Notice of Retraction

After careful and considered review of the content of this paper by a duly constituted committee, this paper has been found to be in violation of IEEE's Publication Principles. IEEE hereby retracts the content of this paper. Reasonable effort should be made to remove all past and future references to this paper.

Short-term and local rainfall probability prediction based on a dislocation support vector machine model using satellite and in-situ observational data

Xunlai Chen, Guangjun He, Yuanzhao Chen, Shuting Zhang, Jinsong Chen, Jing Qian, and Haicong Yu*

***Abstract*—Short-term and local rainfall commonly occurs in southern China, and can result in intensive precipitation in the local region over a very short time. Considerable precipitation could cause an intense flood within a city when the runoff exceeds the capacity of a city drainage system. Predicting rainfall probability accurately is among the important scientific problems in meteorology and remains a challenge. In this study, a dislocation machine learning method based on a support vector machine was used to predict short-term and local rainfall probability using Feng Yun II geostationary meteorological satellite imagery and ground in situ observational data. The results were validated by a traditional threat score, which is commonly used in evaluation of weather predictions. The results indicate that the proposed algorithm can achieve good performance for a 1-h to 6-h forecast. The result implies that prediction accuracy could be improved by using a machine learning method combined with satellite imagery.**

***Index Terms*—Short-term and local rainfall; rainfall probability; dislocation SVM; FY-2G satellite image; TS score evaluation**

This work was supported by the National Natural Science Foundation (91544102), Fundamental Research Foundation of Shenzhen Technology and Innovation Council (JCYJ20160429191127529), the CAS "Light of West China" Program (2016-QNXZ-A-5), Science and Technology Planning Project of Guangdong Province (2017A050 501027), Shenzhen International S&T Cooperation Project (GJHZ20160229194322570).

X. C. is with Shenzhen Meteorological Bureau, and Shenzhen Key laboratory of severe weather in south China, Qixiang Road, Zhuzilin, Futian, Shenzhen, P.R.China. (e-mail: cxlxun@163.com).

G. H. is with State Key Laboratory of Space-Ground Integrated Information Technology, Space Star Technology CO., Ltd., 61rd Yard, Zhichun Road, Haidian District, Beijing, P.R. China, (e-mail: hgjun_2006@163.com).

Y. C is with Shenzhen Meteorological Bureau, and Shenzhen Key laboratory of severe weather in south China, Qixiang Road, Zhuzilin, Futian, Shenzhen, P.R.China. (e-mail: chenyuanzhao@szmb.gov.cn).

S. Z is with Shenzhen Meteorological Bureau, and Shenzhen Key laboratory of severe weather in south China, Qixiang Road, Zhuzilin, Futian, Shenzhen, P.R.China. (e-mail: zhangshuting@szmb.gov.cn).

J. C. is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Xueyuan avenue 1068, Shenzhen University Town, Shenzhen, P.R.China. (e-mail: js.chen@siat.ac.cn).

J. Q. is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Xueyuan avenue 1068, Shenzhen University Town, Shenzhen, P.R.China. (e-mail: jing.qian@siat.ac.cn).

H. Y. was with Center for Assessment and Development of Real Estate; 8007 Honglixi Road, Futian, Shenzhen, P.R.China. (e-mail: haicong_yu@163.com).

*corresponding author.

I. INTRODUCTION

Short-term and local rainfall refers to a common weather phenomenon in which the precipitation intensity is relatively large and the rainfall reaches or exceeds a certain standard within a certain period and local region, such as convective weather[1-3]. In the Pearl River Delta region of China, convective weather is particularly obvious. The Pearl River Delta is located in the low latitude and subtropical regions. It has abundant climatic resources. Solar energy resources, thermal resources and water resources are among the highest in China, but it is also a convective weather-prone area. In spring, there are often low-temperature and rainy, strong convection, spring drought, etc. There may also be a cold wave. The summer is affected by weather systems such as low-lying frontal grooves, tropical cyclones, and monsoon clouds, with heavy rain, thunderstorms, and typhoons. These weather phenomena sometimes occur together with several phenomena, sometimes alone. They have small horizontal scales, short life history, and sudden bursts. Strong convective weather occurs in small and medium-scale weather systems with small spatial scales [4]. The general horizontal range is about a dozen to two or three hundred kilometers, but some horizontal ranges are only tens of meters to ten kilometers. Its life history is short, about one hour to ten hours, and the short ones are only a few minutes to one hour, with obvious suddenness, which may cause heavy rains and other disasters such as urban stagnation and dangerous slope landslides. Gale and heavy rain caused by typhoons can also cause geological disasters and cause large-scale flooding [5].

At present, the forecast of the weather system mainly relies on numerical weather prediction model, which has a good forecasting effect on large-scale weather systems, but shows a large limitation for the forecast of strong convective weather at small and medium scales [6]. Although the current numerical weather prediction model has developed vigorously, its ability to forecast abnormally complex weather changes is still very limited [7]. This is mainly because the current comprehensive understanding of the laws of physics involved in weather changes is not enough, and the level of atmospheric observation and computational techniques are not yet complete. Therefore, the forecast of short-term precipitation and dramatic changes in

the local regions remains need to improve [8]. The nowcasting method based on radar observation data and the rapid assimilation mesoscale forecasting system are the key technologies that constitute the 0-2 hour short-term nowcasting system [9]. Although the forecasting technology based on radar data for identification and extrapolation has been widely used in many nowcasting operational systems, its forecasting time is short, and the forecasting results often have a certain gap with the actual observation, which cannot accurately provide the evolutionary trend of strong convective weather. In addition, the current methods for analyzing the development of radar echoes are based on the reflectivity factor [10]. The dynamics of radar echoes do not fully utilize the velocity field data of Doppler weather radars, and usually do not take into account the evolution of echoes, resulting in large errors in the forecasting.

For the above reasons, there are many difficulties in the short term and local rainfall prediction. For convective weather, its coverage is small and rapid development, so that under the current observation conditions, not only for China, including the rest of the world, its information capture is not comprehensive enough. Hence, there are still some limitations in the understanding of this weather phenomenon. According to the China National Meteorological Administration, based on the numerical weather prediction system and assimilation of various meteorological data and radar echo data, the accuracy of 24-hour rainfall forecast can reach more than 87%. This is a forecast on large spatial scales and long-term scales, but numerical weather prediction model has large limitations on small-scale and short-term scale predictions. There are only a few studies and there is much room for improvement. The forecast TS scores within 1 h and 2 h are less than 10% and 3%, respectively [11]. Therefore, for urban meteorological forecasting, it is particularly important to forecast rainfall at regular, fixed and quantitative levels.

Machine learning technology has made tremendous progress in various fields [12-14]. For the application of machine learning algorithms in weather forecasting, single modeling is a common idea by finding a large and general forecasting model. While in this study, each in-situ observation station is dynamically modeled separately, taking into account the differences in meteorological elements at different sites and at different times. It can solve the problem that single modeling pays too much attention to the overall law to establish a fixed overall predictive function model while ignore the meteorology changes at different sites and different moments. At the same time, the input of the single modeling model is the time series meteorological condition data and the corresponding time series rainfall data, and the output is the future time series rain. In the prediction, because the future meteorological conditions are unknown, in order to obtain the rainfall prediction, it is necessary to first predict the future meteorological conditions (for example, first predicting the future temperature, wind, etc. by numerical weather prediction model). Due to the inherent errors and uncertainties in predicted meteorological conditions, rainfall predictions are also affected by these errors.

This paper proposes a new modeling method based on the perception that limited by the local environmental factors of the city, the meteorological changes at different sites and times are different. For each time rainfall forecast, it is more relevant with the meteorological conditions in the previous period. This is the starting point of this paper. Based on the actual demand for short-term forecasting of higher precision precipitation in Shenzhen, this paper uses support vector machine (SVM) which can solve the nonlinear problems in pattern recognition, classification and prediction, meteorological element data and Fengyun meteorological satellite data in Shenzhen to propose a new modeling method named dynamic dislocation SVM based short-term local rainfall prediction model. The multi-spectral satellite data of a 1-h time interval covering cloud reflectance in the visible, near-infrared, and infrared (IR) wavelengths in an electromagnetic spectrum are used. The main objective of this study was to prove that the prediction accuracy of rainfall probability can be improved by using machine learning technology with high-time-frequency satellite imagery. It fully considers the differences in meteorological elements at different sites and times, and attempts to improve the accuracy of short-term local rainfall forecasts. Compared with the existing numerical weather prediction method, this method predicts from a pure data point of view with a simplified idea, and has more computational cost advantages than the complex numerical weather prediction method.

A preliminary version of this manuscript was published in ISPRS Congress 2016 previously [15]. In this paper, more details about the proposed dynamic dislocation SVM model are explained and the differences with and without dislocation model are compared. In addition, the proposed model is also further compared with some existing machine learning methods to validate the effectiveness of prediction performance.

The remainder of this paper is organized as follows. Section II describes the related works. The study area, the FY data pre-processing and the proposed dislocating SVM model are presented in section III. Section IV are the experiments and discussion, and section V provides the conclusions.

II. RELATED WORK

Improving rainfall probability prediction accuracy is an important issue in weather forecasting. Because rainfall results from the interaction of many large-scale weather systems are with very complex formation mechanisms, and its properties are significantly nonlinear and vary with time, traditional statistical approaches have difficulty in identifying its change law [16]. During the past few decades, Kalman filter [17], neural network [18], and other new techniques have been used to improve the prediction accuracy.

Current research regarding rainfall mainly focuses on long-term rain and its impacts at a large scale. The time series of rain attenuation of a Ku-band signal during rain events at a tropical location was predicted by predicting the fade slope from a modified Van de Kamp (VDK) model [19]. Li et al. assessed the yield forecast performance with increasing nitrogen input to determine when the acceptable predicted yield could be achieved using the CERES-Wheat model [20]. The performance of dynamical seasonal forecast systems was evaluated for the prediction of short-term rainfall anomalies

over equatorial East Africa, based on observational datasets and the Asia-Pacific Climate Center (APCC) Ocean–Atmosphere coupled multi-model ensemble (MME) retrospective forecasts (hindcasts) [21]. Nicholson predicted monthly rainfall during the "Long Rains" over the Greater Horn of Africa using 1- and 2-month time series and the results were evaluated via cross validation [22]. Sharan presented work regarding the prediction of rainfall based on an average of three methods where the historical rain data of Bihar were selected for projection and found that the predicted results were quite close to the actual rainfall data for 2013 [23]. Many other researchers have shown that rain is predictable and this information can be used in many applications [24, 25].

However, regarding short-term and local rainfall prediction, less research has been conducted because it is more difficult to predict. Rainfall probability prediction is among the key issues in short-term and local rainfall prediction. Short-term and local rainfall usually refers to a 0–6 h (mainly a 0–2 h) rainfall event with high spatial and temporal resolution [26]. At present, radar echo extrapolation and fine numerical weather prediction model are the main technology commonly used to predict short-term rainfall [27]. However, the poor accuracy of extrapolation in radar echo inversion significantly affects the accuracy of forecasting. Only using radar extrapolation will result in greater prediction error and a limited forecast period. In addition, the forecast accuracy will decrease rapidly as the extrapolation time increases [28, 29]. Because of the poor spatial resolution of radar data, poor accuracy remains in rainfall forecasting at a local scale [30]. Commonly, the smaller the spatial scale, the greater the error [31]. During the 2008 Beijing Olympic Games, the Meteorological Agency forecasted strong convective weather using an expert system (B08FDP) that has a certain indicative forecasting [32]. Its 30-min and 60-min quantitative precipitation forecasts (QPFs) and TS scores were near 0.3 and 0.2, respectively. However, the random forecast hit rate still remains low [33] as the Shenzhen Meteorological Agency started in 2012 to forecast long-term heavy storms for the city. However, until now, there have still been some limitations in quantitatively short-term rainfall forecasts or other small probability events.

In recent years, the machine learning technology has also been widely used in meteorological element forecasting and predicting. Yang Shuqun used the SVM to predict the precipitation classification, and used the meteorological dataset from 1958-2003 to establish the SVM prediction model, which obtained a good prediction effect [34]. From the current point of view, the SVM classification method is more used since the temporal and spatial correlation characteristics of meteorological data are very strong. Lakshmanan et al. performed real-time mosaic on continental-scale weather and conducted experiments on rainfall estimation [35]; Piani et al. also constructed a series of rainfall prediction models, which were verified by a series of experiments, and the experiment achieved good results [36]. Venkadesh et al. used years of historical temperature data to combine artificial neural networks with genetic algorithms to establish a more accurate artificial neural network model [37]. Jiang combines PSO and GA neural networks to analyze monthly precipitation to establish a predictive model and predict monthly precipitation [38]. Tsymbal et al. conducted an in-depth discussion on online ensemble learning methods [39] by dynamically adjusting the number of base classifiers or adjusting some parameters of the base classifier, the dynamic ensemble and classification of classifiers was finally achieved. A model combining support vector machine, artificial neural network and random forest was established to improve the classification of convective and stratiform rain, and finally reached high overall accuracy of classification [40].

III. METHOD

A. Study area

Shenzhen is the forth-largest city along the South China coast and has abundant precipitation from April to October. Extreme weather occurs frequently and localized heavy rainfall and other weather anomalies are more commonly occurring as the global climate is changing. Different weather can appear within the city itself, for example, a heavy rainfall occurs in the Luohu District while clear skies appear in the Nanshan District. Furthermore, localization was also observed within the time scales of the heavy storm. The rapid change in both the spatial and time domains makes a great challenge for traditional numerical models of nowcasting.

Selected automated meteorological stations are distributed in every district in Shenzhen as shown in Figure 1. The average distance between any two stations is 3.8 km. The observational data cover precipitation, wind speed, wind direction, 2-m local temperature, and air pressure at a 1-h interval.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access

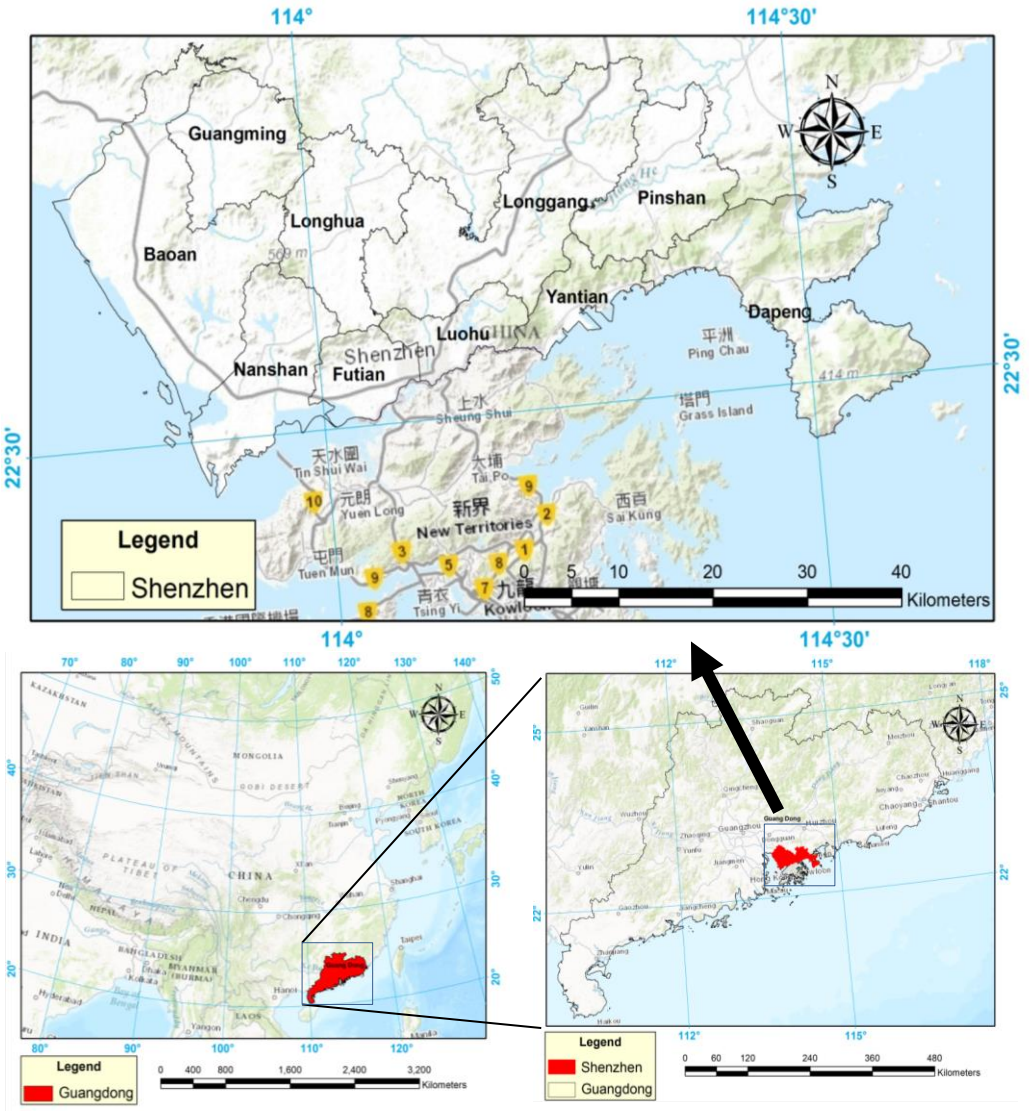


Figure 1. Study area

Satellite imagery from the Feng Yun II-G geostationary meteorological satellite (FY-2G) was used to capture the spatial variation in cloud characteristics at a 1-h time interval. FY-2G is the 8th operational satellite of the Feng Yun II geostationary meteorological satellite series developed by China. The satellite mainly covers cloud reflectance in the visible, near-infrared, and IR wavelengths in an electromagnetic spectrum. Its scanning radiometer has five channels which have two long-wave infrared bands, one medium-wave infrared band, one visible band, and one band in the water vapor absorption region. The FY-2G IR stray light suppression effect increase more than 50%, and the frequency and capabilities of the satellite calibration also increase. These technical improvements effectively improve the inversion accuracy of the FY-2G quantitative products and the application level. It can obtain one-third of the earth's surface images during non-flood seasons every hour and during flood seasons every half hour. In this study, many cloud parameters including top brightness temperature (TB), cloud top temperature (CTT), gradient of the pixel TB (GT), difference in TB (DT), cloud total amount (CTA), cloud type (CT), and middle and upper tropospheric water vapor content (WVC) were used as model inputs to maximally capture the characteristic of the precipitation cloud.

B. Cloud parameter estimation from FY satellite

First, the FY-2G satellite level 1 images needed to be pre-processed, including radiometric and geometric correction. Geometric correction is mainly used to unify the 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 were overlaid with other auxiliary data to analyze the tracking of cloud development. Radiometric correction removed the influence of different atmospheric environments on the image via using auxiliary radiation data of the satellite, thereby enhancing the ability to obtain quantitative information from the images. The data pre-processing of FY satellite images in this study consisted of three steps: the first step was cloud parameter estimation by cloud recognition using the FY satellite data. Cloud

classification and parameter estimation were then conducted according to the texture, shape, smoothness, and brightness temperature. Parameters including cloud type, cloud-top temperature, TB, gradient of the pixel TB, and difference in TB were finally obtained. Table I shows the specific selected parameters used in this study.

TABLE I
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 (10.3-11.3μm) 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)	$GT_{(x,y)} = \left[\left(TB_{(x+1,y)} - TB_{(x-1,y)}\right)^2 - \left(TB_{(x,y+1)} - TB_{(x,y-1)}\right)^2\right]^{1/2}$	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

C. *Dislocation SVM Modelling*

Using the characteristics and advantages of SVM addressing a small sample size, in this study, a multi-time scale multi-source SVM model was built to capture the data variation within a small time window. The input of the model included parameters from the FY-2G images (i.e. TB, CTT, GT, DT, CTA, CT, and WVC) and the parameters from the automated observation station (i.e. wind speed, humidity, temperature, and air pressure). The output of the model was rainfall probability in which "0" represented no rain and "1" represented rain. The basic assumption was that in the time series for each location, rainfall probability is highly related to the nearby historical meteorological conditions within this time series. The influence of historical meteorological conditions that are farther from the current prediction window can be regarded as model variance which can be neglected in real-time prediction. Based on this idea, rainfall probability can be predicted by analysis of the nearby meteorological conditions. The main idea of rainfall probability prediction is shown in Figure 2.

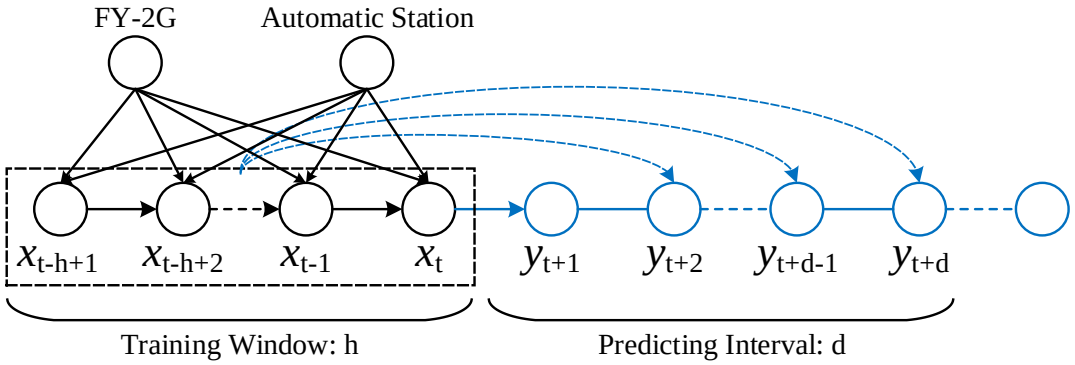


Figure 2. Basic idea of rain probability prediction for future d hours based on nearby weather conditions

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

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

Multi-time scale SVM rainfall prediction was based on the following equations:

$$Y_t = DSVM(X_t) \tag{2}$$

$$X_t = [x_{t-h-d+1}, x_{t-h-d+2}, ..., x_{t-d}]^T \tag{3}$$

$$Y_t = [y_{t-h+1}, y_{t-h+2}, ..., y_t]^T \tag{4}$$

$$y_{t+1} = DSVM(x_{t-d+1}) \tag{5}$$

where $d \in [0,6]$

h=2, 4, 8, 12, 24, 48, and 72

X_{t-1}, Y_{t-1}= training set

DSVM()=dislocation SVM model created by the training set

t = current time

h = training sample size (or time scales)

d = prediction time interval

y_{t+1}= rainfall value to be predicted

x_{t-d+1} =model input of rainfall value to be predicted

Here, the time windows of the training data were set as follows: 2 h, 4 h, 8 h, 12 h, 24 h, 48 h, and 72 h, which means for each prediction, 2-h, 4-h, 8-h, 12-h, 24-h, 48-h, and 72-h historical data were used to predict future rainfall probability. Meanwhile, the prediction intervals were also set as follows: 0 h, 1 h, 2 h, 3 h, 4 h, 5 h, and 6 h; thus, different prediction windows were used for probability estimation.

Once the precipitation was s predicted, the rain probability was calculated as follows:

$$P(y_t) = \frac{1}{1+e^{-y_t}} \tag{6}$$

Then, the 0–1 rain probability was calculated as follows:

$$P = \begin{cases} 1, P(y_t) > 50\% \\ 0, P(y_t) \le 50\% \end{cases} \tag{7}$$

Figure 3 shows the training and prediction diagram of the proposed model. In simple terms, for every time, for example, Time 1 and Time 2, the data before the current time were used to train the SVM model, and then the values after the current time were predicted. For every time, there was a specific SVM model, and thus for the next time, the SVM model was different. This feature ensures that the model is sensitive to and strongly related to the short-term and local weather conditions. To avoid overfitting of the SVM, we used a 10-fold cross-validation.

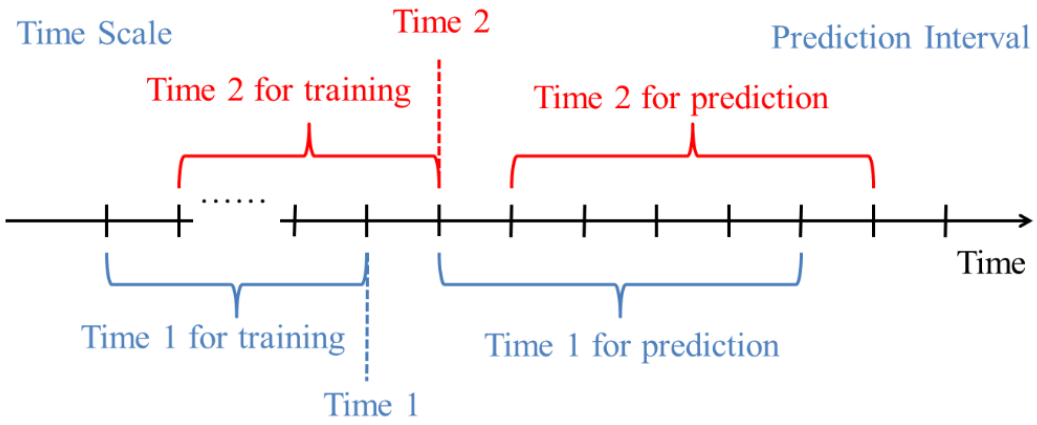


Figure 3. The diagram of the proposed model

As previously mentioned, the proposed model was a dislocation model as seen in Figure 4. For the traditional training strategy, the training x and training y should be simultaneous, i.e. the current independent variables (weather conditions) should be correspond to the dependent variable (precipitation) at the same time. During the testing procedure, the predicted y could be calculated using the input x at the same time. If the input x was unknown, then the input x was estimated first, which introduced more uncertainties, particularly when y for a situation in which the input x was composed of many variables. To solve this problem, we proposed a dislocation training strategy. As seen in Figure 4, the x and y data before the current time were known, and the x and y data after the current time were unknown. For a 1-h prediction, the time of training x was 1 h before training y, and then we inputted the data of the current time (input x) and obtained the predicted y 1 h after the current time. The procedure for a 2-h prediction was the same. The calculation of SVM was the same as that of the traditional SVM model, and the main difference was the composition of the training data. Using the proposed dislocation strategy, we did not need to estimate the input x; hence, no other uncertainty was introduced. This strategy is reasonable because future precipitation is more related to short-term and local weather conditions.

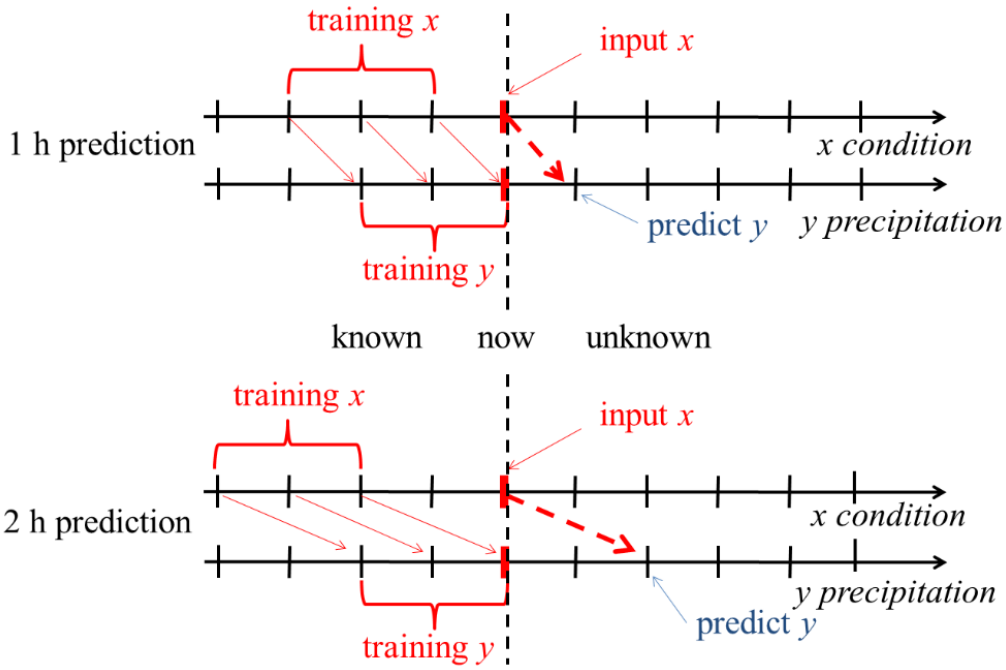


Figure 4. The explanation diagram of the proposed dislocation model

Compared to the traditional SVM model, the proposed method is simpler, faster, and more adaptive. For the traditional SVM model, the general approach is to divide all of the data into two parts, i.e. training data and testing data, and then train an SVM model using all the training data and validate the model using testing data. In many other application fields, this training strategy is effective because some main features are contained in the training data. However, for rainfall prediction, this strategy is not reasonable because the current weather is less related to previous weather conditions occurring long before the current time, and nearby weather conditions may have more influence on future weather. Therefore, we proposed the dislocated SVM model that employed nearby weather data to train a specific model for every time. Because there are much less training data than that used in a general SVM model, the proposed method is simpler, faster, and more adaptive. This is the main feature of our proposed dislocated SVM model and this is also the main difference between our model and a general SVM model.

D. Validation

The traditional means to evaluate forecast accuracy in short-term weather forecasting is threat score (TS), probability of detection (POD) and false alarm ratio (FAR), which can be calculated using the following three equations, respectively:

$$TS_k = \frac{NA_k}{NA_k + NB_k + NC_k} \tag{8}$$

$$POD_k = \frac{NA_k}{NA_k + NB_k} \tag{9}$$

$$FAR_k = \frac{NC_k}{NA_k + NC_k} \tag{10}$$

where NA_k = the number of correct predictions,
NB_k = the number of more predictions, and
NC_k = the number of less predictions.

The meaning of every parameter in the aforementioned equations is shown in Table II.

TABLE II
PARAMETERS FOR VALIDATION

True \ Prediction	Rain	NoRain
Rain	NA_k	NC_k
NoRain	NB_k	ND_k

The previous TS score does not consider the situation that the prediction is no rain and actually no rain occurs. In some applications, this is also very important. Hence, we use an overall accuracy to describe this situation as follows:

$$OA = \frac{NA_k + ND_k}{NA_k + NB_k + NC_k + ND_k} \tag{11}$$

In details, for an observation station, assuming the current time is 11:00, in order to predict the rainfall at 12:00, the previous 6 hours (i.e. time scale) data were used, and the prediction interval was 1 hour. The modeling process consisted of the following steps:

(1) Data preprocessing and training data preparing

The input data included FY-2G satellite parameters (i.e. TB, CTT, GT, DT, CTA, CT, WVC) and meteorological elements (i.e. wind speed, humidity, temperature, and air pressure). For this example, the input data of previous 6 hours were needed for dislocated training data set as:

$$\begin{cases} X^{<11>} = [x^{<5>}, x^{<6>}, x^{<7>}, x^{<8>}, x^{<9>}, x^{<10>}] \\ Y^{<11>} = [y^{<6>}, y^{<7>}, y^{<8>}, y^{<9>}, y^{<10>}, y^{<11>}] \end{cases} \quad (12)$$

where *X* are the input meteorological parameter vectors at 5:00, 6:00, 7:00, 8:00, 9:00, 10:00 and *Y* are the input rainfall vectors at 6:00, 7:00, 8:00, 9:00, 10:00, 11:00. $x^{<5>}$ is the parameter vectors at 5:00 and $y^{<6>}$ is the rainfall at 6:00. It should be noted that the *X* and *Y* are dislocated matched. The detail contents of $x^{<5>}$ are:

$$x^{<5>} = [TB,\ CTT,\ GT,\ DT,\ CTA,\ CT,\ WVC, W, H, T, P]^T \quad (13)$$

where the *W*, *H*, *T*, *P* represent wind speed, humidity, temperature, and air pressure respectively.

(2) Training a dislocation SVM model

A dislocation SVM model at 11:00 can be obtained using the training data set $[X^{<11>}, Y^{<11>}]$ and SVM model:

$$Y^{<11>} = DSVM^{<11>}(X^{<11>}) \quad (14)$$

where $DSVM^{<11>}$ is the trained dislocation SVM model specific at 11:00 for predicting future 1 hour rainfall at 12:00.

(3) Prediction for future 1 hour

Using the trained dislocation SVM model at 11:00 and the meteorological parameter vectors at 11:00, the rainfall at 12:00 can be predicted as:

$$y^{<12>} = DSVM^{<11>}(x^{<11>}) \quad (15)$$

where $x^{<11>}$ is the meteorological parameter vectors at 11:00 and $y^{<12>}$ is the rainfall at 12:00.

(4) Prediction for future other hours

Repeating the above steps and using different time scales and prediction intervals, the rainfall prediction for future other hours can be obtained. For example, if the prediction interval is set as 2 and time scale is 6, the future 2 hour (i.e. 13:00) rainfall can be predicted using the following steps:

$$\begin{cases} X^{<11>} = [x^{<4>}, x^{<5>}, x^{<6>}, x^{<7>}, x^{<8>}, x^{<9>}] \\ Y^{<11>} = [y^{<6>}, y^{<7>}, y^{<8>}, y^{<9>}, y^{<10>}, y^{<11>}] \\ Y^{<11>} = DSVM^{<11>}(X^{<11>}) \\ y^{<13>} = DSVM^{<11>}(x^{<11>}) \end{cases} \quad (16)$$

The training steps for other models using different time scales and prediction intervals are the same as the above steps. The only differences are the contents of every input vectors.

(5) Accuracy evaluation

For every hour in time series and every observation station in each hour, the rainfall can be predicted using the specific dislocation SVM model corresponding to each station at each hour. After all the rainfall in time series are predicted, the accuracy assessment can be achieved using the accuracy indices, i.e. OA, TS, POD and FAR.

IV. EXPERIMENTS

In order to test and verify the performance of the proposed dislocation SVM model, some objective assessment indices (i.e. the overall accuracy, TS score, POD and FAR) were used to analyze the performance and influence of different time scales and prediction intervals. The stability of model was evaluated by the mean and standard deviation of the TS scores. Meanwhile, the comparisons between other machine learning methods with and without dislocation modeling were also discussed.

The FY-2G satellite imagery and in-situ observation data from September 1st, 2014, to August 31st, 2015 were employed to test and validate the proposed method. Unlike satellite data pre-processing, ground data was mainly checked by the quality of the automated station data, including obvious missing data and errors. Data selection was based on choosing high-quality samples to ensure the accuracy of the model output. Because the FY2G image spatial resolution was 5 km × 5 km, however, the automated station was fixed. Thus, we used the nearest neighbor matching principle to select an image pixel which corresponded to the specific automated station location.

A. FY-2G Data Parameters

According to the statistical analysis and precipitation mechanism analysis, the cloud parameter variations could have a correlation with a rainfall event. Figure 3 shows an example regarding CTT, CTA, CT, and WVC of FY-2G parameters at 23:00 on June 28, 2015. Figure 4 shows an example regarding the BT, channel 1 and Channel 4 brightness temperature difference, brightness temperature gradient, and the channel 3 and Channel 4 brightness temperature difference in FY-2G parameters at the same time as those shown in figure 3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access

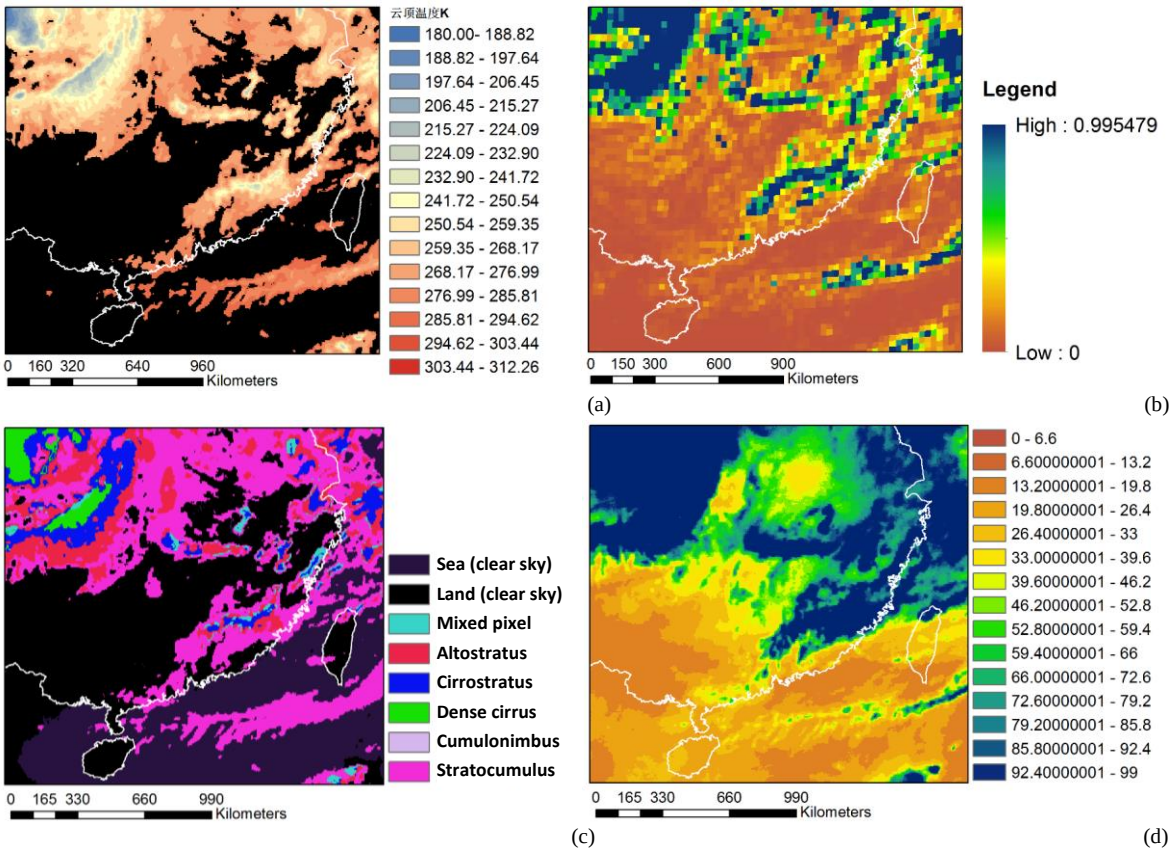


Figure 5. FY-2G parameters, including (a) cloud top temperature, (b) the total amount of cloud, (c) the type of cloud, (d) upper tropospheric water vapor content, at 23:00 on June 28, 2015

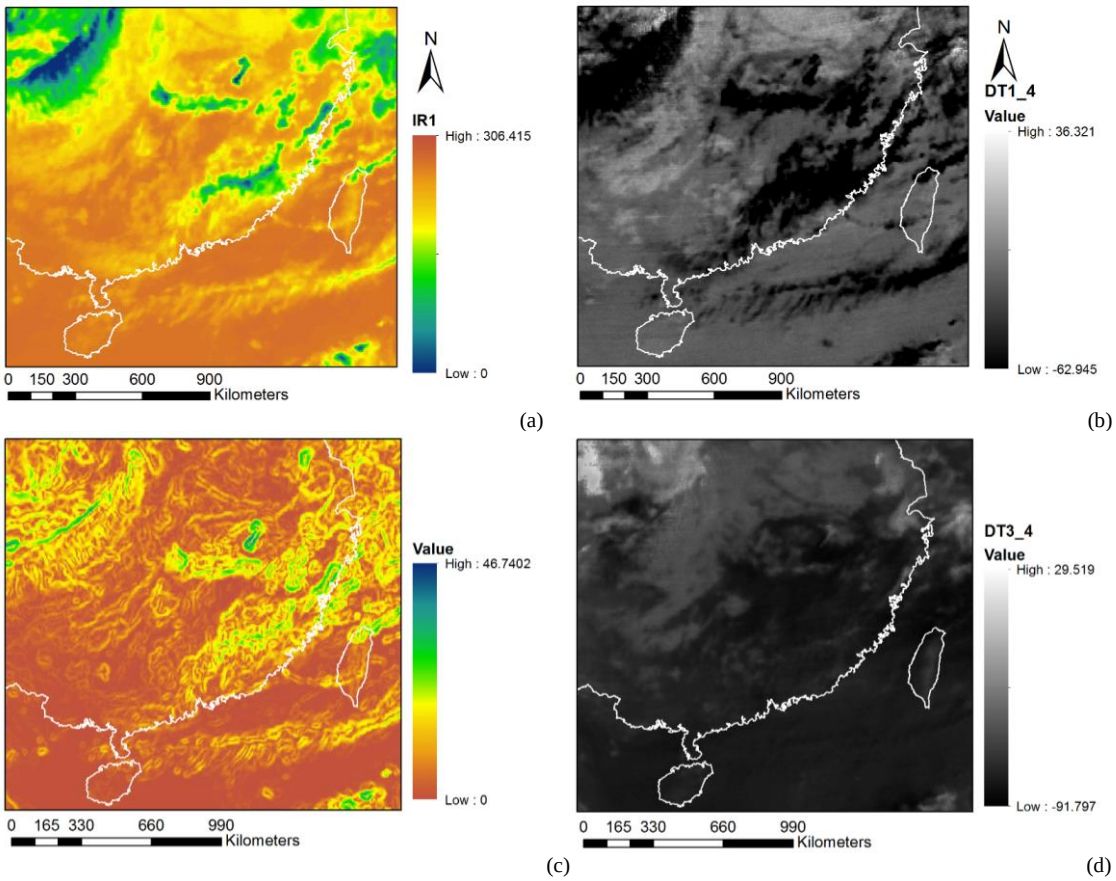
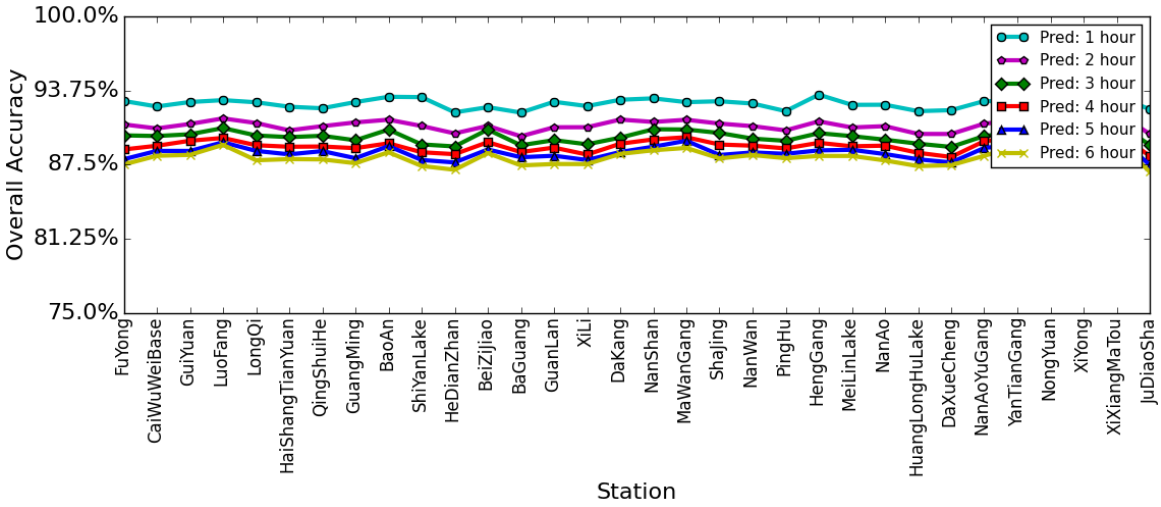


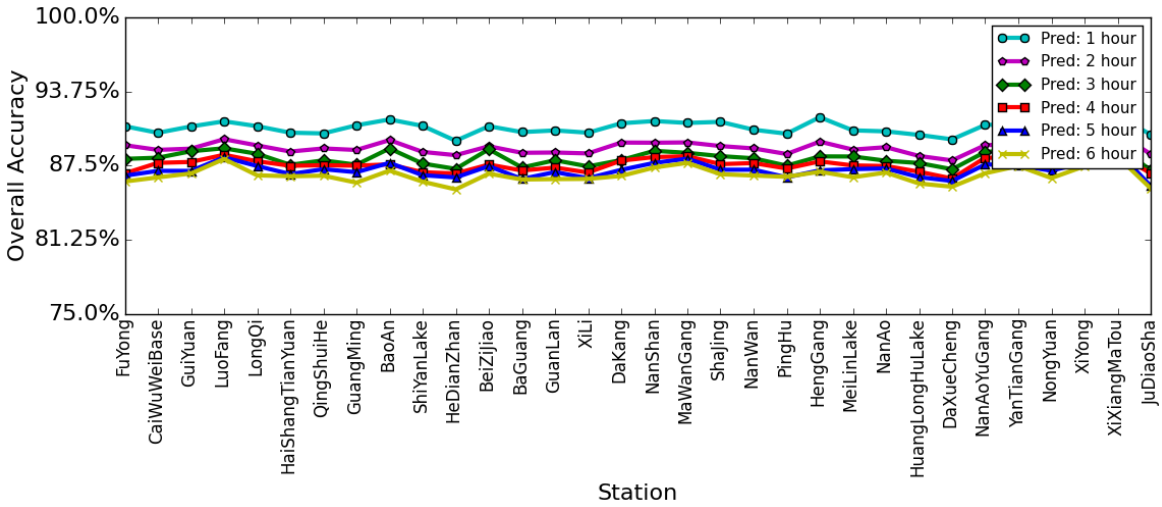
Figure 6. FY-2G parameters, including (a) brightness temperature, (b) channel 1 and Channel 4 brightness temperature difference, (c) brightness temperature gradient, (d) channel 3 and Channel 4 brightness temperature difference, at 23:00 on June 28, 2015

B. Overall accuracy of rain probability prediction

The overall accuracy indicates the overall performance of a model that predicts rain or no rain in the future. Figure 5 shows the overall accuracy of rain prediction at 2-h and 4-h -time scales. All in situ weather stations were employed to evaluate the accuracy. The result shows that the overall prediction accuracy of each station is relatively sufficiently high to practical application. Different colors represent the prediction from 1- to 6-h time intervals in the future. The overall prediction accuracies of most stations for the 1-h prediction interval were greater than 90%. The overall accuracy of the different prediction intervals showed little difference. Among them, for a 1-h prediction interval, the accuracy at Henggang station was greater than 94% and 90% for the 2-h and 4-h time scales, and the accuracy of Hedianzhan station was relatively poorer but still greater than 89%. These results prove the effectiveness of the proposed method.



(a)



(b)

Figure 7. The overall accuracy for future different hours rainfall prediction. (a) 2 hours-time scale (b) 4 hours-time scale

C. Overall TS scores

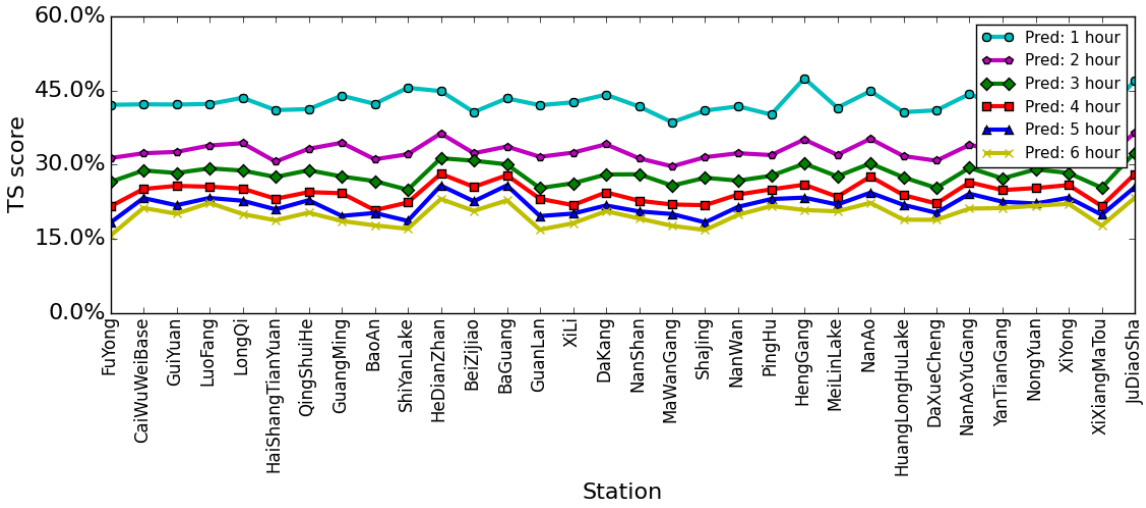
Table 3 shows the TS scores predicted by 2-h time scale. As shown in Table 3, the prediction performance was better and the TS score was much higher at most of the stations compared to the studies of Zhang Lei, 2015. Prediction accuracy varied among the different stations. One reason is that the rainfall at a specific station could be affected by many different factors with varying influence degree. It is clear as shown in Table 3 that the TS scores decreased as the prediction time interval increased. This follows the common sense of rainfall prediction.

TABLE III
TS SCORES OF 2HOURS-TIME SCALE SVM FOR FUTURE DIFFERENT INTERVAL PREDICTION

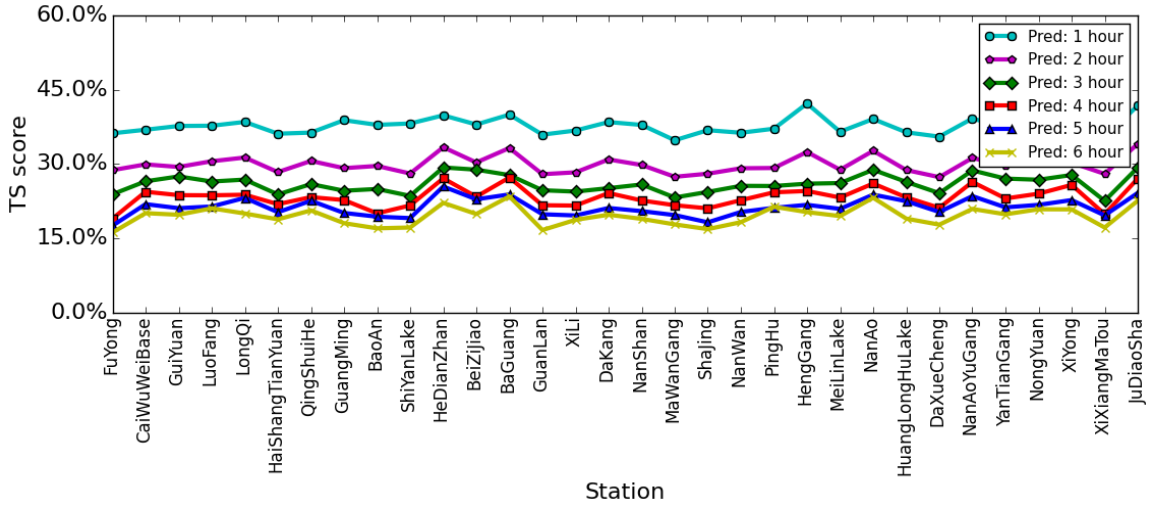
Interval \ Station	Xili	Dakang	Nanshan	……	Nanwan
1 hour	42.63%	44.18%	41.77%	……	41.81%
2 hours	33.33%	34.63%	33.60%	……	32.74%
3 hours	26.74%	28.20%	27.22%	……	26.82%
4 hours	23.06%	25.36%	25.38%	……	23.85%
5 hours	20.09%	22.33%	21.66%	……	21.67%
6 hours	19.15%	20.25%	19.69%	……	19.42%

D. Influence of time scale

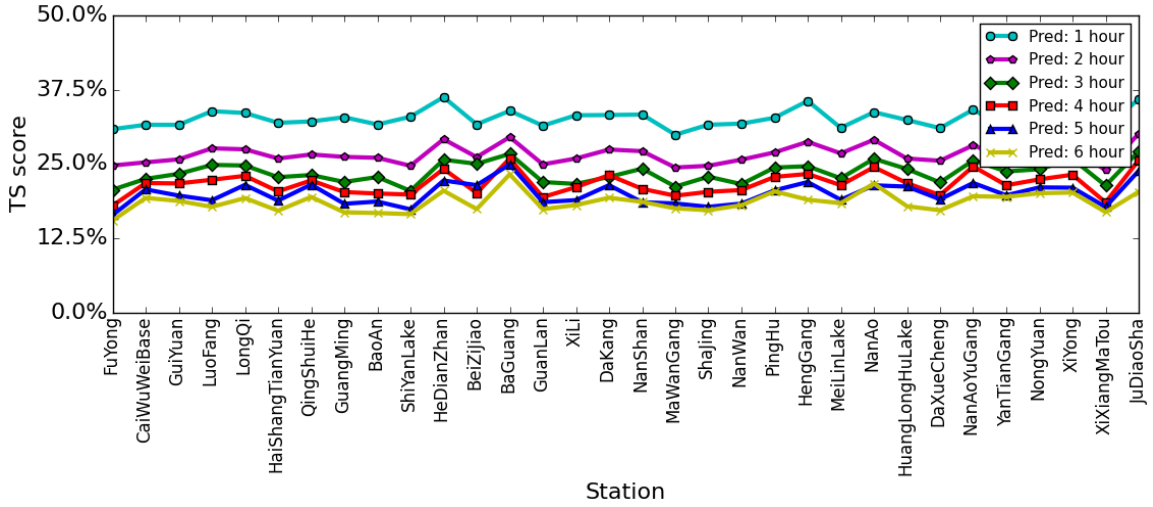
As mentioned previously, nearby weather conditions have a different influence on rain probability prediction. Hence, the influence of time scale was tested. The time scale increased from 2 h to 72 h, and because of the page limitation, only the TS scores for 2 h, 4 h, 8 h, 24 h, 48 h, and 72 h are presented, as shown in figure 8.



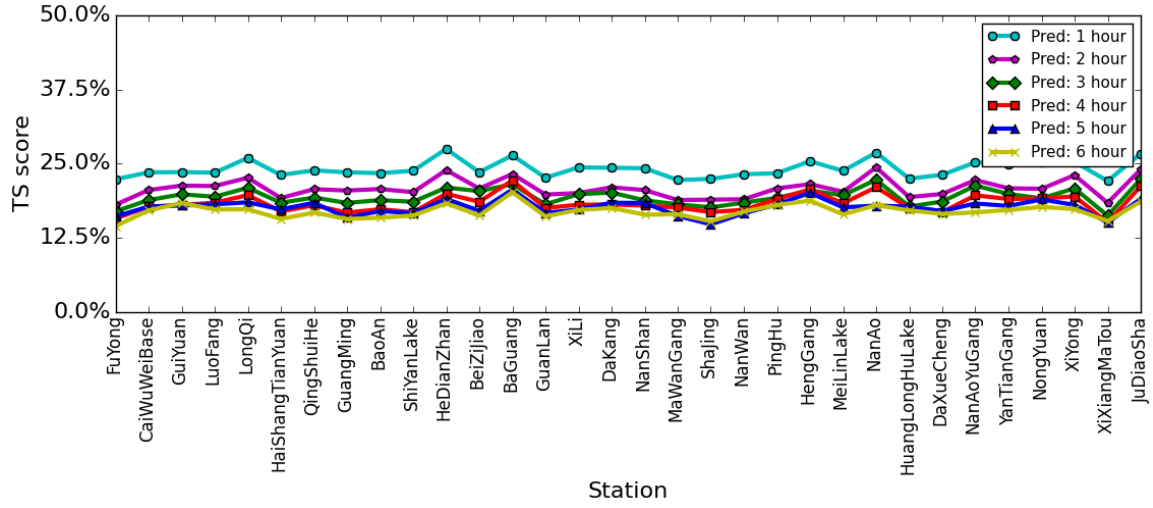
(a)



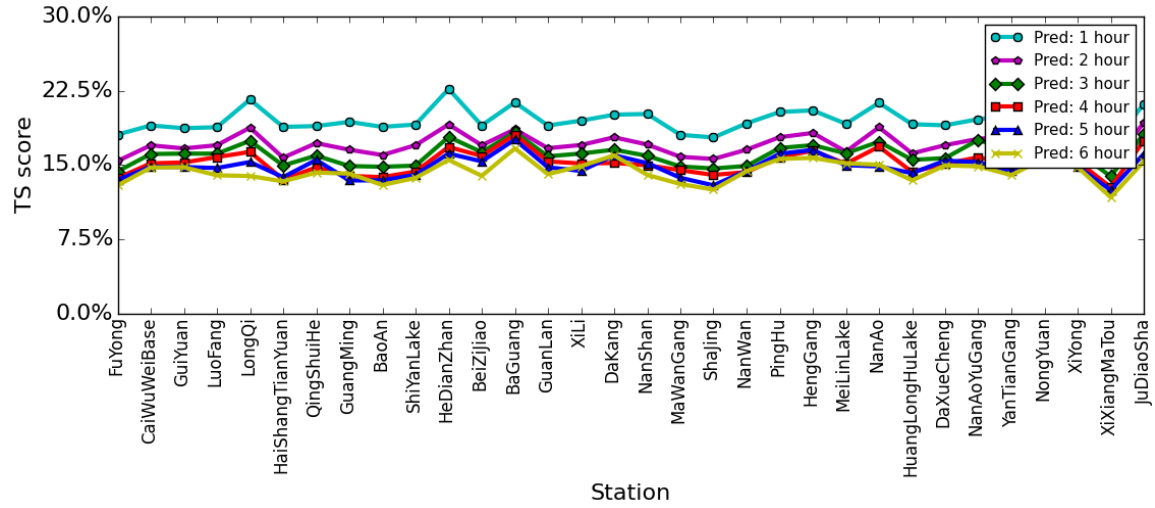
(b)



(c)

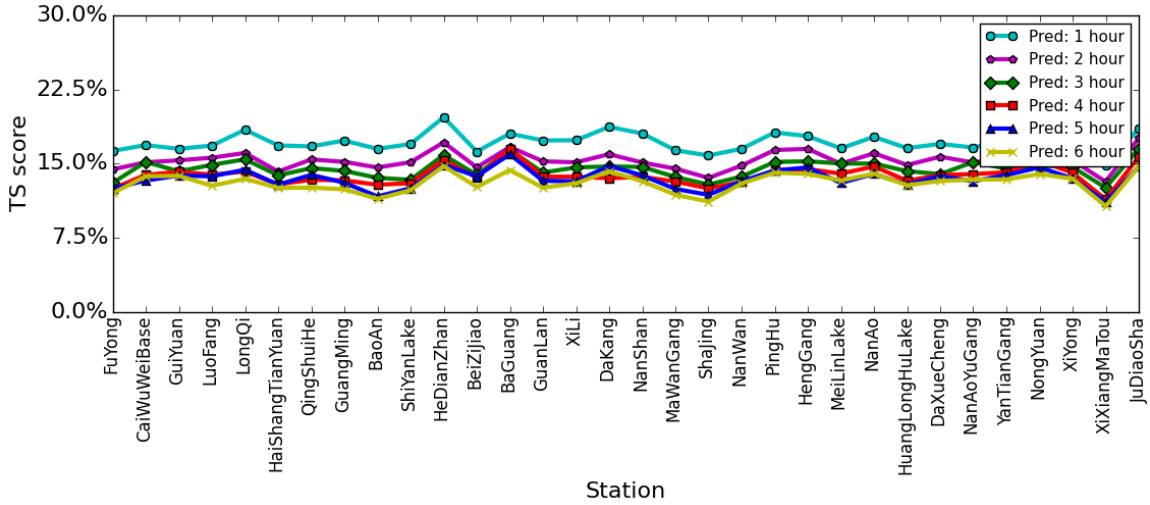


(d)



(e)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access



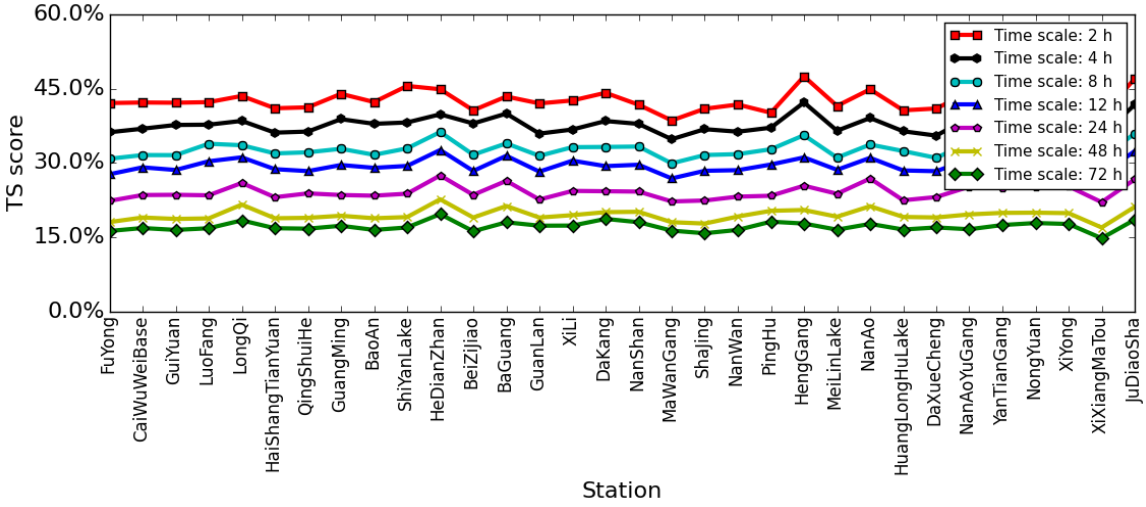
(f)

Figure 8. The TS scores change with increasing of time scale. (a) 2 hours-time scale. (b) 4 hours-time scale. (c) 8 hours-time scale. (d) 24 hours-time scale. (e) 48 hours-time scale. (f) 72 hours-time scale.

Obviously, the 2-h time scale had the highest TS score. This can be seen as evidence of our basic assumption that current rainfall is more likely correlated to nearby weather conditions during the time series. Second, the TS scores of the 4-h, 8-h, 24-h, and 72-h time scales showed a trend that the prediction results decrease in accuracy as time scale increases. This shows that with a training sample increase, many meteorological factors unrelated to future rainfall were introduced into the model resulting in an increase in uncertainties in the model. It is also readily apparent that with increasing prediction interval, the TS score decreases, indicating that weather further into the future is more difficult to predict.

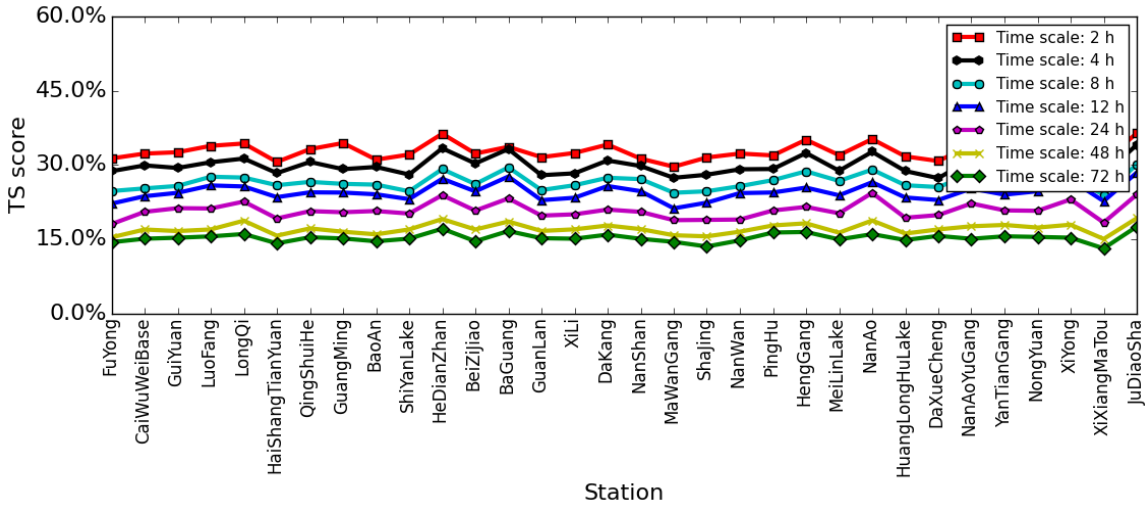
E. Influence of prediction interval

Analogous to the influence of time scale, the prediction interval also had an important influence on final prediction accuracy. During this experiment, the prediction interval increased from 1 h to 6 h and the TS scores were calculated for every prediction interval with different time scales. The results are shown in figure 9.

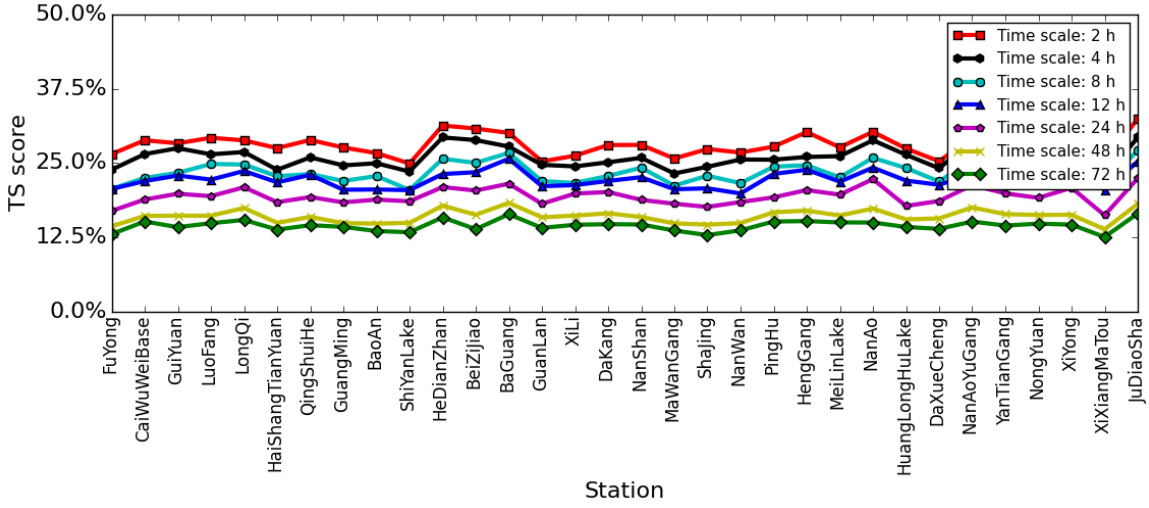


(a)

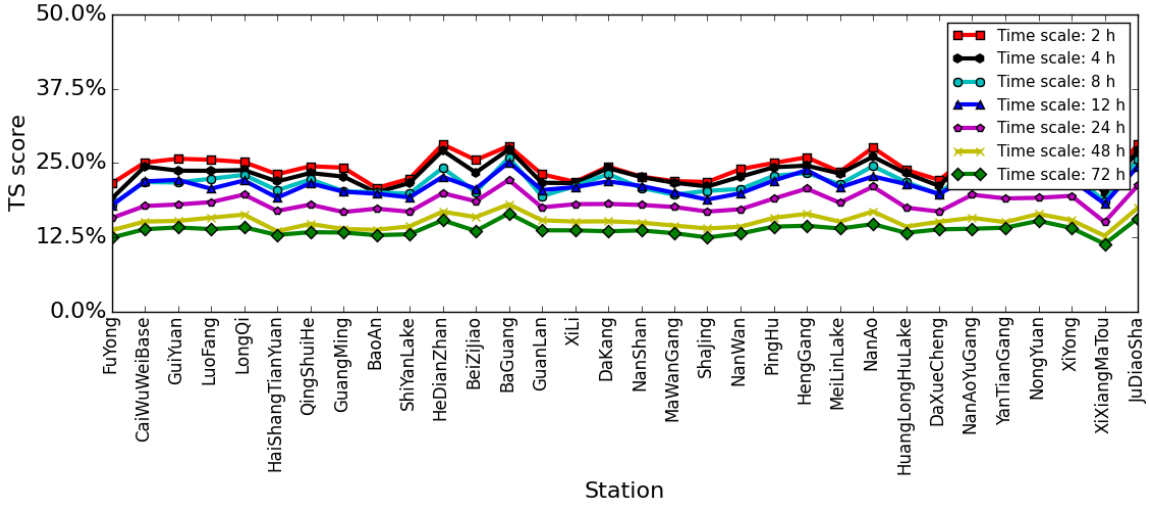
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access



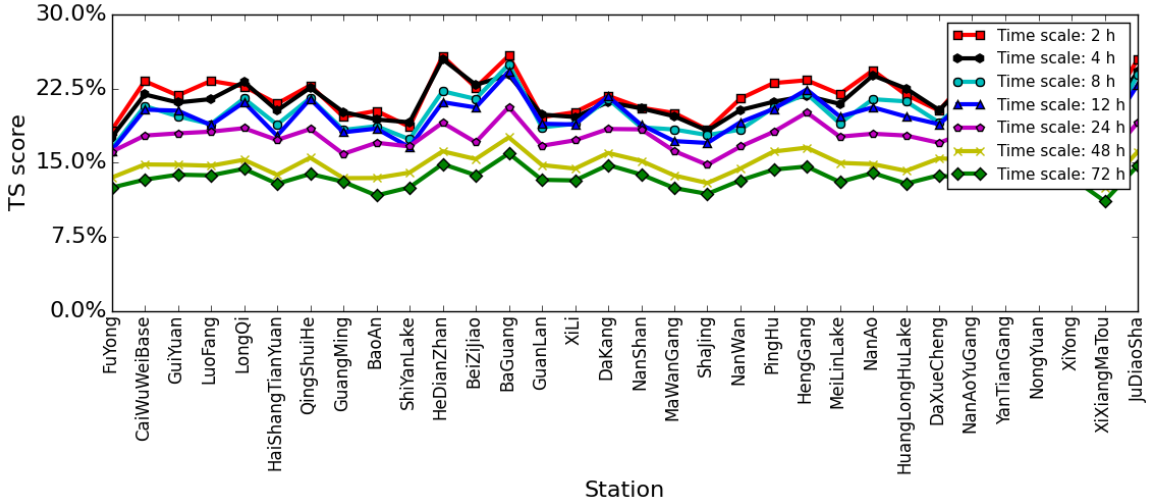
(b)



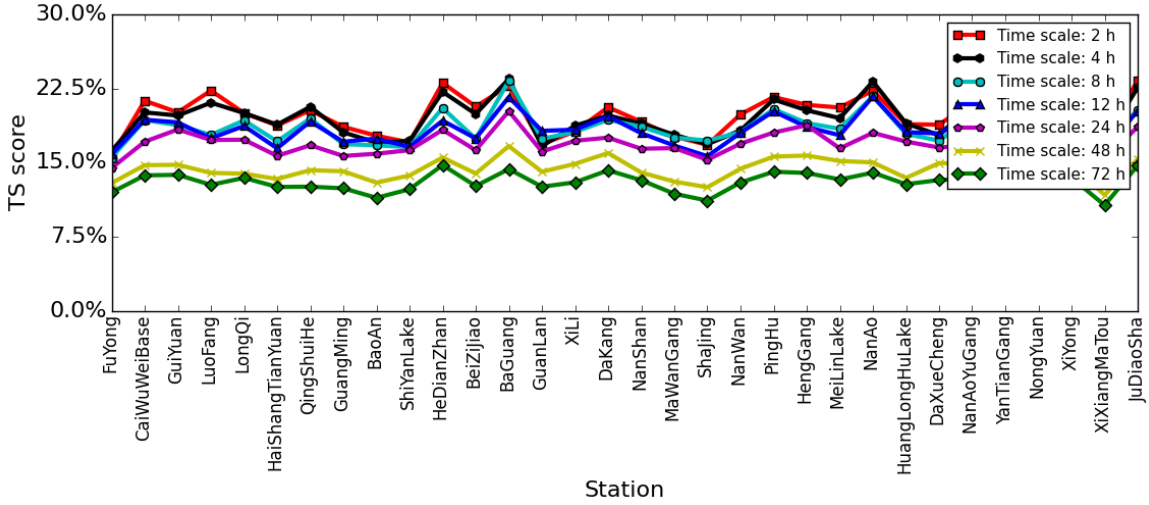
(c)



(d)



(e)



(f)

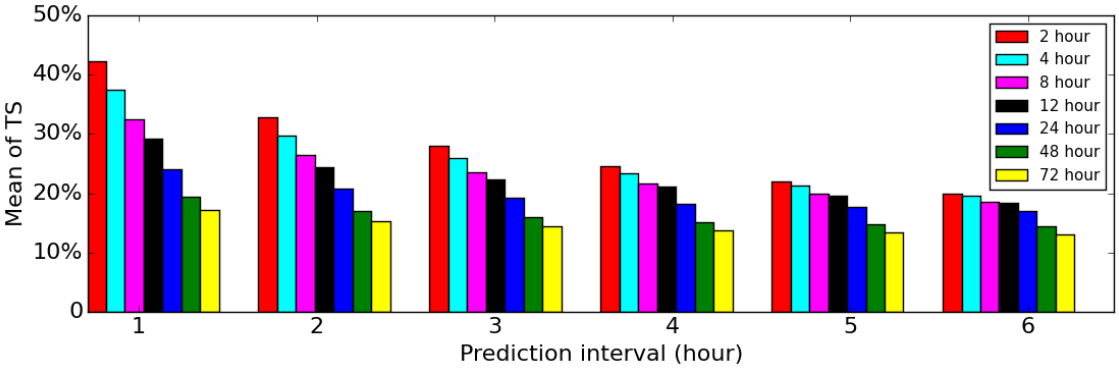
Figure 9. The TS scores change with increasing of prediction interval. (a) 1 hours prediction interval. (b) 2 hours prediction interval. (c) 3 hours prediction interval. (d) 4 hours prediction interval. (e) 5 hours prediction interval. (f) 6 hours prediction interval.

The trend as shown is also obvious and similar to the influence of time scale described in the previous section. As the prediction interval increased, the TS scores decreased. For a specific prediction interval, the TS score decreased when the time scale increased. Even when the prediction interval reached 6 h and time scale was 72 h, the average TS score of all the stations can still reach 10%, which is higher than the results of Lei Zhang (2015). These results also indicate that it is easier to predict the nearby future weather condition, and the nearby weather condition has more influence on the future prediction.

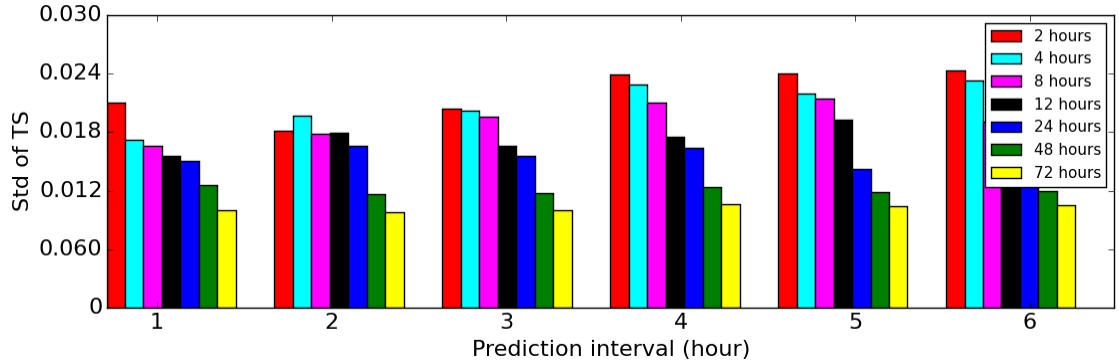
F. Stability of model

The stability of the model will affect the sustainability and applicability of the prediction. Therefore, establishing a robust model is very important. A good model should have similar performance among all the stations under all environmental conditions. To evaluate this performance, we used the mean and standard deviation of the TS scores for all of the stations using different time scales and prediction intervals. The smaller the standard deviation, the more stable the model. On the other hand, the standard deviation of the TS score also exhibits discrete trends as the prediction interval increases. A smaller standard deviation indicates that the majority of the stations' TS scores are near to the overall mean. Figure 10 shows the mean and standard deviation of the TS scores.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access



(a)



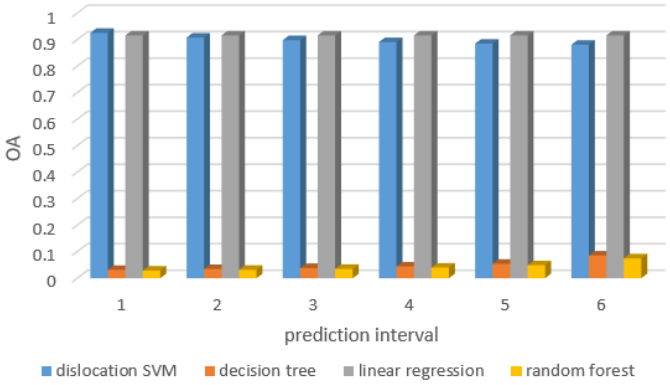
(b)

Figure 10. The stability of proposed method. (a) mean of TS scores about different hours-time scale SVM for all station and (b)standard deviation of TS scores about different hours-time scale SVM for all station

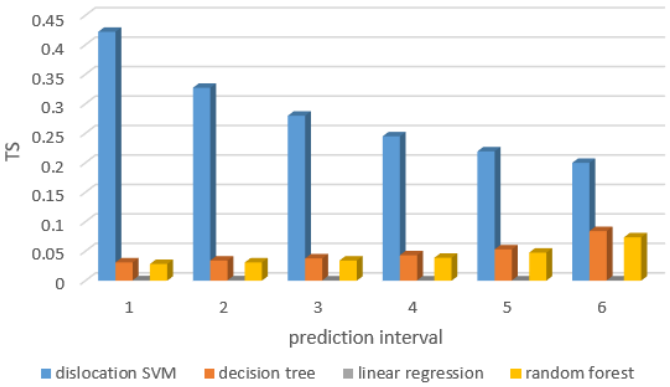
From the mean TS, we can draw the same conclusion that the TS score will decrease with an increase in time scale, and also with an increase in prediction interval. The 2-h time scale and 1-h prediction interval resulted in the highest TS score of more than 40%, and the lowest TS score was still greater than 10%. The standard deviation reflects the deviation degree among all TS scores and their mean. For all time scales and prediction intervals, the standard deviation was only from 0.01 to 0.024, which was sufficiently small to indicate that the proposed dislocation SVM model had good and stable prediction performance, and that the results were not affected by the location and environmental conditions.

G. Comparison with dislocation model

The proposed dislocation model can be used not only on SVM, but also on other machine learning algorithms, such as decision tree, linear regression, and random forest. However, the real performance of different algorithms varies according to their learning capacities. We tested the aforementioned three algorithms using the same data and dislocation model as the proposed method. The average results are shown in figure 11.

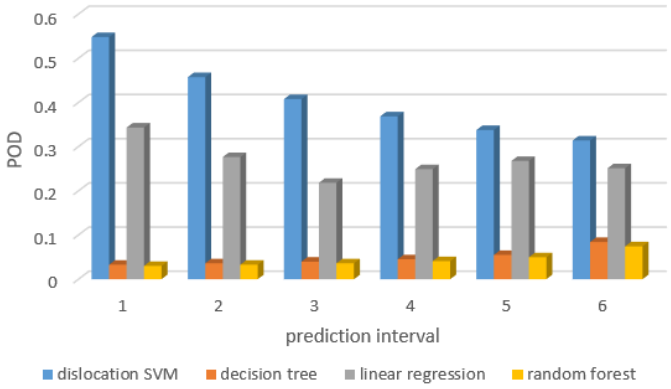


(a)

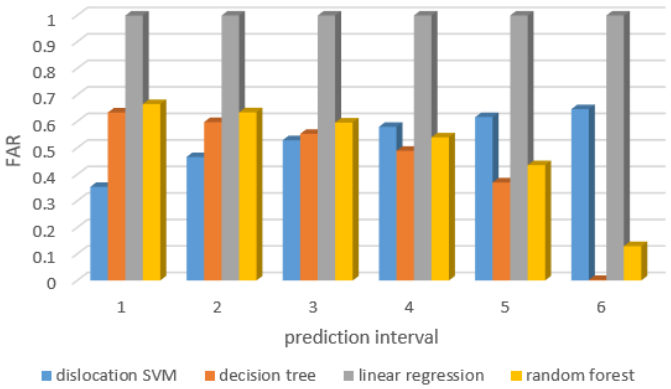


(b)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access



(c)



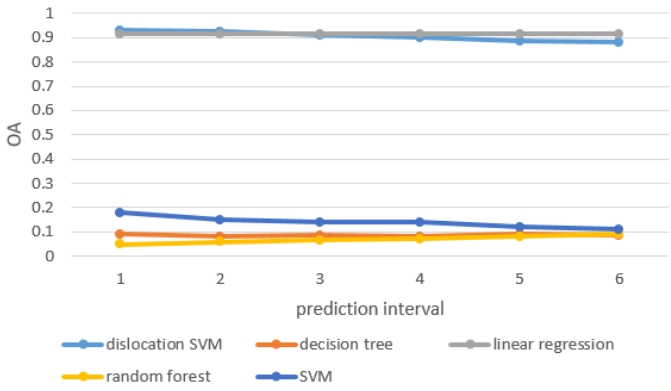
(d)

Figure 11. The comparisons between different algorithms using dislocation model. (a) OA (b) TS (c) POD (d) FAR

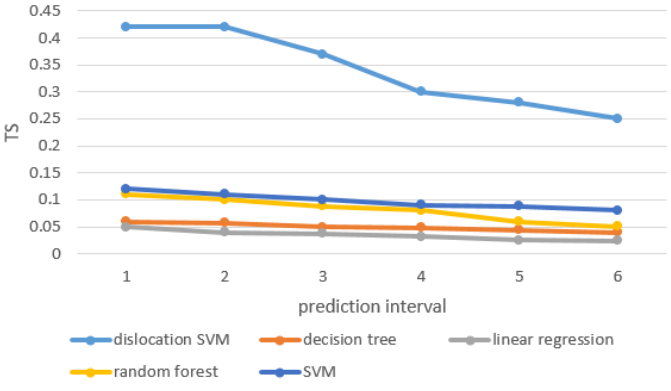
The amount of no-rain (i.e. 0) data was much greater than that of rain (i.e. 1) data. The OA of linear regression was nearly the highest because it predicted almost all of the rain (1) compared to no-rain (0) data, which can be seen from the TS index in which its TS was the lowest. This can also can be seen from FAR index where the FARs of linear regression were all near to 1 for all prediction intervals. The performances of decision tree and random forest were similar to one another because they were both based on the same mathematic principle. As the prediction interval increased, the prediction became increasingly more difficult, which can be seen from the decreasing trend of the TS index using the proposed method. The TS indices of the other three algorithms were much less than the proposed method because they had much more NB and NC prediction values (see table 1), indicating a lower prediction capacity.

H. Comparison without dislocation model

The dislocation model was the key to the proposed method because the nearby weather conditions might have more influence on future weather. However, using the traditional approach, historical data were used to train the model which were then used in the prediction of future data. There were two training strategies. In this section, they were compared. The one-year time-series data were divided into two parts, i.e. the first eight months of data were used to train the algorithms, including decision tree, linear regression, random forest, and SVM, and the last four months of data were used to test the prediction performance of these algorithms. Meanwhile, for the last four months of data, the proposed dislocation model was used directly to predict rainfall for every hour. This meant that two strategies shared the same test data. Finally, the TS, OA, POD and FAR values were calculated using the shared test data. The results are shown in figure 12.

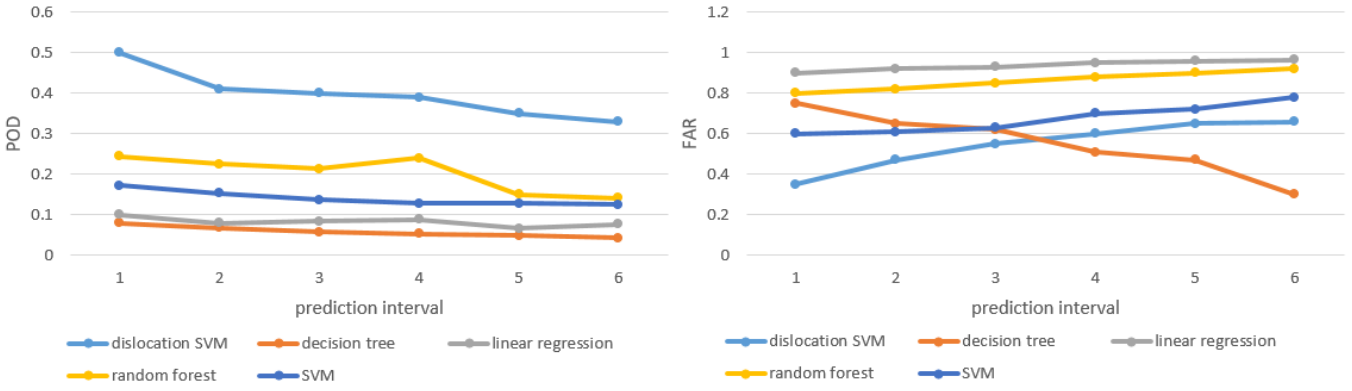


(a)



(b)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2913366, IEEE Access



(c) (d)

Figure 12. The comparisons between different algorithms without dislocation model. (a) OA (b) TS (c) POD (d) FAR

Similar to that shown in figure 11, the linear regression algorithm shows the highest OA and FAR but the lowest TS, which means it provides much more erroneous prediction, i.e. predicting rain to no-rain. The proposed method showed the highest TS and POD and a lower FAR, which means that most of the rain data was correctly predicted. However, the other four algorithms showed much worse performance. These results prove the effectiveness of our strategy and method.

V. Conclusion

In this study, we combined FY-2G satellite data and in situ automated meteorological station hourly observations to establish a rain probability prediction model. Based on multi-source data assimilation analysis, the results proved the effectiveness of the proposed dislocation SVM-based method in short-term local rainfall forecasting. The prediction model also had certain practical value for the next six hours of rainfall forecasting. Using the proposed dislocation SVM modeling, the longer the time scale, the worse the results as predicted by the model, and vice versa. This meant that current rainfall was more relevant with recent historical meteorological conditions. In addition, the smaller the time scale, the lower the calculation costs.

The study proposed a finer and higher density rainfall geospatial forecast method via a combination of FY-2G satellite and the Shenzhen automated meteorological station data. This method was an attempt to solve the limitations of current short-term local rainfall probability forecasting in Shenzhen. It was also useful for assessing frequent small area rainfall anomalies and local short-term rainfall in Shenzhen. Our next study will focus on continuing to optimize the proposed forecasting method, employing more meteorological parameters. This work was expected to further improve the accuracy of rainfall forecasting.

References

[1] A. Swe Swe, S. Yu, O. Shin. (2018). Short-term Prediction of Localized Heavy Rain from Radar Imaging and Machine Learning. *IEIE Transactions on Smart Processing & Computing*, 7(2): 107-115.

[2] Andrade, Fernando J. A., de M. Alvaro A. M., da S.M. Luiz A. R. (2017). Short-Term Rain Attenuation Predictor for Terrestrial Links in Tropical Area. *IEEE antennas and wireless propagation letters*. 16: 1325-1328.

[3] M. Li. (2017). Study of the objective probability forecast method for short-term heavy rain based on ECWMF fine-mesh model. *Journal of Tropical Meteorology*, 33(6): 812-821.

[4] X. Fang, A.M. Shao, X.J. Yue.(2018). Statistics of the Z-R Relationship for Strong Convective Weather over the Yangtze-Huaihe River Basin and Its Application to Radar Reflectivity Data Assimilation for a Heavy Rain Event. *Journal of meteorological research*. 32(4):598-611.

[5] Y.C. Yang. (2018).Practical Method for 4-Dimentional Strategic Air Traffic Management Problem With Convective Weather Uncertainty. *IEEE transactions on intelligent transportation systems*, 19(6): 1697-1708.

[6] G. Nils, J. Tijana, S. Christoph.(2018). Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly journal of the royal meteorological society*, 144(713): 1218-1256.

[7] A. Barszcz, J.A. Milbrandt, J.M. Theriault. (2018). Improving the Explicit Prediction of Freezing Rain in a Kilometer-Scale Numerical Weather Prediction Model. *Weather and forecasting*. 33(3): 767-782

[8] B. Luitel, G. Villarini, G.A. Vecchi. (2018). Verification of the skill of numerical weather prediction models in forecasting rainfall from US landfalling tropical cyclones. *Journal of hydrology*. 556: 1026-1037.

[9] L. Liao, R. Meneghini. (2019). A Modified Dual-Wavelength Technique for Ku- and Ka-Band Radar Rain Retrieval. Journal of applied meteorology and climatology. 58(1): 3-18.

[10] H. Kikuchi, T. Ushio, F. Mizutani. (2018). Improving the Accuracy of Rain Rate Estimates Using X-Band Phased-Array Weather Radar Network. *IEEE transactions on geoscience and remote sensing*. 56(12): 6986-6994.

[11] L. Zhang, M.J. Wang, H. Li. (2015). Discussion of Relative Accuracy of Short-Range Heavy Rain Nowcasting. *Guangdong Meteorology*, 2015, 2: 1-6.

[12] Z. Shao, L. Zhang, L. Wang. (2017). Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass, *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens*., **99**:1-14.

[13] Z. Shao, L. Zhang, X. Zhou, L. Ding. (2014). A Novel Hierarchical Semisupervised SVM for Classification of Hyperspectral Images. *IEEE Geosci. Remote Sens.* 11(9): 1609 - 1613.

[14] J. Liu, X. Zhou, J. Huang, S. Liu, H. Li, S. Wen, J. Liu, (2017). Semantic classification for hyperspectral image by integrating distance measurement and relevance vector machine, *MultiMedia Syst.*, **23**: 95-104.

[15] K. Chen, J. Liu, S.X. Guo, J.S. Chen, P. Liu, J. Qian, H.J. Chen, B. Sun. Short-term precipitation occurrence prediction for strong convective weather using FY2-G satellite data: a case study of Shenzhen, south China. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, XLI-B6: 215-219.

[16] K.W. Chiang, W.C. Peng, Y.H. Yeh. (2009). Study of Alternative GPS Network Meteorological Sensors in Taiwan: Case Studies of the Plum Rains and Typhoon Sinlaku. *Sensors*, 9: 5001-5021.

[17] R.S. Schumacher, A.J. Clark. (2014). Evaluation of Ensemble Configurations for the Analysis and Prediction of Heavy-Rain-Producing

Mesoscale Convective Systems. *Monthly Weather Revie*, 142:4108-4138.

[18] Amarjit, R. P. S. Gangwar. (2008). Implementation of Artificial Neural Network for Prediction of Rain Attenuation in Microwave and Millimeter Wave Frequencies. *IETE Journal of Research*, 54: 346-352.

[19] D. Das, A. Maitra. (2016). Fade-Slope Model for Rain Attenuation Prediction in Tropical Region. *IEEE Geoscience and Remote Sensing Letters*, 13:777-781.

[20] Z.P. Li, M.D. Song, H. Feng. (2016). Within-season yield prediction with different nitrogen inputs under rain-fed condition using CERES-Wheat model in the northwest of China. *Journal of the Science of Food and Agriculture*, 96: 2906-2916

[21] T. K. Bahaga, F. Kucharski, G.M.Tsidu. (2016). Assessment of prediction and predictability of short rains over equatorial East Africa using a multi-model ensemble, *Theoretical and Applied Climatology*, 123:637-649.

[22] S.E. Nicholson. (2015). The Predictability of Rainfall over the Greater Horn of Africa. Part II: Prediction of Monthly Rainfall during the Long Rains. *Journal of Hydrometeorology*, 16: 2001-2012.

[23] A. M. Sharan. (2015). Prediction of Rain in Bihar, India, Based on Historical Rain Data. *Asian Journal of Water Environment and Pollution*, 12: 59-64.

[24] F. A. Semire, R. Mohd-Mokhtar, W. Ismail. (2015). Modeling of rain attenuation and site diversity predictions for tropical regions. *Annales Geophysicae*, 33: 321-331.

[25] K Badron, A.F. Ismail, M.R. Islam. (2015). A modified rain attenuation prediction model for tropical V-band satellite earth link. *International Journal of Satellite Communications and Networking*, 33: 57-67.

[26] X.D. Yu, X.G. Zhou. X.M. Wang. (2012). The Advances in the Nowcasting Techniques on Thunderstorms And Severe Convection. *Acta Meteorologica Sinica*, 03: 311-337.

[27] D.L. Zou, Y.R. Feng, Q.Q. Liang. (2014). Study of 0~3 Hour Short-Term Forecasting Algorithm for Rainfall. *Journal of Tropical Meteorology*, 2: 249-260.

[28] Y.R. Feng, H.S. Xu, Q.Q. Liang. (2013). A 0~6h Quantitative Snow(Rain) Forecast Technique and Its Application in Vancouver Winter Olympics. *Guangdong Meteorology*, 1: 6-13.

[29] A. Atencia, T. Rigo, A. Sairouni. (2010). Improving QPF by blending techniques at the Meteorological Service of Catalonia. *Natural Hazards and Earth System Sciences*, 10: 1443-1455.

[30] B. Casati, G. Ross, D. Stephenson. (2004). A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11: 141-154.

[31] C.H. Hu, J. Huang, Y.Q. Wang. (2015). Verification of Quantitative Precipitation Forecast Between Radar and Numerical Model Based on Intensity-Scale Method. *Journal of Tropical Meteorology*, 2: 273-279.

[32] W.W. James, Y. Feng, M. Chen. (2010). Nowcasting Challenges during the Beijing Olympics: Successes, Failures, and Implications for Future Nowcasting Systems. *Weather and Forecasting*, 25: 1691-1714.

[33] M.X. Chen M.X., F. Gao, R. Kong. (2010). Introduction of Auto-nowcasting System for Convective Storm and Its Performance in Beijing Olympics Meteorological Service. *Journal of Applied Meteorological Science*, 4: 395-404.

[34] S.Q. Yang, J.X. Rui, H.Z. Feng. (2006). application of support vector machine(svm) in rainfall categorical forecast. *Journal of Southwest Agricultural University (Natural Science). 28(2): 252-257.*

[35] *V. Lakshmanan, T.W. Humphrey*.(2014) A MapReduce Technique to Mosaic Continental-Scale Weather Radar Data in Real-Time. IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(2):721-732.

[36] C. Piani, J.O. Haerter, E. Coppola. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99(1-2):187-192.

[37] S. Venkadesh, G. Hoogenboom, W. Potter W, et al. (2013). A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks. *Applied Soft Computing Journal*. 13(5):2253-2260.

[38] L. Jiang, J. Wu. (2013). Hybrid PSO and GA for Neural Network Evolutionary in Monthly Rainfall Forecasting. *Intelligent Information and Database Systems*. Springer Berlin Heidelberg.

[39] A. Tsymbal, S. Puuronen. (2000). Bagging and Boosting with Dynamic Integration of Classifiers. *Principles of Data Mining and Knowledge Discovery*. Springer Berlin Heidelberg.

[40] M. Lazri, S. Ameur. (2018). Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. *atmospheric research*. 203: 118-129.

Xunlai Chen was Senior engineer of Meteorology at Shenzhen Meteorological Bureau. He received the Ph.D. at Sun Yat-sen University in 2007 and B.S. at Nanjing University in 2002. His research interests include Numerical weather prediction, nowcasting technique and numerical simulation of air pollution.

Guangjun He was born in Hubei, China, in 1987. He received the Ph.D. degree in remote sensing of environment from Nanjing University, Nanjing, China, in 2015. From 2015 to 2018, he was an assistant researcher fellow in the State Key Laboratory of Space-Ground Integrated Information Technology, CAST, Beijing, China, where he has been an associate researcher fellow since 2019. His research interests include image processing, multi-sensor information fusion, remote sensing of snow and data mining.

Yuanzhao Chen was Chief forecaster of Meteorology at Shenzhen Meteorological Bureau. His research interests include Numerical weather prediction and radar nowcasting technique.

Shuting Zhang was assistant engineer and weather forecaster of Shenzhen Meteorological Bureau. She obtained master degree in Atmospheric Physics and Atmospheric Environment in Sun Yat-sen University in 2016. Her research interests include Numerical weather prediction, air pollution nowcasting technique and numerical simulation.

J.S. Chen, Ph.D, Research, director of Center for Spatial-information. He graduated from the Chinese Academy of Social Sciences in 2004 with a doctorate degree. From 2004 to 2011, he worked in the Chinese university of Hong Kong as a postdoctoral and professor. Then, he joined the Shenzhen institutes of advanced technology, Chinese academy of sciences. His research interests include spatial and environmental big data processing methods; multi-source spatial and environmental data assimilation and information fusion methods.

Jing Qian received the Ph.D. degrees from Geography department of HongKong Baptist University in 2012. From 2011.01 to 2011.03, she was the visiting scholarship of VITO in Belgium, and from 2014.10 to 2014.12, she was the visiting scholarship of the Technische Univesitaet München in Germany. She has been the associate researcher fellow in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include spatial hydrology, remote sensing and spatial data mining.

Haicong Yu obtained his Ph.D. degrees from Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences in 2012. He received the B.Sc. degree in GIS from Zhengzhou Institute of Surveying and Mapping, Information Engineering University, Zhengzhou, China, in 2006 and the M.Sc. degree in GIS from LIESMARS, Wuhan University, Wuhan, China, in 2008. He is a senior GIS engineer in Center for Assessment and Development of Real Estate, Shenzhen. His current research focuses on applications in image processing and GIS spatial analysis and modeling.