Point and the self-built point. Using excel software, the National Control Point and self-point PM2.5, CO, O\(_3\), SO\(_2\), NO\(_2\) concentration-related data monthly average calculation, calculated within eight months of “two dust four gas” monthly average concentration, made the National Control Point and self-point average monthly concentration comparison chart to PM2.5, for example, as shown in figure 5.

Since the data for November 2018 and June 2019 is less than one month, the average concentration of PM2.5 in these two months is not taken into account. As can be seen, in January the highest average concentration of PM2.5 month, in May the lowest average concentration of PM2.5 month. It can be seen, the concentration of PM2.5 decreases as the seasons change.

5.2.3. Correlation analysis. The correlation coefficient is an indicator of the degree of correlation between variables. Sample correlation coefficient represented by \( r \), the overall correlation coefficient represented by \( \rho \), ranging from the correlation coefficient is \([-1, 1]\). \( \gamma > 0 \) is positively correlated, \( \gamma < 0 \) is negatively correlated, \( \gamma = 0 \) indicates uncorrelated. The greater the value of \(|r|\), the smaller the error \( Q \), the higher the degree of linear correlation between variables; \(|r|\) the closer the value of 0, The Greater The Q, the lower the degree of linear correlation between variables.

First of all, we will conduct a single factor correlation analysis of the meteorological indicators in 2018-2019, respectively, in the air quality, using the correlation coefficient formula:

\[
r_{XY} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}}
\]

Calculated by SPSS20. 0 PM2.5, PM10, CO, NO\(_2\), each correlation coefficient SO\(_2\), O\(_3\) and five meteorological parameters were obtained in Table 2, Table 3, Table 4, Table 5, Table 6, Table 7.

**Table 2.** Correlation coefficient of PM2.5.

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| PM2.5 correlation coefficient R | -.261**    | .304     | .039          | -.388**     | .318**   |
Table 3. Correlation coefficient of PM10.

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| PM10 correlation coefficient R | -.277**    | .417**   | .125*         | -.464**     | .361**   |

Table 4. Correlation coefficient of CO.

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| CO correlation coefficient R | -.355**    | .133*    | .206**        | .235**      | .054     |

Table 5. Correlation coefficient of \( NO_2 \).

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| \( NO_2 \) correlation coefficient R | -.289**    | .136*    | .406**        | -.100       | .105     |

Table 6. Correlation coefficient of \( SO_2 \).

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| \( SO_2 \) correlation coefficient R | -.138*     | .044     | -.131*        | -.064       | .005     |

Table 7. Correlation coefficient of \( O_3 \).

| Indicators                  | Wind speed | Pressure | Precipitation | Temperature | Humidity |
|-----------------------------|------------|----------|---------------|-------------|----------|
| \( O_3 \) correlation coefficient R | -.199**    | -.098    | .371**        | .304**      | -.242**  |

Five Meteorological indicators by a single factor correlation analysis showed that: PM2.5 concentration and temperature, wind speed was negative correlation, positive correlation with humidity; PM10 concentration and temperature and wind speed was negative correlation, positive correlation with humidity and pressure; CO concentration, and temperature and precipitation was positive and negative correlation with wind speed; \( SO_2 \) concentration is negatively correlated with precipitation and wind speed; \( O_3 \) concentration is positively correlated with precipitation and temperature, and humidity and wind speed are negatively correlated.

5.3. Establishing the model and solving the problem three

5.3.1. Prediction of the factors influencing the concentration of "two dust and four gas" in self-built sites and national control points. The use of SPSS regression automatic model to detect the impact of PM2.5, PM10, CO, \( NO_2 \), \( SO_2 \), \( O_3 \) maximum of two meteorological parameters factors, such as the impact of temperature and wind speed on PM2.5 is the largest, as shown in Figure 6.
Then use excel software to calculate the difference between the control point and the self-built Point “Two dust four gas” concentration, as shown in Figure 7.

Using the resulting difference in SPSS software, find the impact of “two dust four gas” concentration of the largest two meteorological parameters scatterplot to arrive at a mathematical model.

Use SPSS software, according to the number of days on the impact of PM2.5 concentration of the largest meteorological factors to predict the results are as follows:

Figure 6 shows that the concentration of PM2.5 maximum impact on the meteorological factors are temperature and wind speed, so we mainly affect the concentration of PM2.5 Maximum Impact Factor temperature and wind speed data integration, and then sorted out by the number of days national control point data and data from the point of PM2.5 concentration by excel software plotted to give a function
of temperature and PM2.5 concentration and wind speed and PM2.5 concentration of a function diagram, as shown in Figure 8 and Figure 9.

![Figure 8. Function diagram of temperature and PM2.5 concentration.](image)

![Figure 9. Function diagram of wind speed and PM2.5 concentration.](image)

To obtain a function of temperature and PM2.5: $y=0.1059x + 1.6908$, wind speed and PM2.5 function: $y=0.1141x - 26.025$, respectively, the self-built point temperature and wind speed data into two functions X, get a different value $y$, and then find their average value, the average value obtained is the calibration value of PM2.5, as shown in Table 8.

|   | PM2.5 | Temperature | Wind speed | Calibration value 1 | Calibration value 2 | Calibration average |
|---|-------|-------------|------------|---------------------|---------------------|---------------------|
| 1 | 50    | 15          | 0.5        | -19.4635            | -25.96795           | -22.715725          |
| 3 | 50    | 15          | 1.9        | -19.4635            | -25.80821           | -22.635855          |
| 4 | 52    | 15          | 0.4        | -19.4635            | -25.97936           | -22.71243           |
| 5 | 49    | 15          | 0.5        | -19.4635            | -25.96795           | -22.715725          |
| 6 | 49    | 15          | 1.4        | -19.4635            | -25.86526           | -22.66438           |
| 7 | 48    | 16          | 1.3        | -19.3494            | -25.87667           | -22.613035          |
| 8 | 48    | 16          | 1.3        | -19.3494            | -25.87667           | -22.613035          |
| 9 | 48    | 16          | 0.2        | -19.3494            | -26.00218           | -22.67579           |
| 10| 52    | 16          | 1.3        | -19.3494            | -25.87667           | -22.613035          |

In the same way, you can get the other “two dust four gas” calibration values for each concentration, as shown in Figure 10.
Figure 10. Calibration value of each concentration of “two dust and four gas”.

Acknowledgments
This work was supported by the Research Projects on Teaching Reform of Vocational Education in Guangxi Province in 2017 (GXGZJG2017A038).

References
[1] Information on http://cumcm.cnki.
[2] W.G.Liu, MATLAB programming and application, third ed., Higher Education Press, Beijing, 2003.
[3] S.C.Yang, J.J.Zhang, Application basis of SPSS statistical software, Guangxi Normal University Press, Guilin, 2010.
[4] Y.Zhou, Design and implementation of air quality monitoring and Analysis System, Electronics World, 18(2019), 164-165.
[5] B.N.Zhang, Z.H.Feng, F.Zhang, Z.H.Du, R.Y.Liu, Q. Zhou, Urban air quality data completion method based on spatio-temporal multi-view BP neural network, Journal of Zhejiang University (Science Edition), 6(2019), 737-744.