Introduction of $PM_{2.5}$ dataset

The dataset contains outdoor fine-grained $PM_{2.5}$ data monitored by low-cost wireless sensor network in Peking University. We use laser $PM_{2.5}$ sensors and the structure of the device is illustrated in Fig.1.

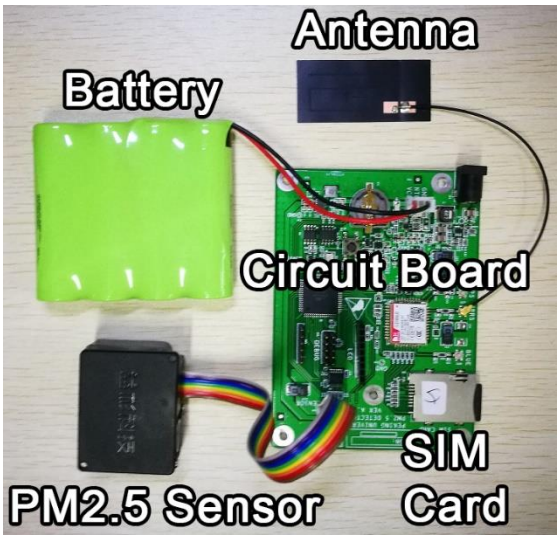

Fig.1 The structure of our device

Monitoring Time Interval:

[1] Form March 1st, 2018 to May 15th, 2018

[2] From June 3rd, 2018 to August 25th, 2018

Monitoring Frequency: 1 hour.

***.csv* files**

In the dataset, each *.csv* file contains $PM_{2.5}$ data collected at each day and there are several columns in it.

The first column records the monitoring times.

The second column contains $PM_{2.5}$ values given by the nearest official meteorological station. The distance between the official monitoring station and Peking University is about 2km.

The rest of columns contain $PM_{2.5}$ values measured by each sensor at monitoring times.

charts

Since the official station is 2km away from Peking University, small deviations between the measured data and the official data are allowed. To illustrate the deviations, we have generated a line chart for each day with both the official data and the measured data. In each line chart, the black line denotes the official data and the red line denotes the average of the measured data. The rest of the lines denotes the measured data of each sensor.

Some typical charts are shown in Fig.2. In these charts, the trends of $PM_{2.5}$ values collected by the official station and our sensors are similar.

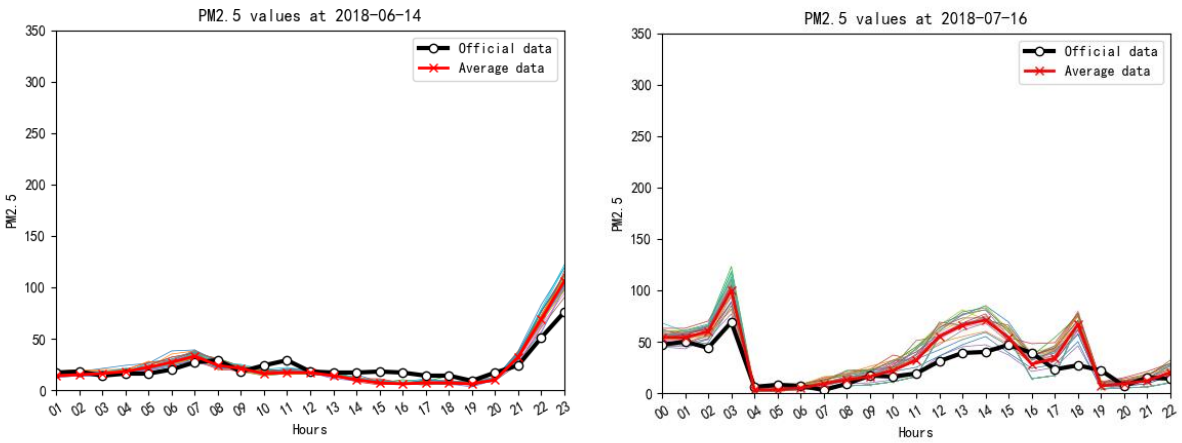

Fig.2 Charts showing similar tendencies

However, some charts show relatively big deviations between two kinds of data. Usually, one reason for the deviations is the instrumental error of laser sensors. Besides, it is believed that high $PM_{2.5}$ data may be modified by the government before published.

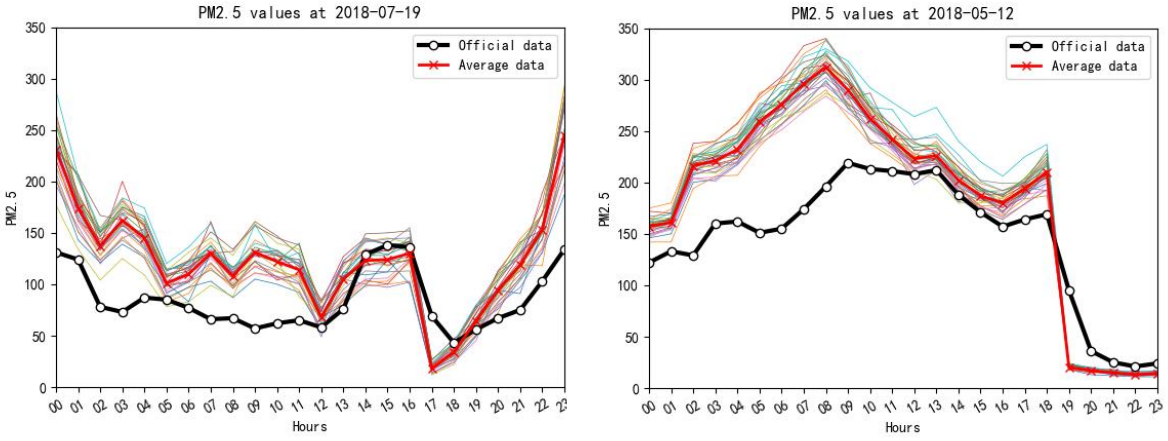

Fig.3 Charts showing abnormal measured data