A study on the application of mass smartphone barometer data in a severe convective weather in China

Tian Wang, R.Li, Q.Zhang, Lili Ding
Moji Co., Ltd, Beijing 100016, China

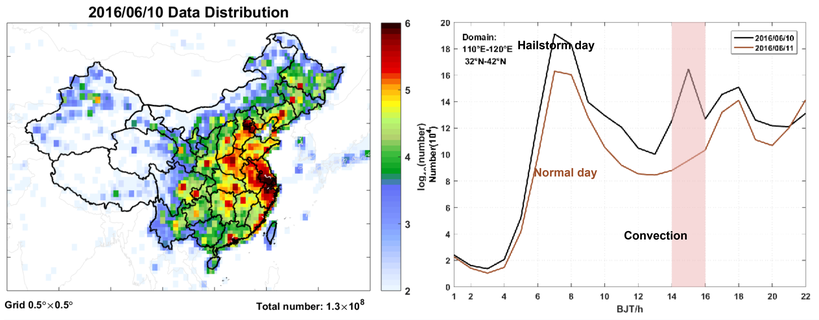
Moji Weather
www.moji.com

1. Introduction

In the age of massive data accumulation, crowdsourcing data has gradually become one of the most important data to improve weather forecasting, while the popularity and develpment of smartphones have made it an important source of crowdsourcing data. Through the MOJI APP, we can get more than tens of millions of smasrtphone barometer data every day.First, comparing the pressure data of the smartphone with the air pressure observation data of the station. It is found that the trends of the two are highly consistent. Secondly, a case of assimilation experiment was carried out, and it was found that the assimilation smartphone barometer data has certain adjustments to the forecast analysis field. And then used the machine learning method to correct the deviation of the data to obtain more accurate data.

2.Data Collection

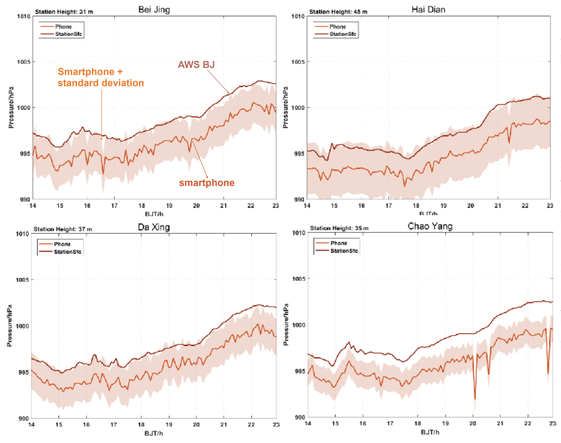
Almost every smartphone is able to capture the local surface pressure in real time. MOJI is the largest weather app service provider all over the world, with over 500 million users woldwide and over 50 million active users getting weather informations through our APP every day. Thus, we have more than tens of millions data per day.



3.Data Quality Control

Eiliminate error code → Range check → Time and space average → Regional check → Reduction to surface

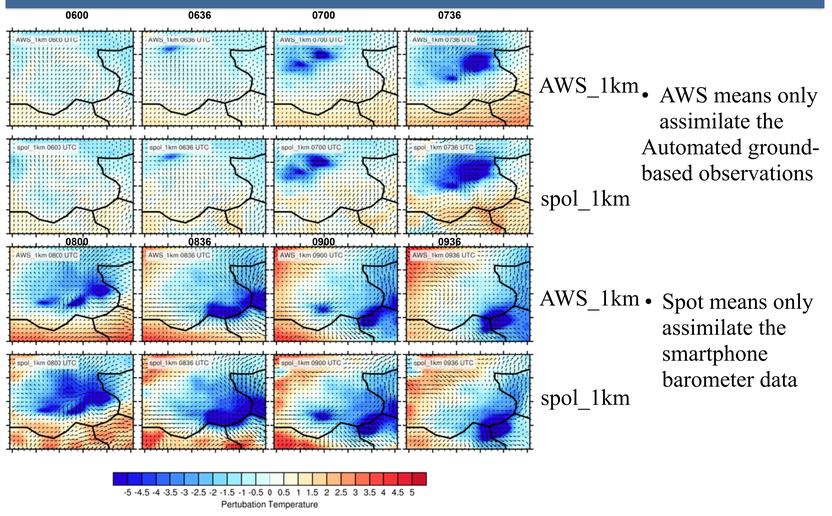
- **Range check**: Air pressure values lower than 500hPa and higher than 1080hPa were removed.
- **Regional check** (self-checking): Remove the values that were significantly different from the neighbors, those points that are greater than three times the standard deviation of positive and negative values.
- **Reduction to surface**: Using the standard deviation to capture some of the error associated with variation in terrain height inside the square grid or the possibility of smartphone data being taken from inside a tall building.



Pressure observation of AWS and smartphone (interpolated to the AWS location) 10 Jun 2016

After the quality control of the above steps, the comparison between the data from the automatic observation station and the data from the smartphone barometer shows that the two trends are highly consistent. Due to the higher resolution of smartphone data, more details of data changes can be reflected. At the same time, there is a systematic error between them, which is about 2-3hpa.

4.Assimilation



- AWS means only assimilate the Automated ground-based observations
- Spot means only assimilate the smartphone barometer data

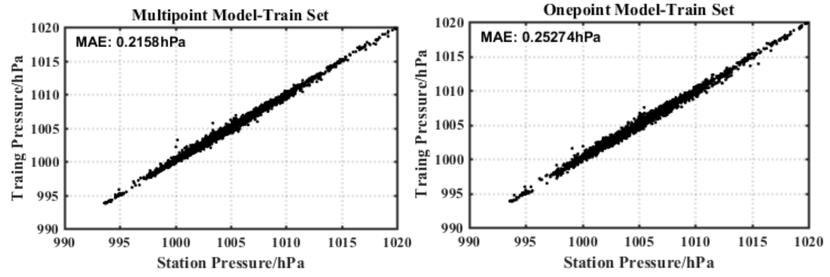
Next we use the Vdras mode for assimilation experiments. The mode resolution is 1km. It can be seen from the above pictures that the results of assimilating smartphone barometer data and the automatic observations are very similar, but the disturbance temperature and the wind field after assimilating the smartphone barometer are strengthened.

5.Bias Correction

Now we have verified the strong correlation between the smartphone barometer data and the actual weather.And the smartphone data have some effect on the assimilation model. Next, we use machine learning method to re-statistics and quality control the data, hoping to get higher quality air pressure data.

Eiliminate error code → Renge check → Time and space average → Reginal check → Statistical eliminate

- **Regional check** (self-checking): Remove the values that were significantly different from the neighbors, those points that are greater than twice times the standard deviation of positive and negative values.
- **Statistical eliminate**: For different months of data, according to the distribution of outliers to eliminate the abnormal value, so that the data basically meet the normal distribution.



Using the open data mining platform WEKA, the method of random forest is used to correct the deviation after the quality control data.

The results of the model training show that the average absolute error values of the multi-point model and the single-point model training are very close, mostly less than 0.5 hPa.

6. Conclusion & Future Work

The conclusions and further plans are as follows:

1. The smartphone data can be used in assimilation, and may improve model analysis if the quality of smartphone pressures is improved.
2. Machine learning has a significant effect on the correction of the smartphone barometer data, and the multi-point model is close to the average absolute value error of the single-point model training.
3. We will try to use the data to our nowcasting which is also based on machine learning. After this it also can be used as real-time data of user's position in MOJI APP, so that the user can see more accurate and real-time data.