International Journal of Energy
ISSN: 2957-9473 | Vol. 4, No. 2, 2024

Discussion on Quality Control Method of Environmental Meteorological Data

Hua Yang

Qinghai Meteorological Information Centre, Xining, Qinghai, 810001, China

Abstract: This paper analyzes and summarizes the data format of environmental meteorological data (PM2.5 and PM10), studies quality control methods for environmental meteorological data, establishes corresponding quality control procedures, utilizes quality control procedures to control the quality of environmental meteorological data in Qinghai Province, and creates quality control documents to further improve the data quality control system in Qinghai Province and provide timely and reliable scientific data for the forecasting and warning system of environmental meteorological disasters.

Keywords: PM2.5; PM10; Quality Control.

1. Introduction

In recent years, with the rapid development of the Chinese economy, aerosol pollution in urban areas has become increasingly serious, leading to frequent occurrences of haze, heavy pollution, and low visibility events that have a significant impact on people's living environment and health. PM2.5 refers to particulate matter in the air with a diameter less than or equal to 2.5μm, also known as inhalable particles; PM10 refers to particles with a diameter below 10μm, also known as inhalable particles. The concentration of both directly affects atmospheric transparency and the degree of damage to the human respiratory system. Therefore, conducting atmospheric composition observations and strengthening the quality control of environmental meteorological observation data is of great importance.

Research by authors such as Zeng Dezhen [1], Che Xuan [2], and Yang Lijuan [3] has analyzed important factors affecting PM2.5 monitoring methods and strengthened quality control of PM2.5 monitoring through the selection of filters, sampling time, instrument maintenance, sampling locations, meteorological conditions, and sampler selection, ensuring the reliability and accuracy of monitoring data. Zhu Hualiang [4] et al. conducted quality control of observation data by analyzing the format of aerosol mass concentration observation data and performing limit value, extreme value, time consistency, and internal consistency checks.

This paper analyzes and summarizes the data format of environmental meteorological data (PM2.5 and PM10), studies quality control methods, establishes corresponding quality control procedures, utilizes quality control procedures to control the quality of environmental meteorological data in Qinghai Province, and creates quality control documents to provide timely and reliable scientific data for the forecasting and warning system of environmental meteorological disasters.

2. Data and Method

Qinghai Province has a total of eight environmental meteorological observation stations (see Table 1). This study focuses on the environmental meteorological data from the Xining station (52866). According to the quality control method of File A (surface monthly data file), the environmental meteorological data undergo checks for format, station parameters, climate threshold values, climate extremes, internal consistency, temporal consistency, and comprehensive judgment of quality control codes. The observation data covers the period from February 2018 to February 2019.

3. Data Format

The environmental meteorological data uploaded by observation stations follow a standardized naming convention. The data file naming and format comply with the meteorological industry standards “Date classifying and coding for atmospheric composition observation (QX/T 124-2011) [5]” and “Date format for atmospheric composition observation (QX/T 132-2011) [6]”.

3.1. Observation Data File Naming Rules

Z_CAWN_I_IIiii_yyyymmddhh0000_O_AER-FLD-ObservationElement-InstrumentTypeCode.txt

The specific explanations for each segment are as follows:

Z: Fixed code, representing domestic exchange data;

CAWN: Fixed code, representing atmospheric components and related data;

I: Indicates that the following IIiii field is the station number;

IIiii: Represents a 5-digit station number;

yyyymmddhh0000: A fixed-length date-time field representing the time the observation data corresponds to (UTC, Coordinated Universal Time, accurate to the hour). Where: yyyy is the year (4 digits); MM is the month (2 digits); dd is the day (2 digits); hh is the hour (2 digits); followed by 4 zero-padded digits.

O: Fixed code, representing observation data;

AER: Fixed code, representing aerosol-related observation data;

FLD: Original data observed at the observation station site;

Observation Element: Encoded as PMMUL, representing multi-element mass concentration, the data file contains mass concentrations of various particle sizes such as PM10, PM2.5. If only one observation element is present, such as PM2.5, it is encoded as PM25.

Instrument Type Code: The beta-ray type atmospheric particle concentration monitor uses the fixed code “BAM”.

txt: Fixed code, indicating a text file.

Table 1. Environmental meteorological observatory in Qinghai province

IIiii	Name of a station	Observation Elements
52737	Delingha	PM2.5
52754	Gangcha	PM2.5
52818	Golmud	PM2.5
52853	Haiyan	PM2.5
52855	Huangyuan	PM2.5
52866	Xining	PM10, PM2.5
52875	Safe and sound	PM2.5
52876	Minhe	PM2.5

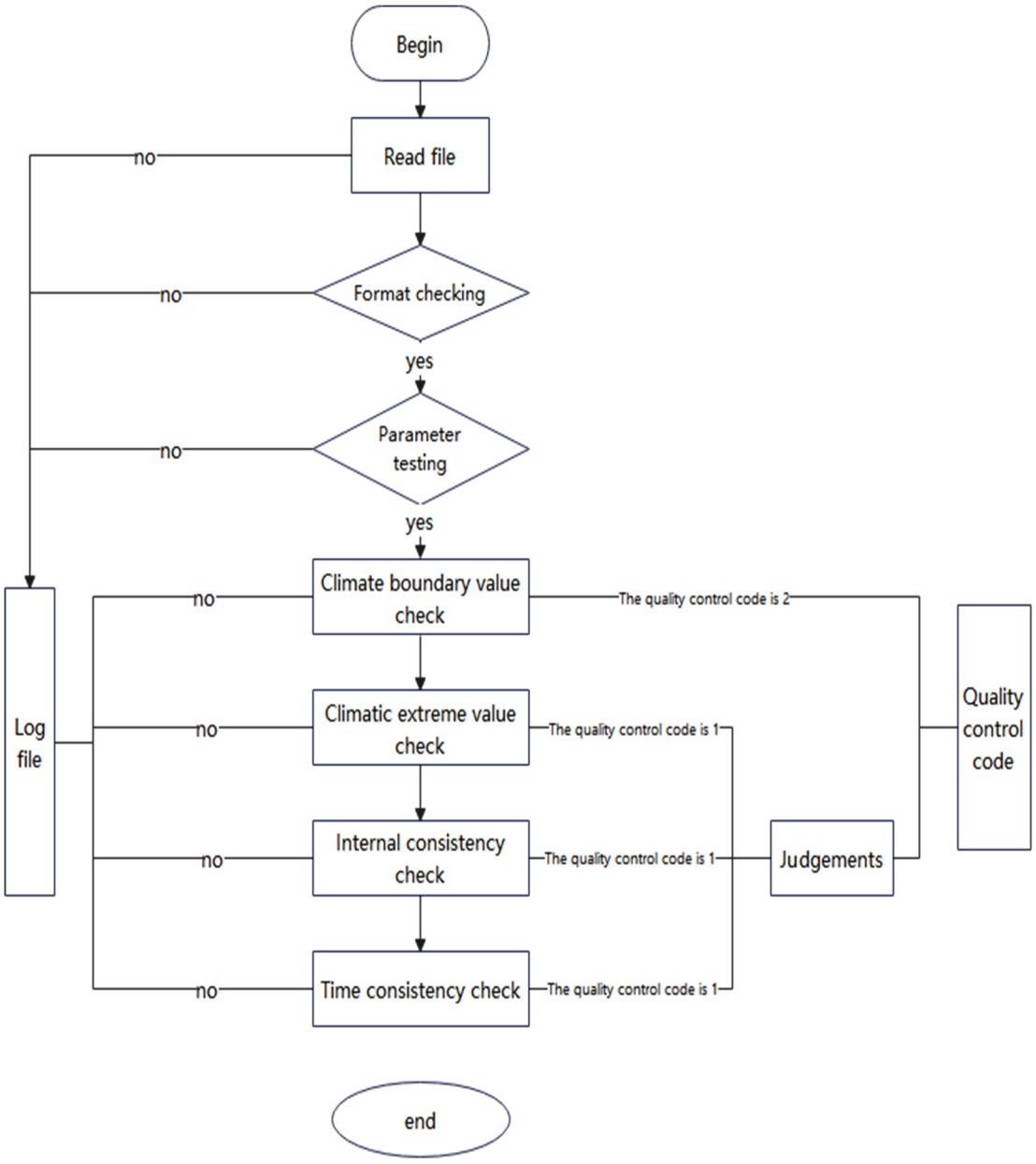


Figure 1. Flow chart of quality control of environmental meteorological observation data

3.2. Observation Data File Format

At the Xining station, the monitored elements are PM10 and PM2.5. Original data is collected every hour, generating one raw observation data record file, i.e., one data file contains 1 piece of raw data, with a total of 23 column parameters. The monitoring time is in Coordinated Universal Time (UTC), and each data record is in a separate row. Different observation data elements within each row are separated by commas using the English input method. If a measurement is missing, it is filled with a fixed value of -999.9, as shown in Table 2.

Table 2. Multi-factor PMMUL hour file data format of aerosol mass concentration data unit

Column	Parameter meaning	Data bit	Data example	Data type	Data precision	Data unit
1	Station number of observation station area	5	52866	Numbers or characters	5	Null
2	Project code	4	2225(Fixed encoding)	Integer	4	Null
3	Year	4	2019	Integer	4	Null
4	Day	3	83	Integer	#	Null
5	Time	4	0400	Integer	4	Null
6	PM10 Concentration (Standard temperature and pressure)	7	176.7	Float	One decimal place	ug/m^3
7	PM2.5 Concentration (Standard temperature and pressure)	7	11.3	Float	One decimal place	ug/m^3
8	PM1 Concentration (Standard temperature and pressure)	7	-999.9	Float	One decimal place	ug/m^3
9	PM10 Measurement data identification	1	N	Character		Null
10	PM2.5 Measurement data identification	1	N	Character		Null
11	PM1 Measurement data identification	1	F	Character		Null
12	Flow measured by PM10 instrument	4	16.8	Float	One decimal place	L/min
13	Temperature measured by PM10 instrument	5	15.1	Float	One decimal place	Degree Celsius
14	Air pressure measured by PM10 instrument	6	755.0	Float	One decimal place	kPa
15	Relative humidity measured by PM10 instrument	5	20.1	Float	One decimal place	%
16	Flow measured by PM2.5 instrument	4	16.9	Float	One decimal place	L/min
17	Temperature measured by PM2.5 instrument	5	14.9	Float	One decimal place	Degree Celsius
18	Air pressure measured by PM2.5 instrument	6	760.0	Float	One decimal place	kPa
19	Relative humidity measured by PM2.5 instrument	5	20.0	Float	One decimal place	%
20	Flow measured by PM1.0 instrument	4	-999.9	Float	One decimal place	L/min
21	Temperature measured by PM1.0 instrument	5	-999.9	Float	One decimal place	Degree Celsius
22	Air pressure measured by PM1.0 instrument	6	-999.9	Float	One decimal place	kPa
23	Relative humidity measured by PM1.0 instrument	5	-999.9	Float	One decimal place	%

4. Quality Control Process

Based on the quality control method in file A, the environmental meteorological data files undergo quality control, with a quality control code added at the end of the file. Information that does not pass the checks is written to a log file. Considering the real-time nature of the data, five quality control codes are set during the entire quality control process: 0 indicates correct data, 1 indicates suspicious data, 2 indicates erroneous data, 8 indicates missing data, and 9 indicates data that has not undergone quality control.

4.1. Content Check

Before conducting quality control on the environmental meteorological data files, a check must be made to ensure that the data file format complies with the specified standards. Any errors in the format check are written to the log file YYYYMMDD.txt. The format check primarily includes file name check, data file size check, data row and group count check, and data length and validity check.

4.2. Parameter Check

Each data record includes parameters such as regional station number, project code, year, day, hour, and minute. Three parameter checks are conducted on each data record:

(1) Verify if the regional station number matches the station number in the file name.

(2) Confirm if the observation time matches the time in the file name.

(3) Ensure that the observation project code matches the fixed code.

Records that pass the parameter check proceed to the next step of the check, while records that do not pass are logged as suspicious or erroneous information, followed by a quality check on the observation data.

4.3. Element Climate Boundary Values, Element Allowable Range Check

The climate boundary values and allowable range of elements check whether a certain element value falls within the climate boundary values and allowable range specified for that element. Climate boundary values refer to values that are impossible from a climatological perspective, while the allowable range of elements refers to the specific range within which meteorological element values are expected to occur.

Table 3. Limits and allowable range of elements in the field of element climatology

Elements	Climatological limit or allowable value of elements
Station pressure	1080hpa—400hpa
Temperature	80℃— -75℃
Relative humidity	0—100％
Maximum precipitation in one hour	400mm
Maximum maximum wind speed	75 m/s
Wind direction	0-360°
PM10	0ug/m3- 5000ug/m3
PM2.5	0ug/m3- 5000ug/m3

Elements that pass the check of climate boundary values and allowable range proceed to the next step of the inspection, while elements that do not pass are flagged with a final quality control code of 2 (error) and no further checks are carried out.

4.4. Climate Extreme Value Check

The climate extreme value check verifies whether a certain element value exceeds the historical maximum and minimum values observed for that element. Element extreme values vary with geographic region and season. The extreme value table developed in this scheme includes monthly maximum (Max) or minimum (Min) values of different elements at each station over the years. The maximum or minimum values for air pressure, temperature, and relative humidity in Table 4 are derived from Quality Control Data in file A, while the observation extreme values for PM10 and PM2.5 in Table 5 are based on historical observational data.

Table 4. Extreme value of each element in document A quality control data sheet of Xining Station in each month

Month	Atmospheric pressure /hPa		Temperature /℃		Relative humidity /%	
	Highest value	Lowest value	Highest value	Lowest value	Highest value	Lowest value
1	778.8	757.7	13.9	-23.3	88	7
2	778.5	758.9	20.8	-20.7	88	3
3	778.3	756.8	26.2	-16.9	95	3
4	782.5	756.7	31.8	-8.8	96	2
5	781.1	757.4	30.7	-2.3	95	7
6	778.0	762.1	31.9	1.6	94	15
7	772.3	762.5	36.5	4.2	98	14
8	777.6	763.0	34	3.7	97	26
9	780.9	767.3	29.2	-1.1	97	18
10	782.6	762.9	25.2	-12.5	97	12
11	779.4	764.0	19.3	-18.2	94	14
12	778.8	762.9	13.3	-23.8	91	8

Table 5. PM observation extreme values in historical data

Elements	Climatic extreme value	
	Highest value	Lowest value
PM10(ug/m^3)	533.5	26.0
PM2.5(ug/m^3)	508.3	1.0

Elements that pass the check of climate boundary values and allowable range undergo a climate extreme value check. Elements that pass the climate extreme value check are assigned a provincial quality control code indicator of 0 for that element, while elements that do not pass are assigned 1, serving as one of the criteria for setting the quality control code in comprehensive assessments.

4.5. Internal Consistency Check

The internal consistency check is based on the fact that there are varying degrees of correlation between elements measured at the same time. It detects whether there is consistency among meteorological elements that have physical relationships. The elements undergoing internal consistency check at Xining Station include air pressure, temperature, relative humidity, PM10, and PM2.5.

Elements that pass the internal consistency correlation check are assigned a provincial quality control code indicator of 0 for that element, while elements that do not pass are assigned 1, serving as one of the criteria for setting the quality control code in comprehensive assessments.

4.5.1. PM Values

In theory, PM10 values should be greater than or equal to PM2.5 values.

4.5.2. Air Pressure

The principle of the internal consistency check for air pressure is that the absolute difference between the air pressure measured by the PM10 instrument and the air pressure measured by the PM2.5 instrument should be less than or equal to a critical value, denoted as 'a'. Otherwise, the data is considered suspicious. The critical value 'a' is determined based on historical observational data. The mean absolute difference between air pressure measured by PM10 and PM2.5 instruments is 3.101hPa, with a standard deviation of 0.50hPa, and the maximum absolute difference is 5hPa. After manual review, no abnormal data was found, and a critical value of 5hPa was selected for the air pressure difference.

4.5.3. Temperature

The principle of internal consistency check for air pressure is that the absolute difference between the temperature measured by the PM10 instrument and the PM2.5 instrument is ≤ critical value 'b'; otherwise, the data is considered suspicious. The critical value 'b' is determined based on historical observational data. The mean absolute difference between the temperature measured by the PM10 instrument and the PM2.5 instrument is 0.6°C, with a standard deviation of 0.17°C and a maximum absolute difference of 1.4°C. As there were no abnormal extreme values, the critical value 'b' for temperature difference is selected as 1.4°C.

4.5.4. Relative Humidity

The principle of internal consistency check for air pressure is that the absolute difference between the relative humidity measured by the PM10 instrument and the PM2.5 instrument is ≤ critical value 'c'; otherwise, the data is considered suspicious. The critical value 'c' is determined based on historical observational data. The mean absolute difference between the relative humidity measured by the PM10 instrument and the PM2.5 instrument is 2.13%, with a standard deviation of 1.22% and a maximum absolute difference of 5.1%. Following the 3σ principle, the critical value 'c' for relative humidity difference is selected as 5.79%.

4.6. Time Consistency Check

Time consistency check is conducted on timestamp data to examine if the rate of change of the same element within a certain time interval exceeds a certain range. The elements that pass the internal consistency check undergo a time consistency check, where elements that pass are assigned a quality control code of 0 and those that do not pass are assigned a code of 1, serving as one of the criteria for setting the quality control code in comprehensive discrimination. The time consistency of each element is checked according to the following rules:

- Hourly change rate of air pressure ≤ 10 hPa/h;
- Hourly change rate of air temperature ≤ 8°C/h;
- Hourly change rate of relative humidity ≤ 50%/h;
- Hourly change rate of PM10 ≤ critical value;
- Hourly change rate of PM2.5 ≤ critical value;
- Check for any changes in 12 consecutive hours or more.

The critical values for hourly changes in air pressure, temperature, and relative humidity are based on the quality control criteria in document A, while the critical values for the hourly change rates of PM10 and PM2.5 are determined based on the original observation data. Due to factors such as rainfall, which can cause significant fluctuations in PM values over a short period, smaller critical values are selected based on the 3σ principle. However, in cases where these changes cannot be captured under special weather conditions, larger critical values for 5-minute changes are selected based on the principle that the probability of PM value changes occurring in 5 minutes is small. For specific values, refer to Table 6.

Table 6. Critical value of PM value change rate at Xining Station

Elements	Average value(ug/m^3)	Standard deviation (ug/m^3)	Critical value(ug/m^3)
PM10	34.51	35.99	142.48
PM2.5	7.57	7.92	31.33

4.7. Comprehensive Judgment of Quality Control Codes

When conducting extreme value checks, internal consistency checks, and time consistency checks on element values, the final quality control code cannot be directly assigned. It is necessary to assign the check results of the element and related elements according to certain principles. After all checks are completed, the following method is used to analyze all check results comprehensively and determine the result.

1. If an element does not pass climate extreme value checks, internal consistency checks, and time consistency checks, meaning all are assigned a quality control code of 1, then the

final quality control code for that element will be 2.

2. If an element passes all climate extreme value checks, internal consistency checks, and time consistency checks, meaning all are assigned a quality control code of 0, then the final quality control code for that element will be 0.

3. If an element does not pass one or two of the climate extreme value checks, internal consistency checks, and time consistency checks, then the final quality control code for that element will be 1.

5. Quality Control Method Implementation Rules

5.1. Quality Control File Naming

To distinguish between original data files and quality control files, the data files that have undergone quality control are renamed according to the following rule:

Z_CAWN_I_IIiii_yyyymmddhh0000_O_AER-FLD-Observation Element- Instrument Type Code-PQC.txt

5.2. Log Files

Information on suspected errors for elements that did not pass a specific test is written to a log file in the format as follows.

52737:2019032000--Row count check, The number of lines in the file doesn't equal the specified number of lines

52737:2019032000--Missing measurements in the fifth set of data

52737:2019032000--Missing measurements in the seventh set of data

...

...

5.3. Quality Control File Format

(a).Original file format is as follows:

52737,2225,2017,344,2100,-999.9,24.8,-999.9
52737,2225,2017,344,2105,-999.9,24.8,-999.9
52737,2225,2017,344,2110,-999.9,24.8,-999.9
52737,2225,2017,344,2115,-999.9,24.8,-999.9
52737,2225,2017,344,2120,-999.9,24.8,-999.9
52737,2225,2017,344,2125,-999.9,24.8,-999.9
52737,2225,2017,344,2130,-999.9,24.8,-999.9
52737,2225,2017,344,2135,-999.9,24.8,-999.9
52737,2225,2017,344,2140,-999.9,24.8,-999.9
52737,2225,2017,344,2145,-999.9,24.8,-999.9
52737,2225,2017,344,2150,-999.9,24.8,-999.9
52737,2225,2017,344,2155,-999.9,24.8,-999.9

(b).Quality control file format is as follows:

52737,2225,2017,344,2100,-999.9,24.8,-999.9
52737,2225,2017,344,2105,-999.9,24.8,-999.9
52737,2225,2017,344,2110,-999.9,24.8,-999.9
52737,2225,2017,344,2115,-999.9,24.8,-999.9
52737,2225,2017,344,2120,-999.9,24.8,-999.9
52737,2225,2017,344,2125,-999.9,24.8,-999.9
52737,2225,2017,344,2130,-999.9,24.8,-999.9
52737,2225,2017,344,2135,-999.9,24.8,-999.9
52737,2225,2017,344,2140,-999.9,24.8,-999.9
52737,2225,2017,344,2145,-999.9,24.8,-999.9
52737,2225,2017,344,2150,-999.9,24.8,-999.9
52737,2225,2017,344,2155,-999.9,24.8,-999.9
QC
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8
9,9,9,9,9,8,0,8

6. Discussion and Suggestions

With the development of society and the continuous improvement of people's living standards, the demand for environmental quality is increasing. The accuracy of monitoring data is becoming more and more important. Therefore, the quality control of PM2.5 and PM10 monitoring data at environmental meteorological observation stations is particularly important. By analyzing the data format of environmental meteorological information, using corresponding quality control procedures to control the monitoring data, and generating quality-controlled result files, it is beneficial to improve the data quality control coverage in Qinghai Province, improve the quality control system, and provide timely and reliable scientific data support for the forecasting and warning systems of environmental meteorological disasters.

References

[1] Zeng dezheng. Quality Control of PM2.5 Monitoring by Gravimetric Method [J]. Science & Technology Information, 2012 (23):413-413.

[2] Che Xuan, Liu Xiaoqiang, Song Wenbo, et al. Monitoring techniques and quality assurance/quality control measures for PM_ (2.5) [C]//Proceedings of the 6th Executive Council Meeting of Hebei Environmental Science Society and Seminar on Promoting the Implementation of Cleaner Production Promotion Law to Enhance Cleaner Production Audit Ability and Quality. 2012.

[3] Yang Lijuan. Discussion on methods to improve the quality and timeliness of atmospheric composition data upload [C]//S6 Atmospheric Composition and Weather Climate Change. 2012.

[4] Research on quality control technology of aerosol mass concentration observation data in Anhui Province.

[5] Date classifying and coding for atmospheric composition observation (QX/T 124-2011).

[6] Date format for atmospheric composition observation(QX/T 132-2011).