

SHORT-TERM PRECIPITATION OCCURRENCE PREDICTION FOR STRONG CONVECTIVE WEATHER USING FY2-G SATELLITE DATA: A CASE STUDY OF SHENZHEN, SOUTH CHINA

Kai Chen^a, Jun Liu^{a,*}, Shanxin Guo^a, Jinsong Chen^a, Ping Liu^a, Jing Qian^a, Huijuan Chen^a, Bo Sun^a
^aShenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 518055, Shenzhen China
(kai.chen, jun.liu, js.chen, liuping, jing.qian, hj.chen, sunbo)@siat.ac.cn

Youth Forum

KEY WORDS: Short-term precipitation, SVM, FY-2G satellite image, Precipitation occurrence prediction

ABSTRACT:

Short-term precipitation commonly occurs in south part of China, which brings intensive precipitation in local region for very short time. Massive water would cause the intensive flood inside of city when precipitation amount beyond the capacity of city drainage system. Thousands people's life could be influenced by those short-term disasters and the higher city managements are required to facing these challenges. How to predict the occurrence of heavy precipitation accurately is one of the worthwhile scientific questions in meteorology. According to recent studies, the accuracy of short-term precipitation prediction based on numerical simulation model still remains low reliability, in some area where lack of local observations, the accuracy may be as low as 10%. The methodology for short term precipitation occurrence prediction still remains a challenge. In this paper, a machine learning method based on SVM was presented to predict short-term precipitation occurrence by using FY2-G satellite imagery and ground in situ observation data. The results were validated by traditional TS score which commonly used in evaluation of weather prediction. The results indicate that the proposed algorithm can present overall accuracy up to 90% for one-hour to six-hour forecast. The result implies the prediction accuracy could be improved by using machine learning method combining with satellite image. This prediction model can be further used to evaluated to predicted other characteristics of weather in Shenzhen in future.

1. INTRODUCTION

Short-term precipitation nowcasting usually refers to 0-6 hours (mainly focus on 0-2 hours) rainfall event occurrence with high spatial and temporal resolution (Yu, 2012). Nowadays, radar echo extrapolation and fine numerical forecast model are the main technology commonly used to predict short-term precipitation (Zou, 2014). However, the poor accuracy of extrapolation in radar echo inversion significantly affects the accuracy of forecasting. Only using radar extrapolation will result in larger prediction error and limited forecast period, and also the forecast accuracy will be decline rapidly with the extrapolation time increasing (Feng, 2013; Atencia, 2010). Due to the poor spatial resolution of radar data, poor accuracy remains in precipitation forecasting in local scale (Casati, 2004). Commonly the smaller spatial scale is, the higher the error remains (Hu, 2015). During the 2008 Beijing Olympic Games, Meteorological Agency diagnosed strong convective weather through expert systems (B08FDP) which has a certain indicative forecasting (James, 2010). Its 30 minutes and 60 minutes of QPF (quantitative precipitation forecast) and TS scores were close to 0.3 and 0.2 respectively. However, the random forecast hit rate still remains low (Chen, 2010). Shenzhen Meteorological Agency has started to forecast long-term heavy storm for city area since 2012. But until now, there are still some limitation in quantitatively short-term rainfall forecast or other small probability events. The forecast TS score within 1 hour and 2 hours are below 10% and 3% respectively (Zhang, 2015).

In this study, we use multi-spectral satellite data of 1-hour time interval covers the cloud reflectance in visible, near-infrared and infrared wavelength in electromagnetic spectrum. The

prediction model also combines with hourly data in situ automatic meteorological station in Shenzhen area. The dynamic SVM regression was used to predict precipitation occurrence based on nearby observations in time series. The main objective of this study is to prove that the prediction accuracy of precipitation occurrence can be improved by using machine learning technology with high time frequency satellite imagery.

2. STUDY AREA AND DATA OVERVIEW

Shenzhen is the fourth economic city located in the South China coast with abundant precipitation from April to October. Extreme weather occurs frequently and trend to appear localized heavy rainfall and other weather anomalies as the global climate changing. Localized heavy storm was observed from in situ observations. Different weather could appear within the Shenzhen city, for example, a heavy rainfall occurs in Luohu district while the clear sky appears in Nanshan District. Furthermore, the localization was also observed in time scales of the heavy storm. The rapid changing in both spatial and time domain shows a big challenge for traditional numerical model of nowcasting.

The selected automatic meteorological stations distributed in every district area in Shenzhen are shown in Figure 1. The average distance between two stations is 3.8km. The observation covers precipitation, wind speed, wind direction, 2m local temperature, and air pressure with 1h interval.

Satellite image from FY-2G was used to capture the spatial variation of cloud characteristics with 1-hour time interval. FY-2G is the 8th operational satellites of Fengyun II geostationary

* Corresponding author

meteorological satellite series developed by China. The satellite mainly covers the cloud reflectance in visible, near-infrared and infrared wavelength in electromagnetic spectrum. Its scanning radiometer have five channels which has two long-wave infrared band, one medium-wave infrared band, one visible band and one band in water vapour absorption region. FY-2G IR stray light suppression effect increased more than 50%, the frequency and capabilities of satellite calibration also increased. These technical improvements effectively improved the inversion accuracy of FY-2G quantitative products and the quantitative application level. It can obtain one-third of the earth's surface image in a non-flood season every hour and in a flood season every half hour. In this study, Many cloud parameters include Top Brightness Temperature(TB), Cloud Top Temperature, Gradient of the pixel TB, Difference of TB, Cloud Total Amount, Cloud Type, Middle and Upper Tropospheric Water Vapour Content were used as model input to maximally capture the characteristic of precipitation cloud.

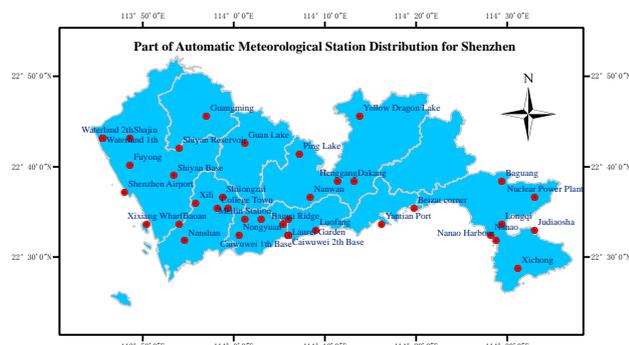


Figure 1. Spatial locations of some automatic meteorological stations in Shenzhen

3. STUDY METHODS

3.1 Data Pre-processing

Firstly, FY-2G satellite level 1 images need to be pre-processed, including radiometric correction and geometric correction. Geometric correction is mainly to unify FY satellite imagery data to the same geographic coordinate system based on ground control points and satellite orbit data. After geometric correction, all the images can be overlaid with other auxiliary data to analysis the tracking of cloud development. Radiometric correction is to remove the influence by different atmospheric environment on the image through using auxiliary radiation data of satellite, thereby enhancing the ability to obtain quantitative information from images. The proposed model consists of three steps: the first step is the cloud parameters estimation in FY satellite data by cloud recognition. The cloud classification and parameters estimation were implemented according to the texture, shape, smoothness and brightness temperature, and the parameters including cloud type, cloud-top temperature, top brightness temperature(TB), Gradient of the pixel TB, Difference of TB, are obtained. The following Table 1 shows the specific selected parameters used in this study.

Unlike satellite data pre-processing, ground data was mainly checked in the quality of automatic station data, include obvious data missing and errors. Data selection based on choosing samples with good quality to ensure the accuracy of model output in this study. Since the FY2G images spatial resolution is 5 km x 5 km, however, the automatic station is fixed, so we use nearest neighbour matching principle to select image pixel which corresponding to the specific automatic station location.

Table 1. FY-2G satellite inversion parameters list

Parameters	Computing methods	Parameter descriptions	Spatial resolution	Time resolution
Top brightness temperature (TB)	Calculated based on the FY-2G IR1-4 band and radiometric calibration lookup table.	This parameter reflects the brightness temperature of underlying surface satellite observed	5km	0.5h
Cloud Top Temperature (CTT)	Calculated by the FY-2G thermal infrared band (10.3-11.3 μ m) and radiometric calibration lookup table.	This parameter reflects the cloud top k's temperature of satellites observed	5km	0.5h
Gradient of the pixel TB (GT)	$GT_{(x,y)} = \left[(TB_{(x+1,y)} - TB_{(x-1,y)})^2 - (TB_{(x,y+1)} - TB_{(x,y-1)})^2 \right]^{1/2}$	This parameter reflects the gradient changes of FY-2G IR1-4 band brightness temperature	5km	0.5h
Difference of TB (DT)	$DT_{ab(x,y)} = TB_{a(x,y)} - TB_{b(x,y)}$	$TB_{a(x,y)}$, $TB_{b(x,y)}$ represent the brightness temperature of the same object in different bands respectively	5km	0.5h
Cloud Total Amount (CTA)	$A_c = (I - I_{clr}) / (I - I_{cld})$	A_c = Cloud Total Amount I_{clr} = Clear sky pixel radiation I_{cld} = Clouds pixel radiation	30km	1h
Cloud Type (CT)	Get through cluster analysis by emissivity differences of different types clouds in the infrared and water vapour channels	Pixel types include earth's surface, low clouds, altostratus, cirrostratus, dense cirrus, cumulonimbus, etc.	5km	1h
Middle and Upper Tropospheric Water Vapour Content (WVC)	Inversion from the water vapour channel observation data and the relative humidity in the radiative transfer model	This value represents the relative humidity of effective area, and the data is susceptible to digital simulation accuracy	5km	3h

3.2 Multi-Source Data Modelling

According to the characteristic and advantage of SVM dealing with small sample size, in this study, a multi-time scale multi-source data SVM model was built to capture the data variation in small time window. The input of model includes parameters from FY-2G images, i.e., Top Brightness Temperature (TB), Cloud Top Temperature (CTT), Gradient of the pixel TB (GT), Difference of TB (DT), Cloud Total Amount (CTA), Cloud Type (CT), Middle and Upper Tropospheric Water Vapour Content (WVC), and from ground automatic station, i.e., Wind Speed, Humidity, Temperature, Air Pressure. The output of the model is precipitation occurrence in which “0” represents no rain and “1” represents rain occurs. The basic assumption is: in time series for each location, the rainfall occurrence is highly related to the nearby historical meteorological conditions in this time series. The influence of historical meteorological conditions which farther away from the current prediction window can be regarded as model variance that can be neglected in real-time prediction. Based on this idea, the precipitation occurrence can be predicted by analysis the nearby last-minute meteorological conditions. The main idea of precipitation occurrence prediction was shown in Figure 2.

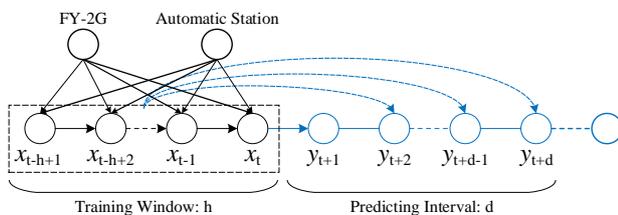


Figure 2. Basic idea of precipitation occurrence prediction for future d hours based on nearby weather conditions

SVM algorithm with radial basis function (RBF) was chosen in this study. The RBF function is as following:

$$K(x, x_i) = \exp\left\{-\frac{|x-x_i|^2}{\sigma^2}\right\} \quad (1)$$

Multi-time scales SVM rainfall prediction based on following equations:

$$Y_t = R^*(X_t) \quad (2)$$

$$X_t = [x_{t-h-d+1}, x_{t-h-d+1}, \dots, x_{t-d}]^T \quad (3)$$

$$Y_t = [y_{t-h+1}, y_{t-h+2}, \dots, y_t]^T \quad (4)$$

$$y_{t+1} = R^*(x_{t-d+1}) \quad (5)$$

where $d \in [0,6]$
 $h=2,4,8,12,24,48,72$
 X_{t-i}, Y_{t-i} = training set
 R^* =SVM model created by training set
 t = current time
 h = training sample size (or time scales)
 d = prediction time interval
 y_{t+i} = rainfall value to be predicted
 x_{t-d+i} = model input of rainfall value to be predicted

Here the time windows of training data were set as: 2h, 4h, 8h, 12h, 24h, 48h, 72h, which means for each prediction, 2h, 4h, 8h, 12h, 24h, 48h, 72h historical data were used to predict future rainfall occurrence. Meanwhile, The prediction intervals were

also set as: 0h, 1h, 2h, 3h, 4h, 5h, 6h, which means the different prediction windows used for occurrence estimation.

3.3 Validation

The traditional way to evaluate the forecast accuracy in the short-term weather forecasting is TS score which can be calculated as following equation.

$$TS_k = \frac{NA_k}{NA_k + NB_k + NC_k} \times 100\% \quad (6)$$

where NA_k = the number of correct prediction
 NB_k = the number of more prediction
 NC_k = the number of less prediction

The validation table is shown in Table 2, which can also be used to validate the output of prediction.

Prediction \ True	Rain	NoRain
Rain	NA_k	NC_k
NoRain	NB_k	—

Table 2. Classification of rainfall forecast test

4. RESULTS AND DISCUSSION

4.1 FY-2G Data Product

According to the statistical analysis and precipitation mechanism analysis, the cloud parameters variation could have correlation with rainfall event. Figure 3 shows an example about the cloud top temperature, the total amount of cloud, the type of cloud, upper tropospheric water vapour content of FY-2G parameters at 23:00 on June 28, 2015. Figure 4 shows an example about the brightness temperature, the channel 1 and Channel 4 brightness temperature difference, brightness temperature gradient, channel 3 and Channel 4 brightness temperature difference of FY-2G parameters at the same time with Figure 3.

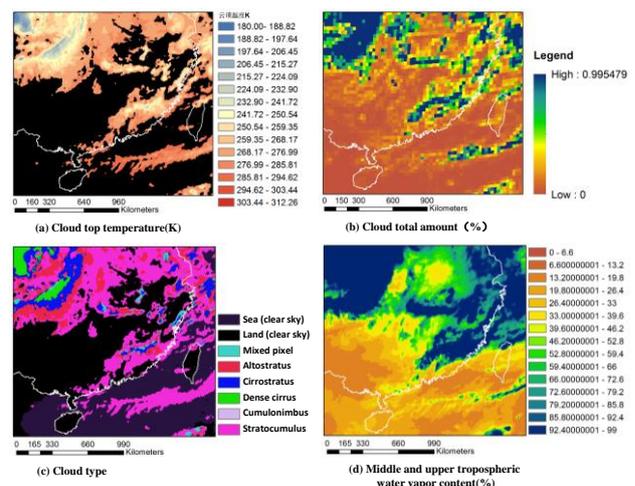


Figure 3. FY-2G parameters, including (a) cloud top temperature, (b) the total amount of cloud, (c) the type of cloud, (d) upper tropospheric water vapor content,

at 23:00 on June 28, 2015

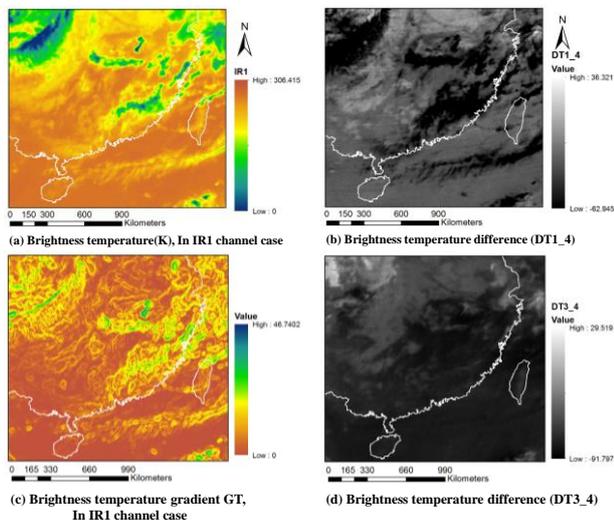


Figure 4. FY-2G parameters, including (a) brightness temperature, (b) channel 1 and Channel 4 brightness temperature difference, (c) brightness temperature gradient, (d) channel 3 and Channel 4 brightness temperature difference, at 23:00 on June 28, 2015

4.2 Precipitation Occurrence Prediction

Table 3 shows the TS scores predicted by 2 hours training window. As table 3 shows, the prediction performance better and the TS score is relatively high in the most of station compare to the studies present by Zhang Lei, 2015. Prediction accuracy varies in different stations. One reason is that the precipitation in the specific station could be affected by different factors with different influence degree. It's clear in table 3 that the TS scores reduced as the prediction window size increase. This follows the common sense of rainfall prediction.

Station Interval	Xili	Dakang	Nanshan	Nanwan
1 hour	42.63%	44.18%	41.77%	41.81%
2 hours	33.33%	34.63%	33.60%	32.74%
3 hours	26.74%	28.20%	27.22%	26.82%
4 hours	23.06%	25.36%	25.38%	23.85%
5 hours	20.09%	22.33%	21.66%	21.67%
6 hours	19.15%	20.25%	19.69%	19.42%

Table 3. TS scores of 2 hours time scale SVM for future different hours rainfall prediction

Figure 5 shows the overall accuracy of occurrence prediction with 2 hours training window. All in situ weather stations are employed to evaluate the accuracy. The overall accuracy represents the ratio between accurate prediction times of rainfall and no rainfall and the total number of times. The result shows that the overall prediction accuracy of each station are relatively high enough to be practical applied. Different colour represents the prediction within 1 hour to 6 hours prediction in the future. The overall prediction accuracy of most stations are more than 87.5%. The overall accuracy of different prediction intervals shows a little difference. Among them, the accuracy of Henggang station is above 94%, the accuracy of Judiaosha station is relatively poor and with only 91%. These results prove the effectiveness of the proposed method.

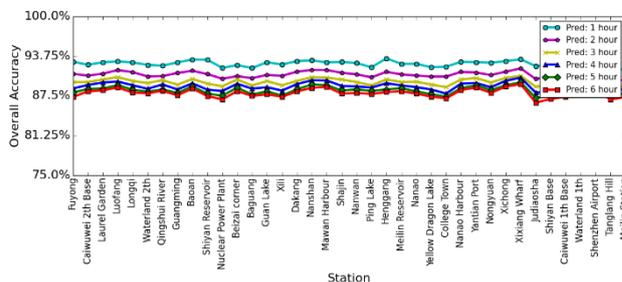


Figure 5. The overall accuracy of 2 hours time scale SVM for future different hours rainfall prediction

Figure 6 and 7 show the TS scores of different sizes of training window (i.e., time scale) with SVM for precipitation occurrence prediction. Obviously, 2 hours time scale gets the highest TS score and also obtains the highest prediction accuracy. This can be seen as the evidence of our basic assumption that the current rainfall has more likely correlated to the nearby weather conditions in time series. Secondly, TS scores of 4 hours, 8 hours, 12 hours, 24 hours, 48 hours, 72 hours time scale presents the trends that the prediction results decrease as the increase of prediction window size. This shows that with the training samples increases, many meteorological factors unrelated to future rainfall were introduced in the model which brings the increasing uncertainties to the model.

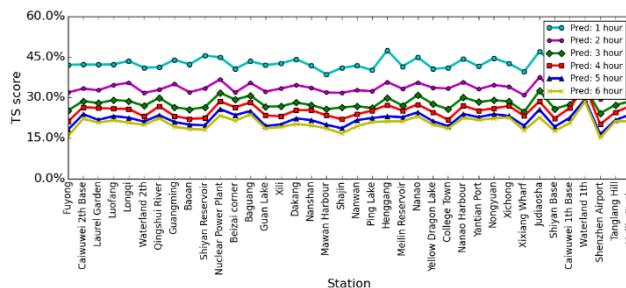


Figure 6. TS scores of 2 hours time scale SVM for future different hours rainfall prediction

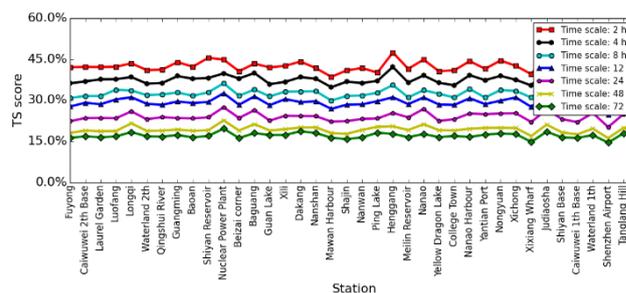


Figure 7. TS scores of different time scale SVM for future 2 hours precipitation prediction

The stability of the model will affect the sustainability and applicability of prediction. Therefore, establishing a robust model is very important. The standard deviation of TS scores was calculated within different time scale for different prediction window. The smaller the standard deviation is, the more stability the model remains. On the other side, the standard deviation of TS score also reflects the discrete trends as the prediction window increase. A smaller standard deviation represents the majority of stations TS scores are close to the overall mean. Figure 8 shows the standard deviation of TS scores.

It can be seen from Figure 8 that the model established in this paper have good stability. The results are not affected by the location and environmental conditions of stations. They have shown a good and stable prediction.

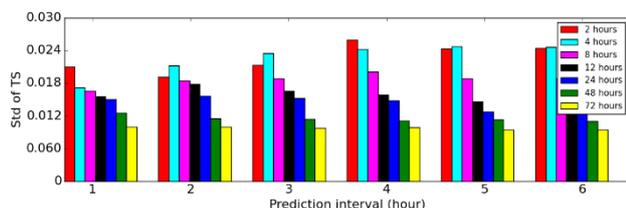


Figure 8. The standard deviation of TS scores about different hours time scale SVM for all station

5. CONCLUSIONS

In this study, we combined FY-2G satellite data and in situ automatic meteorological station hourly observations to establish a precipitation occurrence prediction model. Based on multi-source data assimilation analysis, the results prove the effectiveness of the proposed SVM-based method in short-term rainfall forecasting. The mainly conclusion of this study are as follows:

The study proposed a finer and higher density precipitation forecast method on geospatial through combination of the FY-2G satellite and the shenzhen automatic meteorological station data. This method is a attempt to try to solve the limitations of current short-term rainfall forecast of shenzhen. It's also useful for frequent small areas rainfall anomalies and local short-term rainfall in shenzhen.

The prediction model also has certain practical value for the next six hours of rainfall forecast. In Multi-time scales SVM rainfall prediction modelling, the greater the time scale is, the worse the results will be predicted by the model, and vice versa. That means current rainfall was more relevant to the recently history meteorological conditions. And the smaller time scale is, the lower calculations will be.

The next study will focus on continuing optimizing the proposed forecast method, and employing more meteorological parameters. All of these work are expected to further improve the accuracy of rainfall forecast.

ACKNOWLEDGEMENTS

This work is jointly supported by the National Natural Science Foundation program (No. 41301403, 41471340), the Basic Research Program of Shenzhen (No. JCYJ20150831194441446, JCYJ20150630114942312 and JCYJ20150630114942260), the program of Shenzhen key laboratory of severe weather in south China (ZDSYS20140715153957030 and SY5Z0211001) and the program of Guangdong Provincial Science and Technology Department (20140218).

REFERENCES

Yu X.D., et al., 2012. The Advances in the Nowcasting Techniques on Thunderstorms And Severe Convection. *Acta Meteorologica Sinica*, (03), pp. 311-337.

Zou D.L., et al., 2014. Study of 0~3 Hour Short-Term Forecasting Algorithm for Rainfall. *Journal of Tropical Meteorology*, (02), pp. 249-260.

Feng Y.R., et al., 2013. A 0~6h Quantitative Snow(Rain) Forecast Technique and Its Application in Vancouver Winter Olympics. *Guangdong Meteorology*, (01), pp. 6-13.

Atencia A, et al., 2010. Improving QPF by blending techniques at the Meteorological Service of Catalonia. *Natural Hazards and Earth System Sciences*, 2010, Vol. 10, pp. 1443-1455.

Casati B., et al., 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11(02), pp. 141-154.

Hu C.H., et al., 2015. Verification of Quantitative Precipitation Forecast Between Radar and Numerical Model Based on Intensity-Scale Method. *Journal of Tropical Meteorology*, (02), pp. 273-279.

James W. W., et al., 2010. Nowcasting Challenges during the Beijing Olympics: Successes, Failures, and Implications for Future Nowcasting Systems. *Weather and Forecasting*, 25(6), pp. 1691-1714.

Chen M.X., et al., 2010. Introduction of Auto-nowcasting System for Convective Storm and Its Performance in Beijing Olympics Meteorological Service. *Journal of Applied Meteorological Science*, (04), pp. 395-404.

Zhang L., et al., 2015. Discussion of Relative Accuracy of Short-Range Heavy Rain Nowcasting. *Guangdong Meteorology*, (02), pp. 1-6.